The role of RNA structure in Chikungunya virus early replication events

Catherine Theresa Kendall

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

School of Molecular and Cellular Biology

Faculty of Biological Sciences

The University of Leeds

December, 2019

Acknowledgements

Many thanks to the White Rose BBSRC DTP for funding this PhD, including my three-month stint with the Feds. Thanks also to Professor Andres Merits (University of Tartu) and Professor Alain Kohl (University of Glasgow) for reagents and contributing to the formulation of this project.

First and foremost, I would like to thank my supervisor Dr Andrew Tuplin. Andy has been instrumental to my success, valuing my ideas and encouraging my development. His sense of humour and passion for the project made the good moments great and his kind support made the bad moments bearable. Thank you, Tuppers, from the bottom of my heart.

Thank you to my co-supervisor Professor Nicola Stonehouse for her fastidious advice, valuable input and friendly counsel.

I would also like to thank Dr Julie Aspden and Professor Mark Harris for their guidance.

I would like to thank Dr Marietta Müller, Samuel Bradley, Dominic Banda, Henna Khalid, Oliver Prosser, Kate Loveday, Oliver Antoniak, Emily Levitt and Trang Nguyen of the Tuplin group, and Dr Lorna Kelly and Dr Ray Li, for their valuable input and collaboration.

To Rebecca, Mum, Sam and Tim, thank you for believing in me and patiently trying to understand my explanations (and diagrams!).

To Frank, thanks for making me laugh when I was too tired to go on.

To Dr Ethan Morgan, thank you for the many useful discussions throughout the project and for proofreading this thesis. More than that, thank you for your unwavering belief in me and making life outside of work worthwhile. Your support means everything to me.

Abstract

Chikungunya virus (CHIKV) is a pathogenic, single-stranded, positive-sense RNA virus transmitted to humans by *Aedes* spp. mosquitoes. After decades of low-level endemic circulation, CHIKV has re-emerged to establish local transmission on five continents, infecting upwards of 3,000,000 people (1). There are no currently available vaccines or direct-acting anti-viral therapeutic agents. A greater understanding of the CHIKV replication cycle is essential, as much of what is known about the replication cycle is assumed from studies of related but divergent viruses, which have provided conflicting reports. Preliminary work carried out by A. Tuplin (University of Leeds) suggested a highly ordered structured region at the 5′ end of the CHIKV genome, spanning ~300 nt including the 5′ untranslated region (UTR) and the 5′ coding sequence of *nsp1*. The aim of this project was to determine the phenotypic importance of secondary structure in this region for the CHIKV lifecycle in human and mosquito cells at multiple stages of viral replication and to investigate the sequence and structure requirements for functional interactions. This study represents the first investigation of functional elements within the 5′ UTR and adjacent nsP1-coding region in CHIKV.

Taking a structure-led reverse genetic approach, in both infectious virus and sub-genomic replicon systems, the wild-type secondary structure of the 5′ 300 nt of the CHIKV genome was found to be essential for genome replication in human- and *Ae. albopictus*-derived cells. Six RNA stem-loops were determined to individually enhance CHIKV genome replication - including novel structures analysed for the first time in this study. Comparative analysis in human and mosquito-derived cell lines revealed that the novel stem-loop SL47 in the 5′ UTR functions in a host-independent manner while stem-loops in *nsp1* function in a host-dependent manner. Stem-loops were demonstrated to function within the positive-strand genomic RNA, via predominantly structure-dependent mechanisms. Furthermore, single-host passaging studies suggested strong selection pressure to regenerate secondary structures and highlighted potential differences in translational recoding between host species. Finally, the potential for tertiary structure formation was explored. In addition to furthering knowledge of fundamental aspects of the molecular virology of this important human pathogen, this study will inform rational design of a genetically stable attenuated vaccine.

List of Figures

Abbreviations

A226V	alanine to valine at position 226 in envelope protein E1
Ae.	*Aedes*
AGO	autocidal gravid ovitraps
AUD	alphavirus unique domain
BHK-21	baby hamster kidney fibroblasts
CaMV	cauliflower mosaic virus
CD2AP	CD2-associated protein
CDC	center for disease control
CHIKV	Chikungunya virus
CP	capsid protein
CRE	conserved replication element
CRM1	chromosomal maintenance 1 protein
CS	cyclisation sequences
CSE	conserved sequence element
CTE	*cis*-acting constitutive element
DAR	downstream of AUG
DEET	N,N-diethyl-3-methylbenzamide
DENV	dengue virus
DEPC	diethyl pyrocarbonate
DMS	dimethylsulphate
DR1-3	direct repeat sequences 1-3
DSH	downstream stable hairpin
E1-3	envelope proteins 1-3
EEEV	Eastern equine encephalitis virus
ECSA	East Central South African
EEEV	eastern equine encephalitis virus
eEF1A	eukaryotic translation elongation factor 1 α
eIF	eukaryotic translation initiation factor
EILV	Eilat virus
EMSA	electrophoretic mobility shift assay
FDA	Food and Drug Administration
FHL1	four and a half LIM domain protein 1
FWD	forward primer
G3BP1 and 3	GTPase Activating Protein (SH3 Domain) Binding Protein 1/3
gRNA	genomic ribonucleic acid
GTase	guanylyltransferase
HCV	Hepatitis C virus
HDX	hydrogen-to-deuterium exchange
HIV-1	Human immunodeficiency virus 1
HuR	human-antigen R
HVD	hypervariable domain
IC	infectious clone
IFIT-1	interferon-induced protein with tetratricopeptide repeats 1

IFN	interferon
IGR	intergenic region
IOL	Indian Ocean lineage
IRES	internal ribosome entry site
L-15	Leibovitz media
m^7G	N^7-methylguanosine
mAbs	monoclonal antibodies
MAYV	Mayaro virus
MC	methylcellulose
MHV	Mouse Hepatitis virus
MOI	multiplicity of infection
MOPS	3-N-morpholinopropanesulphonic acid
MTase	guanine-7-methyltransferase
MV	Measles virus
Mxra8	matrix remodeling-associated 8
NC	nucleocapsid
NLS	nuclear localization sequence
NMIA	N-methylisatoic anhydride
NMR	Nuclear Magnetic Resonance
NS5B	non-structural protein 5B
NSC	NSC260594, a quinolinium derivative
nsP	non-structural protein
NXF1	Nuclear RNA Export Factor 1
ONNV	O′nyong nyong virus
ORF	open reading frame
P( )	passage( )
pE2	precursor of envelope protein 2
PFU	plaque-forming units
PK	pseudoknot
Pk1/Pk2	pseudoknot-1, pseudoknot-2
PKR	protein kinase R
PLB	passive lysis buffer
PPi	pyrophosphate
ppp	triphosphate
PRF	programmed ribosomal frameshifting
PV	Poliovirus
qRT-PCR	quantitative reverse transcription polymerase chain reaction
RdRp	RNA-dependent RNA polymerase
REV	reverse primer
RIG-I	retinoic acid-inducible protein I
RNAi	RNA interference
rRNA	ribosomal ribonucleic acid
RRV	Ross River virus
RTE	RNA transport element
SFV	Semliki Forest virus
sgRNA	sub-genomic ribonucleic acid

SHAPE	selective 2′-hydroxyl acylation analysed by primer extension
sHP	small hairpin
SINV	Sindbis virus
SL	stem-loop
TAR	*trans*-activation response element
Tat	*trans*-activator of transcription
TATase	terminal adenosyl transferase
TCR	termination codon readthrough
TF	trans-frame
TIM-1	T-cell immunoglobulin and mucin domain 1 protein
TLS	tRNA-like structure
TM	trans-membrane
TPB	tryptose phosphate broth
TYMV	turnip yellow mosaic virus
U	enzyme units
UAR	upstream of AUG
UTR	untranslated region
VEEV	Venezuelan equine encephalitis virus
VLP	virus-like particle
WA	West African
WEEV	Western equine encephalitis virus
WHO	World Health Organisation
WT	wild-type
YIPP	yeast inorganic pyrophosphatase
ZIKV	Zika virus

Table of Contents

Chapter 1. Introduction

1.1 General introduction

Chikungunya virus (CHIKV) is an arthropod-borne (arbo-) virus which poses a significant threat to public health. After ~60 years of endemic circulation in Africa and Asia, CHIKV has entered naïve populations in Europe and the Americas, resulting in nearly 3,000,000 cases of Chikungunya fever (1). At present there are no licensed vaccines or direct-acting therapeutic agents against CHIKV. Very little is known about the mechanism of CHIKV replication and the requirements for productive infection of both the human and mosquito host. Therefore, a better mechanistic understanding of CHIKV replication could identify potential attenuation sites for vaccine design or novel drug targets for prophylactic use or post-exposure treatment. Conserved RNA structures within the CHIKV genome represent a potential target.

In this chapter, CHIKV is described within virus phylogeny, followed by a discussion of its emergence, clinical features, vector specificity and control. The molecular virology of the virus is then described in detail including the viral genome, encoded proteins and replication cycle. Additionally, RNA structure is introduced, primarily in the context of mammalian virus replication. Finally, known and predicted RNA structures in the CHIKV genome are described and are summarised in Fig 1.21.

1.1.1 The *Togaviridae* family

The *Togaviridae* are a family of positive-sense, single-stranded, enveloped RNA viruses. The togavirus genome consists of a single RNA molecule with 5′ and 3′ untranslated regions (UTRs), terminating in a 5′ m^7G (N^7-methylguanosine) cap and 3′ polyadenylate tail. The genome contains two open reading frames (ORF) separated by an intergenic region. The *Rubivirus* genus was previously included in the *Togaviridae*, but has since been recategorised as *Matonaviridae* (2). The *Togaviridae* family currently comprises a single genus, the *Alphavirus* genus consisting of 31 recognised species in distinct antigenic complexes based on antibody cross-reactivity (Fig 1.1). Alphaviruses are transmitted by mosquitoes with the exception of two aquatic alphaviruses: Salmon pancreatic disease virus and Southern elephant seal virus (3,4). Whole-genome phylogenetic analysis of the entire *Alphavirus* genus places marine alphaviruses

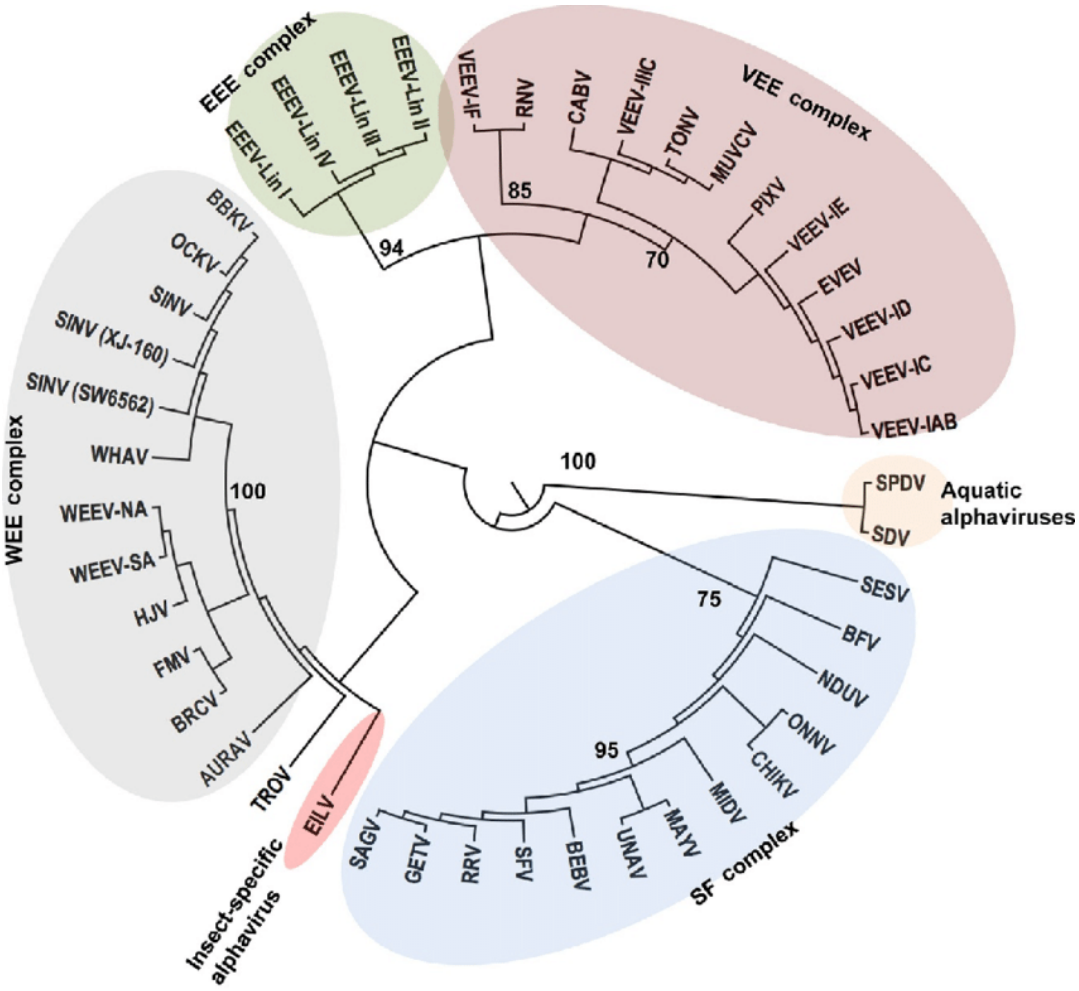


Figure 1.1: The *Alphavirus* genus (5)**.** Phylogeny of the *Alphavirus* genus based on whole-genome analysis, excluding the hypervariable domain of nsP3. Bootstrap values are denoted for major branchpoints, indicating the number of times per hundred that the same branch was observed during re-sampling of the dataset. Viruses are grouped as sero-complexes (coloured ovals, labelled), where cross-reactivity is observed during haemagglutination assays. The aquatic alphaviruses are basal (peach). The WEE complex contains New World (WEEV) and Old World (SINV) viruses due to the recombination event that gave rise to WEEV.

at a basal position, suggesting that the ancestral alphavirus was aquatic (6). While marine alphaviruses are thought to be transmitted by lice, the majority of recognised alphavirus species are transmitted by haematophagous mosquitoes to vertebrates during bloodmeals (3,7).

1.1.1.1 New World and Old World alphaviruses

Alphaviruses have historically been classified into Old World and New World alphaviruses based on endemic regions, although geographical classification is becoming less useful given recent translocation events. New World alphaviruses are endemic to the USA, the Caribbean and South America, where they infect mosquitoes, horses, humans, and birds. Examples include Venezuelan equine encephalitis virus (VEEV), eastern equine encephalitis virus (EEEV) and western equine encephalitis virus (WEEV). New World alphaviruses are associated with neurotropic disease, characterised by acute onset fever and headache progressing to altered mental state, paralysis, seizures and coma. The severity of disease in humans varies greatly within the New World alphaviruses: ~70% mortality for EEEV and ~4% for WEEV (8). In contrast, infection with Old World alphaviruses is rarely fatal and is typically associated with arthritogenic disease, characterised by maculopapular rash, severe fever and chronic musculoskeletal pain. Prior to recent epidemics, Old World alphaviruses such as CHIKV, Semliki Forest virus (SFV) and Sindbis virus (SINV) were restricted to Africa, Asia and Australia. Reservoirs of the Old World alphaviruses are much more diverse, including non-human primates, rodents, horses, wallabies and kangaroos (9,10).

The serological cross-reactivity of SINV (an Old World virus) and WEEV (a New World virus) groups them into the WEE complex (Fig 1.1). This is due to a recombination event, whereby an EEEV-like ancestral virus acquired SINV-like envelope proteins (11). Aside from this recombination event, New World and Old World alphaviruses are fairly divergent, sharing ~40% sequence identity in structural proteins and 60-80% identity in non-structural proteins (12). However, conservation of RNA structures has been demonstrated by selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) mapping in viruses as divergent as VEEV and SINV (13). SHAPE mapping determines the likelihood of base-pairing at single nucleotide resolution via reactivity of each nucleotide with N-methylisatoic anhydride (NMIA) and is described in more detail in 1.3.2.

1.1.1.2 Insect-specific alphaviruses

The first and only recognised insect-specific alphavirus, Eilat virus (EILV), was discovered in 2012 in *Anopheles* mosquitoes (14). During the last two years, four additional insect-specific alphaviruses have been discovered: Taï forest and Mwinilunga alphaviruses in Africa, and Agua Salud and Caainguá alphaviruses in South America (15–18). Almost every recognised alphavirus species is pathogenic to humans, livestock or wildlife. Bias in virus discovery has resulted in an underrepresentation of non-pathogenic, insect-specific and marine alphaviruses within the genus. Further expansion of the *Alphavirus* genus will provide insight into virulence factors of currently recognised members.

1.1.1.3 CHIKV phylogeny

CHIKV comprises three genotypes based on nucleic acid sequence and geographical origin: Asian, West African (WA) and East, Central and South African (ECSA) (19) (Fig 1.2). CHIKV is well conserved across the genotypes. Intra-lineage comparisons demonstrate up to 99.8% nucleotide sequence identity within envelope protein genes, while inter-lineage comparisons indicate genetic divergence of up to 15% (20). However, even the most divergent CHIKV strains share a minimum of 95% amino acid (aa) identity in the envelope proteins (21). Despite genetic similarity between genotypes, relatively few amino acid substitutions are necessary to alter key processes of pathogenesis and virulence. For example, Asian CHIKV exhibited increased neurovirulence compared to ECSA CHIKV following intracerebral inoculation of mice as a result of a 7 aa deletion in nsP3 (22). Comparison of pathology in vertebrates and altered vector specificity suggests the emergence of a distinct lineage within the ECSA genotype, the Indian Ocean lineage (IOL) (23) (Fig 1.2).

Comparative studies of virulence in the three genotypes, including the ECSA-derived IOL lineage, demonstrated higher virulence for WA genotypes and lower virulence associated with Asian CHIKV infection of mice (24). Vaccination of cynomolgus macaques with an experimental CHIKV/IRESv1 vaccine based on the IOL lineage (discussed in 1.1.5.1) provided cross-protection against lethal challenge with CHIKV of all genotypes and similar cross-protection was observed in mice. Cross-protection against all CHIKV lineages following vaccination agrees with observations that induction of humoral immunity following natural CHIKV infection is associated

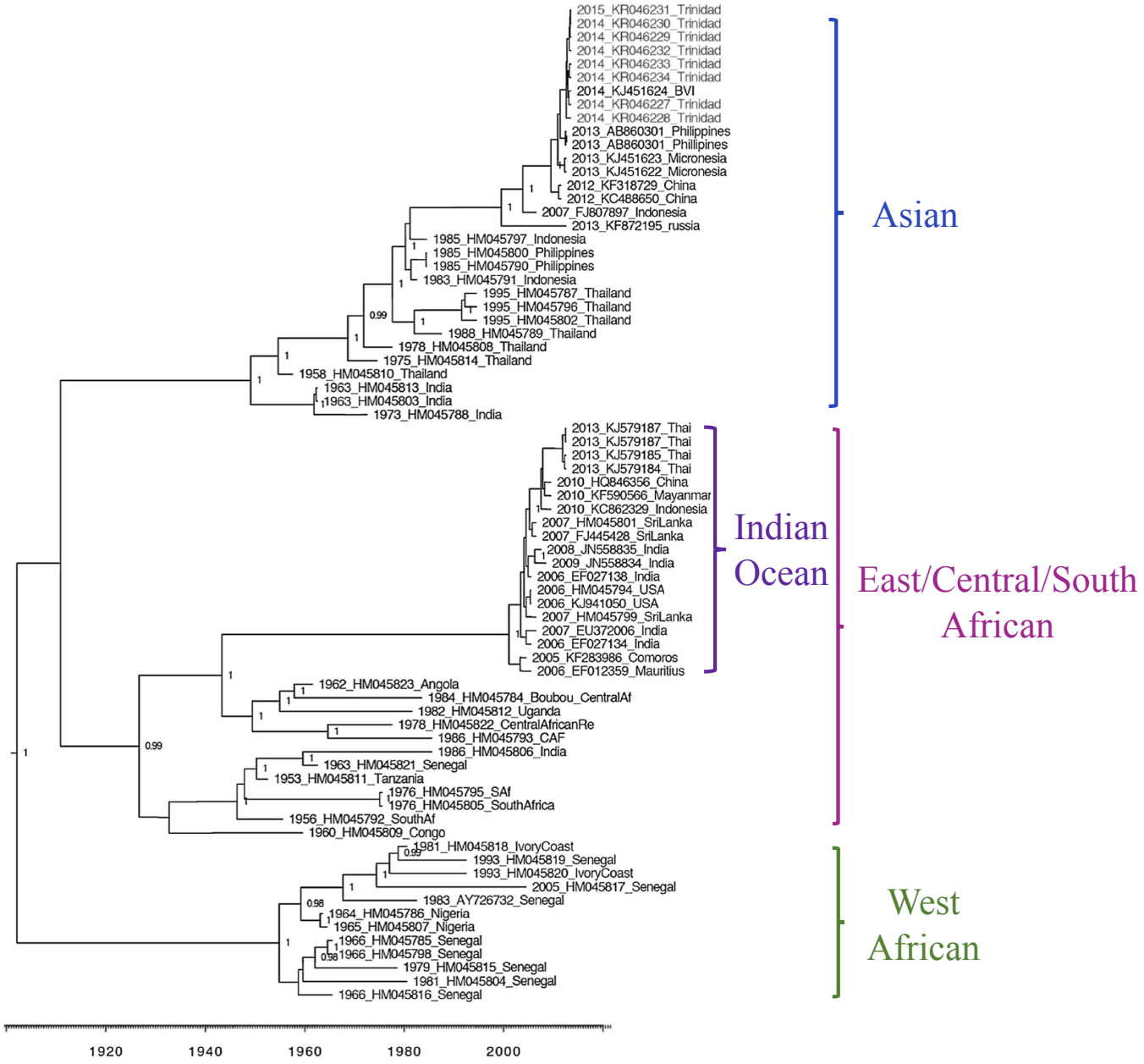


Figure 1.2: Phylogenetic tree of CHIKV strains based on amino acid identity of the complete coding region (adapted from (23)). Three genotypes are shown: Asian, West African, East, Central and South African (ECSA). The Indian Ocean lineage arose from the ECSA, thus is shown within the genotype, but exhibits more severe pathogenicity. The scale bar (bottom) represents time in years, with CHIKV strains isolated between 1953 and 2015.

with lifelong immunity (25,26). In spite of high genetic similarity between genotypes, lineage-specific differences in the length and identity of the 3′ untranslated region (UTR) have been determined in the number of repeated *cis*-acting structures and sequences (27,28) (discussed in 1.3.3). In contrast, the identity and length of the 5′ UTR is highly conserved across the lineages.

1.1.2 CHIKV epidemiology

1.1.2.1 Emergence of CHIKV

CHIKV was first isolated in 1952 in the Newala district of modern-day Tanzania (29,30). Following the Newala epidemic, frequent epidemics occurred between 1952 and 1980 in fifteen African states (23,31) (Fig 1.3). CHIKV presumably emerged as a zoonotic infection from non-human primates, transmitted by arboreal *Aedes* spp. mosquitoes. CHIKV has been isolated from *Ae. furcifer* and *Ae. africanus* in sylvatic cycle with non-human primates: African green monkeys, mandrills, red-tail monkeys, Guinea baboons and a bushbaby (32–37). In contrast, CHIKV in Asia is primarily transmitted by the peridomestic *Ae. aegypti* and *Ae. albopictus* in an urban transmission cycle with no identified reservoir host. The first Asian epidemic of CHIKV was reported in Thailand in 1958, although recent whole-genome phylogenetic analysis of CHIKV strains estimates introduction as early as 1920 (38,39). Subsequent endemic spread was described in Southeast Asia and India (40,41) (Fig 1.3). In endemic regions of Africa and Asia, major outbreaks ceased from ~1980-2000.

1.1.2.2 Re-emergence of CHIKV

A resurgence in CHIKV infections began with an epidemic in the Democratic Republic of Congo in 2000 (42). In 2004, a large epidemic began in Kenya and spread rapidly to islands in the Indian Ocean including La Réunion in 2005 (Fig 1.4A). The La Réunion CHIKV strain diverged from the ECSA sufficiently to be termed the Indian Ocean lineage (IOL). The IOL was associated with more severe disease: neurological symptoms, fulminant hepatitis and higher mortality (43,44). The IOL also developed adaptive mutations enabling higher rates of replication in *Ae. albopictus* (45). Naïve populations, where *Ae. albopictus* was the only competent vector, exhibited high attack rates, between 35-70% (44,46). Subsequently, the IOL circulated in India, China and Southeast Asia as well as infecting urban populations of African *Ae. albopictus* (20,47–49).

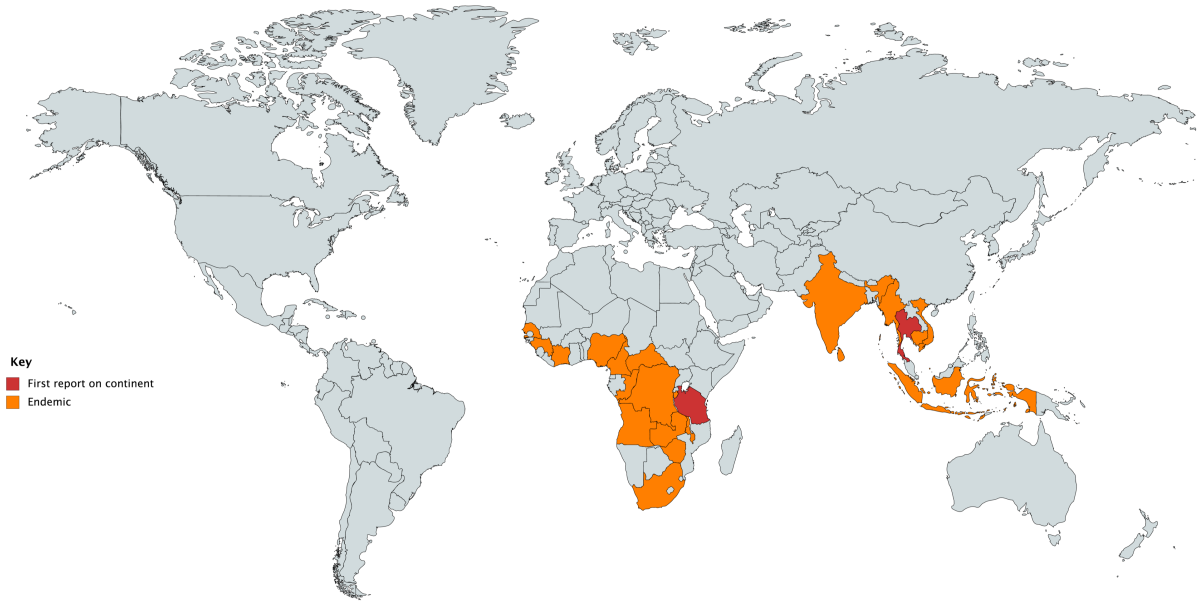


Figure 1.3: Map of CHIKV endemic areas prior to 2004. Countries which had reported local transmission of CHIKV prior to 2004 are coloured according to the first reports in Africa and Asia (red) and subsequent reports of clinical disease (orange). Grey colouring indicates no local transmission. Imported cases are not shown. Map produced using mapchart.net using data collated from: (19,23,31).

A second CHIKV re-emergence took place in 2013 when an epidemic strain of Asian genotype CHIKV circulating in China and the Philippines was introduced to the Caribbean island of Saint Martin (50). Within 3 years, autochthonous transmission of CHIKV was reported in 45 countries in the Caribbean, North, South and Central America (1) (Fig 1.4B). A third translocation event occurred in 2014 when an ECSA strain from Cameroon was introduced to Brazil and Haiti (28,51–53) (Fig 1.4C). Concurrent epidemics resulting from distinct CHIKV lineages highlight the need for pan-lineage CHIKV treatments and vaccines with targets such as the RNA-dependent RNA polymerase and conserved RNA structures (described in 1.2.5.4 and 1.3.2).

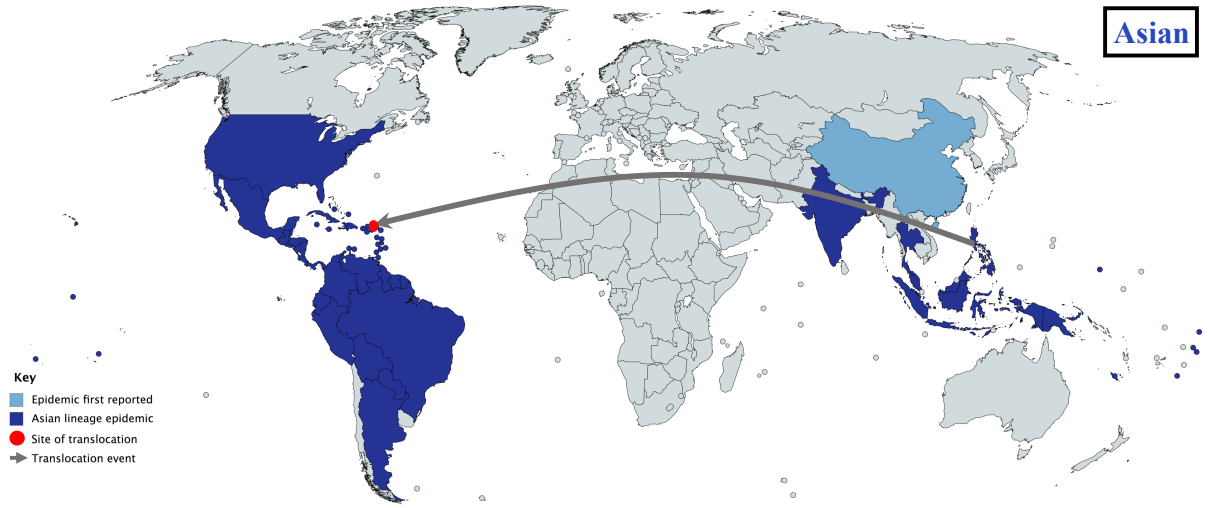
As of 2019, only two European countries have reported autochthonous transmission of CHIKV, by *Ae. albopictus* in both instances (54,55) (Fig 1.5). A CHIKV outbreak in northern Italy in the summer of 2007 resulted in 217 symptomatic cases (54). The A226V ECSA CHIKV strain was isolated from local *Ae. albopictus* populations, confirming autochthonous transmission (56). Vector control measures limited the outbreak, halting local transmission. A second outbreak occurred in central Italy a decade later in the summer of 2017, leading to a secondary outbreak in southern Italy with a total number of symptomatic cases exceeding 400 (57).

Autochthonous transmission of CHIKV in Europe was also demonstrated during three small epidemics of Chikungunya fever in southern France in 2011, 2014 and 2017 (55,58,59). Temperate *Ae. albopictus* captured in France were found to transmit CHIKV with high efficiency, comparable to that of *Ae. aegypti* (60). Local transmission was prevented in France by the same aggressive vector control as in Italy (61). The CHIKV-competent vector *Ae. albopictus* is present in almost 20 European countries, rendering Europe vulnerable to CHIKV epidemics. Furthermore, global warming has enabled survival of vector species in previously inhospitable regions, as demonstrated by the recent discovery of *Ae. albopictus* eggs and adults in southern England (62,63). European epidemics would involve immunologically naïve populations, which are associated with increased morbidity and mortality (64–66). Although arboviral research often utilises *Ae. aegypti*-derived cells, *Ae. albopictus* represents the most likely vector of CHIKV epidemics in Europe.

(A)


IOL
Key
Epidemic first reported
Indian Ocean Lineage epidemic
Site of translocation
Translocation event

(B)


Asian
Key
Epidemic first reported
Asian lineage epidemic
Site of translocation
Translocation event

(C)

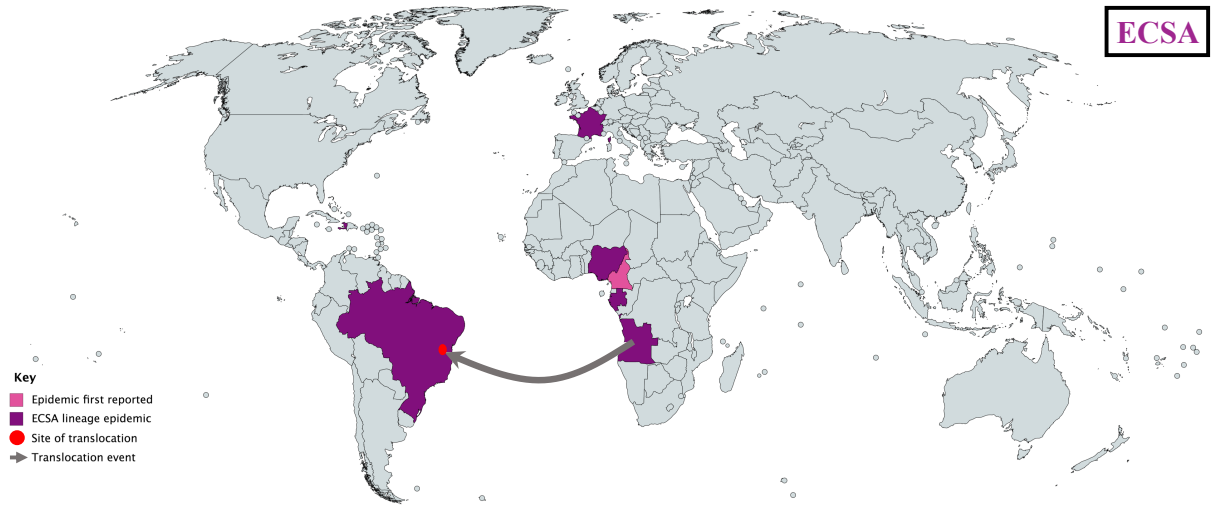


Figure 1.4: Map of CHIKV re-emergence from 2004-2019 during three separate pandemics. The lineage responsible for each pandemic is displayed in the top right (black box). Major translocation events (grey arrows) and significant locations (red dots) are displayed for each pandemic. **(A)** The 2005-2006 La Réunion (red dot) epidemic of the Indian Ocean lineage began in Kenya (light purple) and was reported in many countries throughout Asia (dark purple). **(B)** The 2012 Asian genotype epidemic began in China (light blue) and spread throughout the Americas, beginning in Saint Martin (red dot). Countries reporting local transmission are coloured (dark blue). Alaska remains grey due to geographical separation and lack of cases. **(C)** The 2014 ECSA epidemic in Brazil (red dot) began in Cameroon (light pink) and was transmitted by travelers returning from Angola, spreading to several countries (dark pink). Map produced using mapchart.net using data collated from: (1,19,23,28,31,52,53).

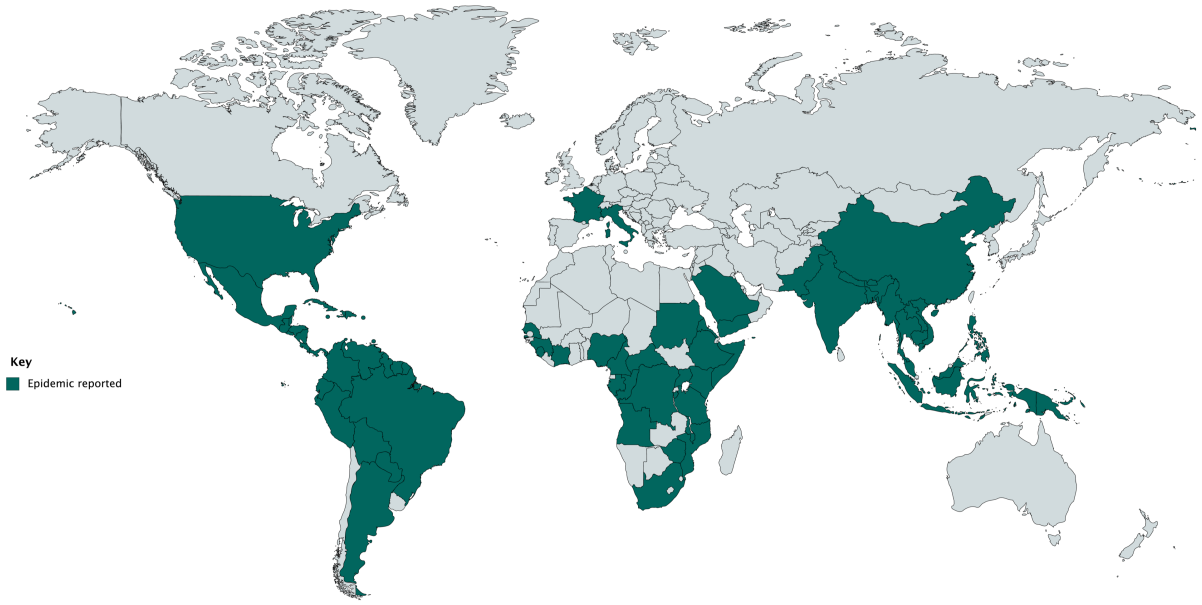


Figure 1.5: Worldwide distribution of CHIKV in 2019. Countries which have reported autochthonous, local transmission of CHIKV as of September 2019 are coloured (green). Grey colouring indicates no local transmission. Imported cases are not shown. Map produced using mapchart.net using data from the centre for disease control (CDC), 2019 (67,68).

1.1.3 The pathogenesis of CHIKV infection in humans

The name 'Chikungunya' derives from the Kimakonde dialect of southern Tanzania, meaning "to walk bent over" due to the distinctive bent posture of sufferers. Following transmission by a mosquito bite, CHIKV replicates in fibroblasts and macrophages at the site of inoculation, producing viraemia up to 10^8 virions/mL of blood (69) (Fig 1.6). The virus then disseminates to the muscle, joints, brain, liver and lymphoid tissue (70,71).

After an incubation period of 2-4 days, a sudden onset of clinical disease is observed in the majority of infections; around 25% remain asymptomatic (72–74). Increasing viraemia coincides with the acute phase of infection, characterised by high fever (39-40°C), maculopapular rash and debilitating, symmetric polyarthralgia in the wrists, ankles, hands and feet (75) (Fig 1.7). The painful symptoms of CHIKV are a result of tissue injury induced by cytopathic infection and the inflammatory immune response (69). Additional symptoms include arthritis, headache, myalgia, back pain, conjunctivitis and gastrointestinal distress. Fatal CHIKV infections and severe complications such as hepatitis, haemorrhage, encephalopathy and myocarditis were recorded for the first time during the La Réunion epidemic of 2005-6 (254 deaths) and the Indian epidemic of 2006 (76–78). Pregnant women, neonates, the elderly and patients with pre-existing chronic illnesses were particularly at risk for complications (78,79).

CHIKV infection is typically self-limiting. Rising viral titre induces a strong type I interferon response leading to clearance of viraemia within one week and resolution of acute symptoms within two weeks (80). CHIKV-specific adaptive immunity (immunoglobulins IgM and IgG) is detectable at the time of viral clearance, associated with life-long immunity against all CHIKV lineages (26). Following the acute phase, ~60% patients experience chronic arthralgia for an average of 9 months (81–83). Post-Chikungunya chronic inflammatory rheumatism is more common in people over 35 and those with previous injury, where arthritis may result in deformity (84,85) (Fig 1.8). Ongoing chronic inflammation in the joints is thought to be an induced autoimmune response; CHIKV was not detectable in the synovial fluid of 38 patients with chronic arthralgia 22 months post-infection (86) However, macrophages have been identified as a cellular reservoir of persistent CHIKV infection in the muscle, joints and liver of chronically infected cynomolgus macaques (70) and one La Réunion patient (82).

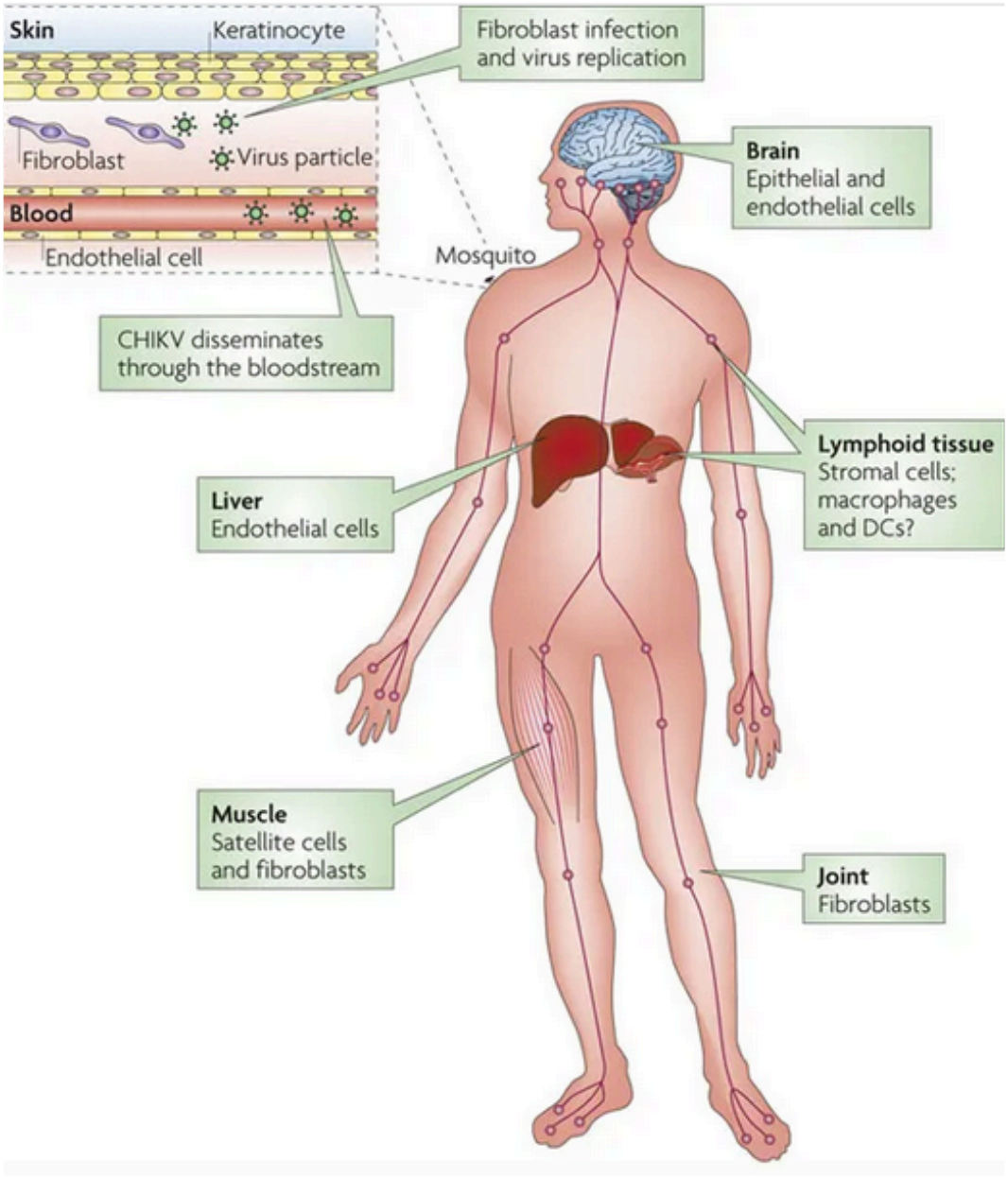


Figure 1.6: Dissemination of CHIKV in vertebrates (87)**.** CHIKV is transmitted by mosquito bite, followed by replication in fibroblasts in the skin (dashed box). CHIKV then disseminates into cells of the liver, muscle, joints, lymphoid tissue and brain, as labelled.

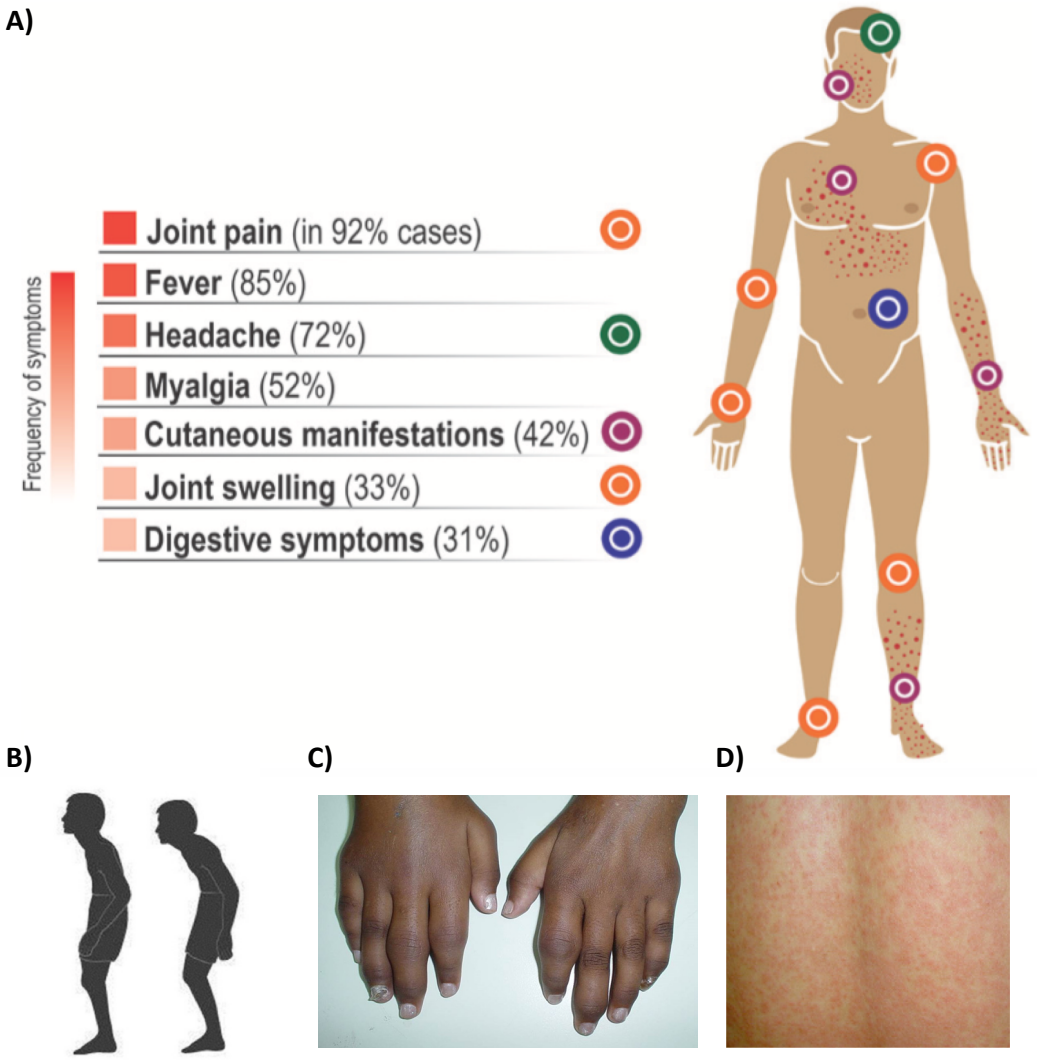
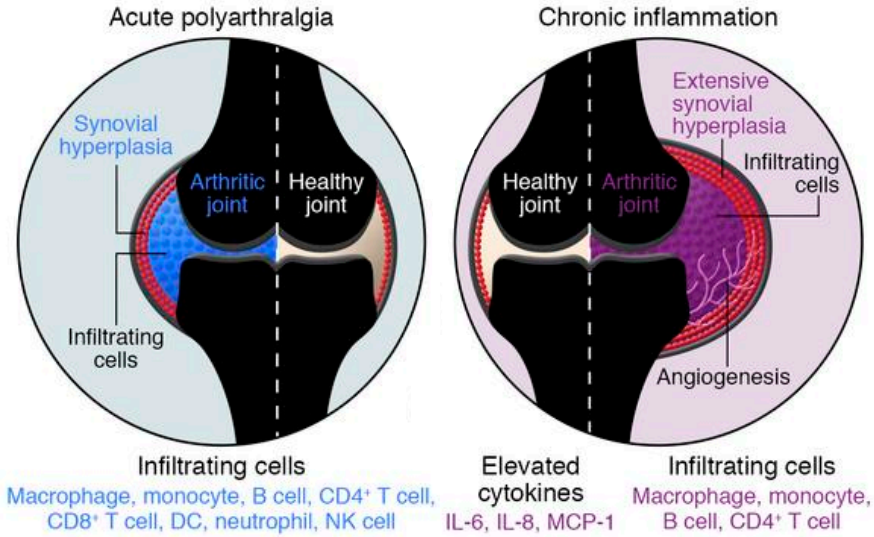


Figure 1.7: Clinical presentation of Chikungunya fever (adapted from (88)**). (A)** Percentage of symptomatic patients displaying a range of symptoms, in decreasing order of frequency: (89,90). A schematic representation of a symptomatic patient is shown, coloured circles represent the typical location of symptoms. Symptoms of Chikungunya fever include **(B)** a stooped posture adopted to relieve arthralgia, **(C)** swollen, painful joints, especially wrists and phalanges, and **(D)** maculopapular rash.

A)



B)

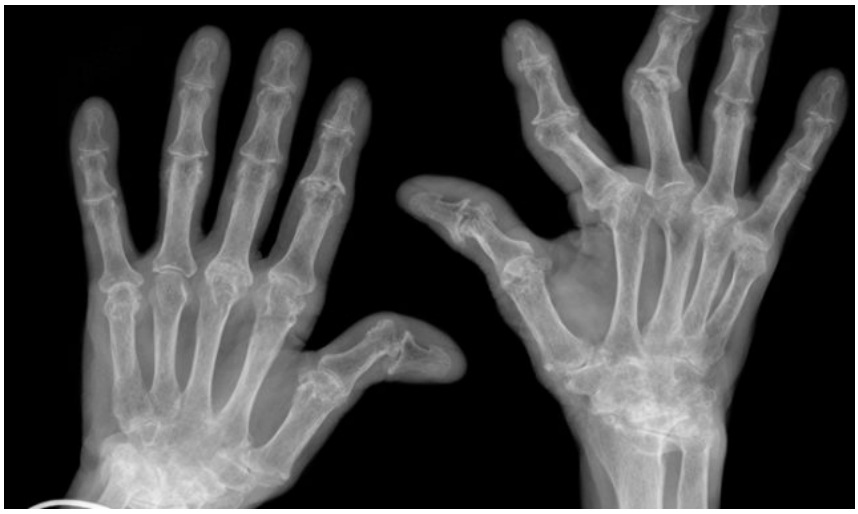


Figure 1.8: Model of post-infection chronic inflammation and arthritis. (A) Schematic representation of acute inflammation and post-Chikungunya chronic inflammation. The synovial fluid of the chronically inflamed joint retains a subset of leukocytes and increased levels of pro-inflammatory mediators such as IL-6 which contribute to prolonged inflammation, swelling and pain. The pathology of the joint mirrors that of rheumatoid arthritis (adapted from (91)) **(B)** Hand X-Ray of patient with chronic, debilitating post-infection arthritis (85).

1.1.4 Replication of CHIKV within *Aedes* spp. mosquitoes

In addition to replication in humans and non-human primates, CHIKV replicates within *Aedes* spp. mosquitoes; they are biological vectors and do not mechanically transfer CHIKV by passive means (Fig 1.9). After a female mosquito ingests an infectious bloodmeal, CHIKV penetrates the midgut by infection of epithelial cells or, more rarely, by extravasation directly into the haemocoel (92). Dissemination of the virus through the haemocoel leads to rapid infection of secondary tissues including the fat body, muscle, haemocytes and salivary glands (93). Infectious CHIKV in the salivary glands is detectable 48 hours post-ingestion by immunostaining and plaque assay (7,94). In contrast to the acute pathogenic infection seen in vertebrates, CHIKV infection of *Aedes* spp. mosquitoes is typically lifelong and asymptomatic despite extensive, disseminated viremia (95). One exception to apathogenic CHIKV infection in mosquitoes was observed during experimental infection with the IOL epidemic strain. *Ae. albopictus* populations from La Réunion infected with A226V-IOL CHIKV died 7-9 days sooner than uninfected counterparts, representing loss of a quarter of the typical lifespan. Interestingly, no pathogenesis was observed during infection of *Ae. albopictus* or *Ae. aegypti* from nearby Mayotte island, suggesting intra-species differences in vector competence and pathogenesis. Mayotte island, located a similar distance from Madagascar, experienced an epidemic during the same period. *Ae. albopictus* from La Réunion exhibited a higher concentration of CHIKV in the salivary glands than those from Mayotte, in addition to higher concentrations of CHIKV disseminated through the organism (96).

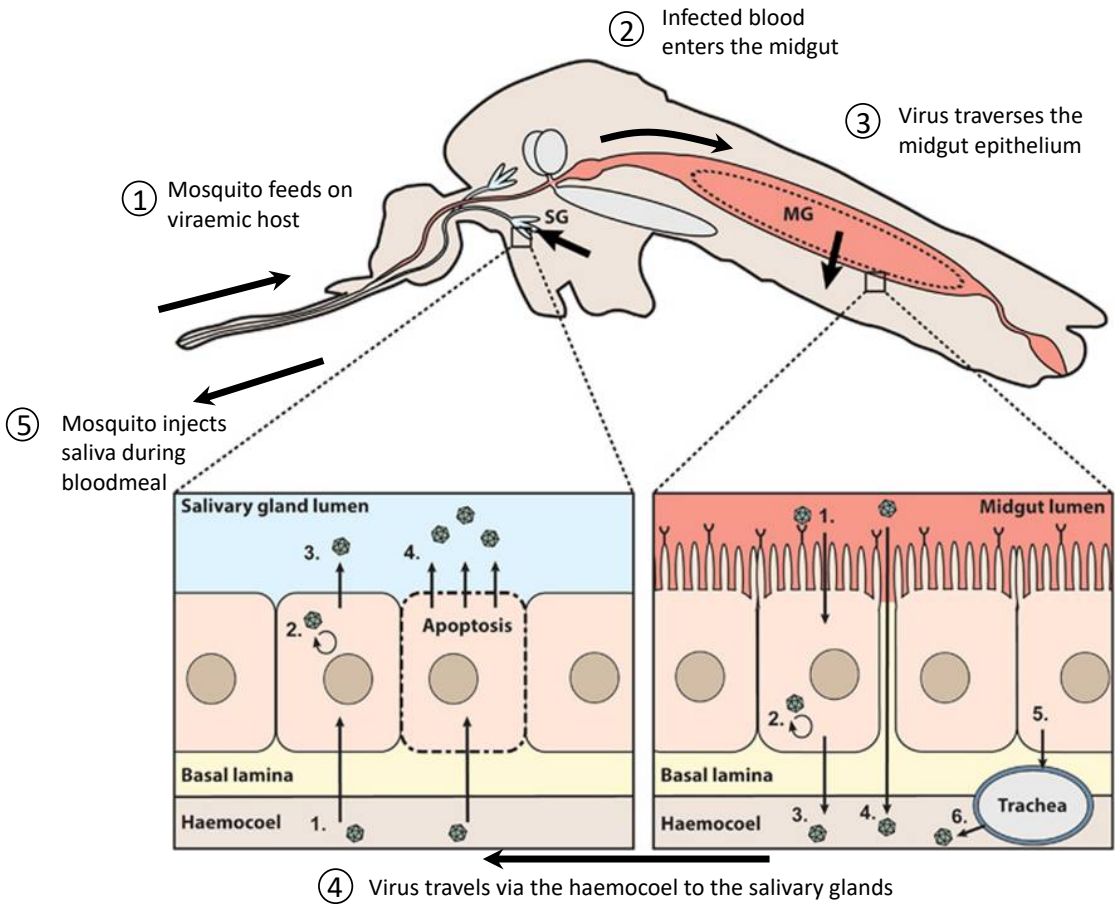


Figure 1.9: Schematic representation of CHIKV replication within *Aedes* spp. mosquitoes (adapted from (97)). A longitudinal cross section of an *Aedes* spp. mosquito is shown, with labelled midgut (MG) and salivary glands (SG). The passage of virions is denoted by black arrows and described in 5 stages (circled numbers). Within the midgut, the peritrophic membrane formed after a bloodmeal is shown (dashed circle). Escape of CHIKV from the midgut lumen (bottom right): (1) CHIKV infects midgut epithelial cells, (2) replicates and (3) buds into the haemocoel through the basal lamina. Alternatively, (4) 'leaky' midgut may allow direct extravasation or (5) CHIKV may infect the trachea before (6) entering the haemocoel. Entry into the salivary gland lumen (bottom left): (1) CHIKV crosses the basal lamina and infects salivary gland epithelium, (2) replicates and (3) buds into the lumen or (4) is released during apoptosis of infected cells.

Vector competence refers to the suitability of a species to transmit a pathogen. This property depends on appropriate cellular biology, anatomy and behaviour of the vector species. A competent vector must display intracellular permissivity to infection, high levels of replication within the organism and efficient, frequent transmission to susceptible hosts (98). *Aedes* spp. mosquitoes facilitate high, continuous levels of CHIKV replication and dissemination to the salivary glands (34). Within the *Aedes* genus, the behaviour of *Ae. aeygpti* confers the greatest vector competence, especially compared to arboreal and rural mosquitoes (99). *Ae. aegypti* is exclusively anthropophilic and takes multiple bloodmeals per gonotrophic cycle, enabling frequent transmission between humans (100). *Ae. aegypti* also demonstrates a preference for artificial larval habitats in urban environments, such as water containers, increasing levels of contact with humans. In comparison, *Ae. albopictus* is more ecologically flexible; it is found in urban and rural environments, tropical and temperate climates and uses a range of natural and artificial larval habitats (101). *Ae. albopictus* is not exclusively anthropophilic and typically ingests a single blood meal per gonotrophic cycle, reducing the likelihood of human-to-human transmission. However, *Ae. albopictus* is more susceptible to CHIKV infection than *Ae. aegypti,* resulting in a greater proportion of infected mosquitoes and higher titres in the salivary glands. Experimental infection of *Ae. aegypti* and *Ae. albopictus* from a range of North and South American countries determined transmission efficiency of 83% and 97% respectively (102). In addition, *Ae. albopictus* is invasive and widespread, particularly in more temperate regions where *Ae. aegypti* is not established (103).

Despite species-specific differences, *Ae. aegypti* and *Ae. albopictus* share several important characteristics; they are both aggressive, diurnal species with desiccation-resistant eggs, enabling survival in urban environments (104) (Fig 1.10). One particularly devastating similarity is the ability of *Ae. aegypti* and *Ae. albopictus* to support productive co-infection with CHIKV and DENV. Oral inoculation of each species resulted in infectious CHIKV and DENV particles in the saliva which could be transmitted by a single bite, with no significant increase in mortality for the mosquito (105,106). Triple experimental infection with CHIKV, DENV and ZIKV was supported by *Ae. aegypti* and serological studies indicated the presence of triple infections in patients in South America (106,107).

Prior to recent epidemics, *Ae. aegypti* was considered the primary vector of CHIKV and *Ae. albopictus* a secondary vector. During the La Réunion epidemic, adaptive mutations in the CHIKV

envelope proteins E1 and E2 significantly increased infectivity for *Ae. albopictus.* A single E1-A226V mutation increased CHIKV vector infectivity, dissemination, and transmissibility to mice (45). Epistatic mutations E2-G60D, E2-L210Q and E2-I211T increased infectivity for midgut epithelium in the context of the E1-A226V mutation (108,109). The emergence of more severe CHIKV pathogenesis, including the first reports of fatal infections, coincided with CHIKV adaptation to *Ae. albopictus* (45,75). However, transmission by *Ae. albopictus* during this epidemic involved a naïve population. Thus, it is difficult to disentangle the relative importance of the novel vector of ECSA CHIKV and the immunologically vulnerable human population.

Aedes aegypti

- Urban
- Tropical and subtropical climates
- Feeds on humans
- Multiple bloodmeals per gonotrophic cycle
- Artificial water sources for egg-laying
- Moderately susceptible to CHIKV

Aedes albopictus

- Urban and rural
- Tropical and temperate climates
- Feeds opportunistically
- Single bloodmeal per gonotrophic cycle
- Natural and artificial water sources for egg-laying
- Moderately to highly susceptible to CHIKV

Figure 1.10: CHIKV vectors (redrawn from (19)). Typical behaviour and ecological features of the two CHIKV vectors *Ae. aegypti* (left) and *Ae. albopictus* (right). Scales below indicate the relative contribution of behavior and susceptibility in determining level of vector competence for each species. Photographs by James Gathany, CDC Public Health Image Library.

1.1.5 Treatments, vaccines and control

1.1.5.1 Vaccines

Currently, there are no licensed vaccines or anti-viral therapeutic agents for prevention or treatment of CHIKV infection (110). The first live vaccine candidate, developed by 18 plaque-to-plaque passages in human embryonic lung cells, was attenuated during Phase 1 trials (111,112). Unfortunately, transient arthralgia was reported by 8.5% of recipients during Phase 2, resulting in abandonment over safety concerns despite 98% seroconversion (113). Later studies determined that attenuation of the passaged strain was due to two point mutations in the E2 envelope protein, which reverted to the wild-type sequence during infection of mice (114).

Several vaccine strategies have been evaluated including live-attenuated vaccines generated by deletions (115), virus-like particles (116), DNA (117) and subunit vaccines (118). Two approaches in particular have utilised host alternation of CHIKV. Firstly, an Eilat virus/CHIKV chimeric alphavirus (EILV-CHIKV) was designed, composed of the non-structural genes of the insect-specific EILV and the structural genes of CHIKV. EILV-CHIKV can be generated easily in mosquito cells but is attenuated in mice, inducing protective levels of neutralising antibodies within 4 days post-infection (119). Secondly, a CHIKV strain with a picornavirus internal ribosome entry site (IRES) in place of the sub-genomic promoter cannot replicate in mosquitoes (120) (sub-genomic promoter described in 1.3.2.2.1). Furthermore, in mice and cynomolgus macaques, CHIKV-IRESv1 structural protein expression is reduced, attenuating viral replication and protecting against challenge with all CHIKV lineages (121,122). One particularly promising live-attenuated vaccine candidate is in preparation for Phase 3 trials, combining the Schwartz measles virus (MV) vaccine strain with the structural genes of CHIKV to form the recombinant MV-CHIKV. During Phase 2 trials of the MV-CHIKV vaccine in 263 recipients, a single immunisation induced seroconversion in 50-93% of recipients, while a second immunisation within six months produced high titres of neutralising antibodies in 86-100% with minor adverse effects (123).

1.1.5.2 Anti-viral therapeutics

In the absence of specific anti-viral compounds, non-specific antipyretics, anti-inflammatory drugs and analgesics are the only available treatments for symptomatic patients. Screens of existing food and drug administration (FDA)-approved compounds in cell culture has revealed several compounds which prevent productive CHIKV infection due to inhibited entry or replication, including flavonoids (124), flavaglines (125), suramin (126) and natural products such as curcumin (127). In particular, known anti-virals have been investigated for activity against CHIKV, including ribavirin (128). A small clinical study in 10 patients suffering with chronic post-CHIKV arthritis confirmed the efficacy of ribavirin treatment; 200mg ribavirin twice per day significantly reduced joint swelling and arthralgia in 80% subjects (129). Combination therapy with interferon (IFN)-α and ribavirin synergistically suppressed CHIKV replication in cell culture compared to monotherapy (130). Another promising approach is the use of monoclonal antibodies (mAbs). Derivatives of the extremely potent human mAb 4N12 were protective against a lethal CHIKV dose when administered up to 3 days post-infection in mice and rhesus macaques (131,132).

1.1.5.3 Vector control

At present, control of CHIKV epidemics relies on personal preventative measures and reduction of mosquito populations. During CHIKV outbreaks, individuals are encouraged to apply topical repellents such as DEET (N,N-diethyl-3-methylbenzamide) and minimise skin exposure (133). 20% concentrated DEET repels *Aedes* spp. for up to 10 hours and insecticide-treated nets have been shown to reduce the incidence of vector-borne diseases (134,135).

As described above, *Ae. aegypti* and *Ae. albopictus* deposit eggs in refuse and water containers. Community-based involvement in reducing the number of accessible breeding sites has proven effective in reducing *Ae. aeypti* populations (136). Potential breeding sites which cannot be removed can be treated with larvicides such as organophosphates and growth regulators (137,138). During emergencies, adult mosquito populations are reduced by space and residual insecticide spraying, along with mass trapping in autocidal gravid ovitraps (AGO) (139,140). One study of integrated vector control in Puerto Rico combined source reduction, larvaciding and mass trapping, demonstrating ~85-95% reduction in *Ae. aegypti* around target houses (140). A recent study in Puerto Rico demonstrated that AGO traps reduced rates of CHIKV infection from 70% to 20%, as

determined by seroconversion, suggesting that vector control can reduce incidence of CHIKV-related disease (141). Despite their efficacy, current control methods are expensive, labour-intensive and require high levels of compliance from individuals, communities and governmental agencies. Additionally, resistance to insecticides has been documented in *Ae. aegypti* (142).

Controlled testing of a unique approach to vector control is underway, involving the release of male mosquitoes which are sterile or cannot produce viable offspring with wild-type field females. Release of irradiated *Ae. aeypti* males has been shown to reduce vector populations, but irradiation lowers their mating competitiveness (143). Infection of male mosquitoes with the maternally inherited endosymbiont *Wolbachia* does not reduce competitiveness and results in sterile matings with females carrying an incompatible strain of *Wolbachia,* due to cytoplasmic incompatibility (144). However, any anomaly during rearing which led to the release of lab-infected females would overcome the engineered cytoplasmic incompatibility. A combination of triple-strain *Wolbachia* infection and low-level irradiation of males, to prevent fertility in any accidentally released females, has proven effective in eliminating *Ae. albopictus* populations in a field trial in China (145). An alternative approach was taken by Oxitec Ltd, whereby transgenic males with a lethal tetracycline-repressed gene were reared in the presence of tetracycline prior to release. Male OX513A mosquitos successfully reduced *Ae. aegypti* populations by 95% in field trials in Jacobina, Brazil (146). However, portions of the transgenic OX513A genome were detected in the *Ae. aegypti* population 12-24 months later, indicating that lethality of offspring was not complete (147).

1.2 CHIKV

1.2.1 Virion structure

CHIKV is a small, enveloped virus of ~70 nm in diameter (148). The icosahedral nucleocapsid is formed from 240 capsomeres with *T*=4 symmetry, encapsidating a single-stranded, positive-sense RNA genome (Fig 1.11A). The virion surface consists of a host-derived lipid bilayer studded with 80 trimeric glycoprotein spikes (Fig 1.11B). The structure of the spikes, each a trimer of E1-E2 heterodimers, was determined by X-ray crystallography (149). The nucleocapsid, formed from pentameric and hexameric capsomeres, tightly associates with the lipid bilayer via direct interactions of capsomeres with the trans-membrane helix of the viral envelope protein E2 (Fig 1.11C). Cryo-electron microscopy studies of mature and immature CHIKV virus-like particles confirm that E3 is cleaved from released virions during maturation (150–152). The structural proteins are discussed in more detail in 1.2.4.

1.2.2 Genome organisation

The CHIKV genome is a single-stranded, positive-sense RNA molecule roughly 11.8 kb in length (153). The genome comprises two open reading frames (ORFs) flanked by 5′ and 3′ untranslated regions (UTRs) and separated by a non-coding intergenic region (IGR) (Fig 1.12A). The 5′ UTR is 76 nt in length and contains a 5′ type-0 N^7-methylguanosine cap for initiation of cap-dependent translation. The 3′ UTR varies in length between ~500 and ~900 nt and includes a 3′ polyadenylate tail. ORF-1 encodes the non-structural proteins nsP1–4, which form distinct modules of the viral replicase complex that is responsible for CHIKV RNA synthesis. ORF1 is translated directly from the genomic 49S RNA as a polyprotein. An opal stop codon between *nsp3* and *nsp4* regulates expression of nsP4, the RNA-dependent RNA polymerase (RdRp) (43,154). ORF2 encodes six proteins including three major structural proteins (C, E1 and E2). In contrast to ORF1, ORF2 is translated from a sub-genomic 26S RNA transcribed from a sub-genomic promoter in the negative strand. Expression of the 26S RNA depends on successful genome replication, allowing temporal control of structural gene expression (12,155) (Fig 1.12B). In addition to protein-coding sequences, the CHIKV genome harbours important RNA elements within coding and non-coding regions (described in detail in 1.3.3).

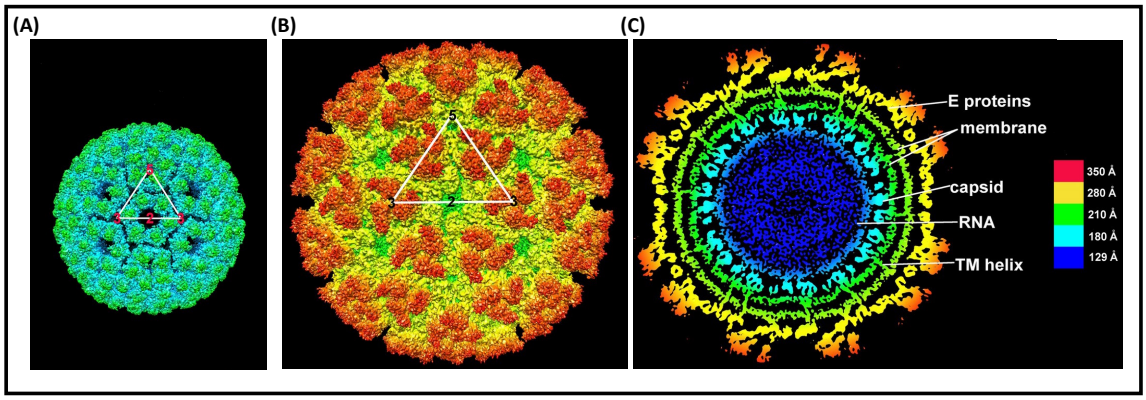


Figure 1.11: Cryo-electron microscopy (cryo-EM) structure of the mature CHIKV virion at ~5.3 Å (adapted from (151)). Surface-shaded figure of **(A)** nucleocapsid and **(B)** ectodomain. White triangles indicate one icosahedral asymmetric unit. (**C**) Cross-section of a virion showing density above 1.5 σ. Each figure is coloured according to the radial distance from the centre of the virus, as shown by the key (far right). The RNA genome is shown in blue, the capsid and membrane in turquoise and green, and the envelope (E) proteins in yellow and red. The trans-membrane (TM) helix of E2 can be seen associating with the nucleocapsid.

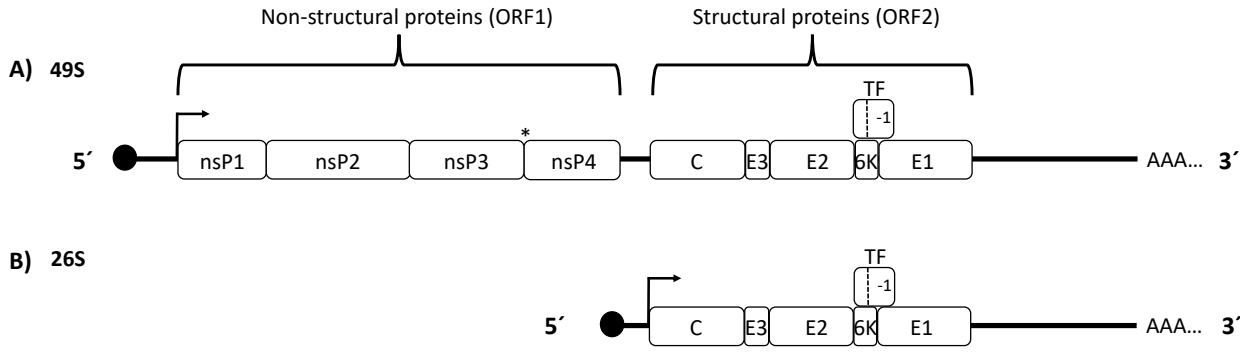


Figure 1.12: CHIKV genome organisation (redrawn from (156)). (A) Schematic representation of the CHIKV genome (49S) displaying two open reading frames (ORFs) encoding the non-structural and structural proteins. The methylguanosine cap is denoted by a black circle and the polyadenosine tail by AAA…. The position of an opal termination codon between *nsp3* and *nsp4* in ORF1 is denoted (*). **(B)** Schematic representation of the sub-genomic 26S RNA encoding the structural proteins. The AUG start codon of genomic 49S RNA and 26S sub-genomic RNA are denoted by black arrows.

1.2.3 Replication cycle

The CHIKV envelope protein E2 binds to the cell adhesion molecule matrix remodeling-associated 8 (Mxra8) at the plasma membrane, in addition to attachment factors such as glycosaminoglycans (157–160) (Fig 1.13). After E2-mediated binding to surface receptors, the CHIKV virion is internalised via an Eps15-dependent, clathrin-independent mechanism and delivered to early endosomes (161). In mosquito cells, early ($Rab5^+$) and late ($Rab7^+$) endosomes are both required for CHIKV entry (162). Endosomal acidification protonates histidine residues on E2, loosening interactions between E1/E2 and exposing the E1 fusogenic peptide (149,163). Intercalation of the fusogenic peptide with the endosomal membrane causes formation of E1 homotrimers. Multiple E1 homotrimers associate to form a fusion pore, releasing the nucleocapsid into the cytosol within 10 seconds of exposure to acidic pH (164,165).

Upon entry into the cytosol, the nucleocapsid is bound by the 60S ribosomal subunit, inducing uncoating within 1-2 minutes. Each 60S subunit irreversibly sequesters 3-6 capsid monomers, resulting in rapid disassembly of the nucleocapsid and releasing the viral genome for translation (166,167). Several capsid monomers are thought to remain associated with the viral genome, as specific capsid-binding sites within the viral RNA are necessary for efficient early translation. The mechanisms behind this enhancement are currently not well understood (168). The CHIKV genome undergoes canonical cap-mediated translation of ORF1 to produce the non-structural polyprotein variants, P123 and P1234. An opal stop codon between *nsp3* and *nsp4* regulates the expression of P1234 via termination codon readthrough, resulting in ~10% frequency of expression of the RNA-dependent RNA polymerase (RdRp) nsP4 (169). Efficient translation of the viral genome in mosquito cells has an additional requirement for a triple tandem sequence motif in the 3′ UTR (170). Nascent P1234 is anchored to proximal endosomal membranes by nsP1 and efficiently cleaved by nsP2 *in cis* to produce a complex of P123 and nsP4 (171,172). Cleaved nsP4 is highly unstable in the cytosol and is stabilised by inclusion in the replication complex (173). nsP3 enzymatically activates the P123/nsP4 complex to form a functional replicase which is competent for negative strand synthesis (174). The initiation of negative strand synthesis is concurrent with the formation of membrane invaginations termed spherules and the recruitment of cellular factors (175–177). In mammalian cells, the P123/nsP4 complex traffics to the plasma membrane for spherule formation. In mosquito cells, live cell imaging, fluorescent reporter viruses

and transmission electron microscopy (TEM) analysis demonstrate that replication occurs in spherules exclusively on intracellular cytopathic vacuoles (178,179). The key differences between CHIKV replication in mammalian and insect cells are highlighted below; the most important differences involve sites of spherule formation, specialization of cytopathic vacuoles and the extent of internal budding (Fig 1.14).

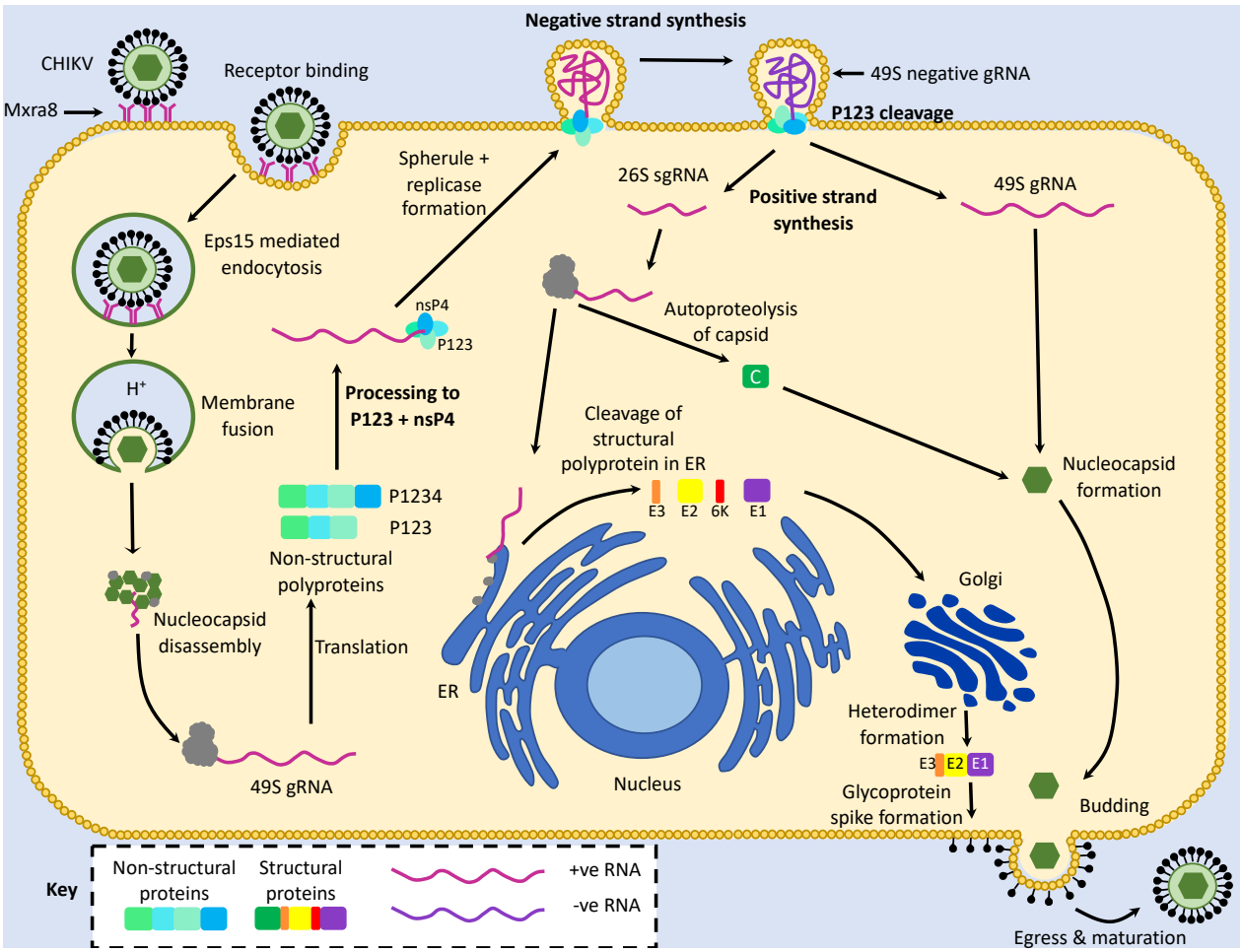


Figure 1.13: CHIKV replication cycle (redrawn from (180)**).** CHIKV binds to cellular receptors (pink Y shapes) via glycoprotein spikes (black) and is endocytosed. Decreasing pH (H^+) in the endosome (dark green circle) results in membrane fusion and release of the nucleocapsid into the cytosol. Capsid (C) monomers (green) associate with ribosomal subunits (grey) and release the viral genome (49S gRNA) (pink line). Translation of the viral genome yields the non-structural polyproteins P123 and P1234 which are processed into P123 and nsP4. P123 and nsP4 form the negative strand replicase which binds to the 3′ of the 49S gRNA before internalisation into invaginations at the plasma membrane where negative strand synthesis occurs. When the cytosolic concentration of P123 is high enough to support an efficient reaction *in trans*, P123 cleaved into mature non-structural proteins: nsP1-4. Cleavage of P123 rearranges the replicase to favour binding to the negative strand (purple line), resulting in positive strand synthesis. The 49S gRNA and a 26S sub-genomic (sg) RNA which encodes the structural polyprotein are released into the cytosol. Translation of the sgRNA releases the C protein by autoproteolysis, followed by co-

translational translocation of precursor of envelope protein 2 (pE2), 6K, and E1 into the endoplasmic reticulum (ER). The C proteins and 49S gRNA accumulate in the cytoplasm and assemble into nucleocapsids (green hexagons). The envelope proteins E1-3 are modified during transit through the secretory pathway and form E1-E2 heterodimers with E2-associated E3. Glycoprotein spikes accumulate on the plasma membrane and E2 cytoplasmic domains interact with the nucleocapsid. CHIKV virions bud from the membrane and undergo maturation soon after release, shedding E3.

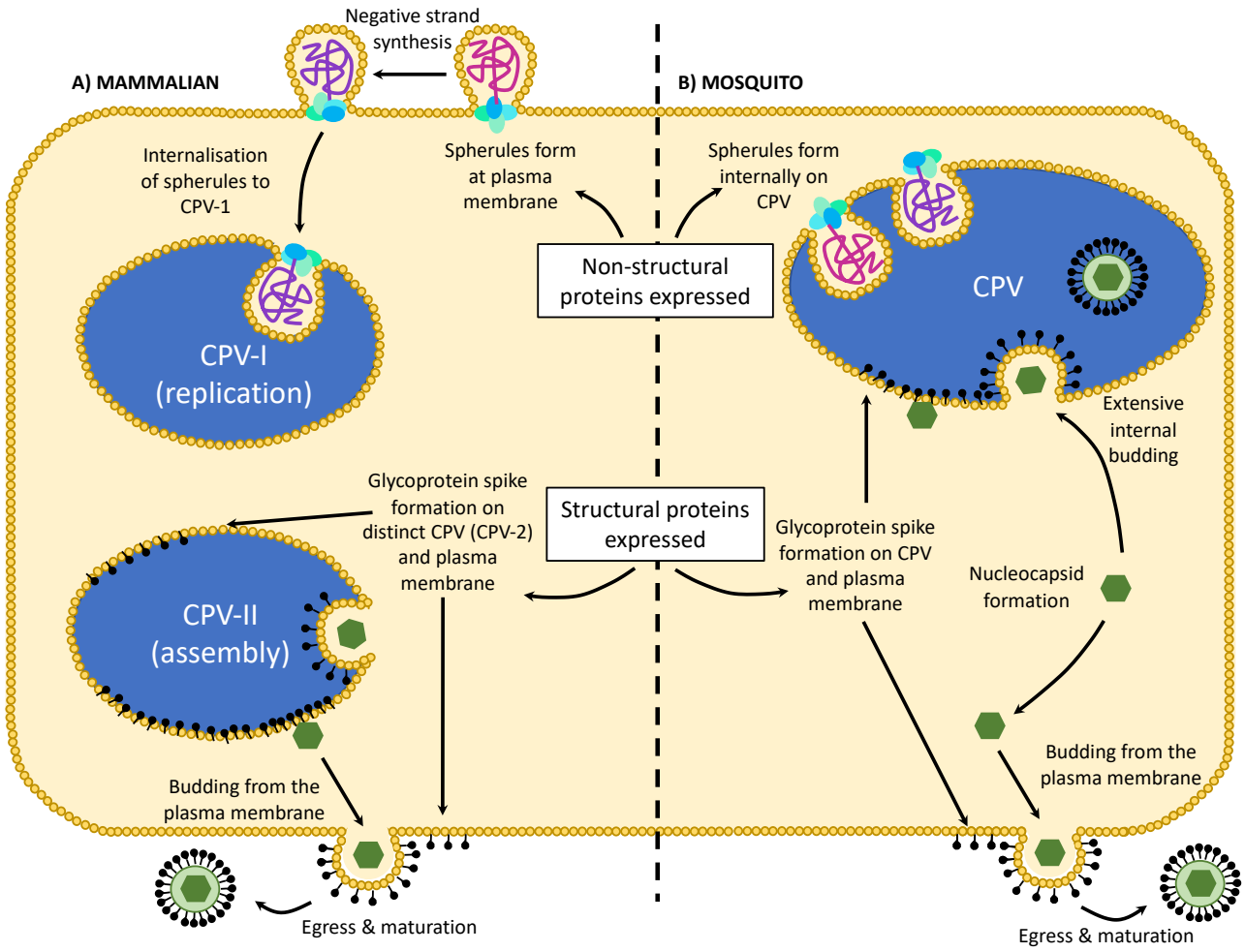


Figure 1.14: Differences in CHIKV replication in mammalian and mosquito cells. Expression of CHIKV non-structural proteins in mammalian cells (A) results in spherule formation on the plasma membrane followed by internalization of spherules onto the surface of specialized cytopathic vacuoles (dark blue) (CPV-I). Expression of structural proteins results in accumulation of viral glycoproteins on the plasma membrane and in arrays on the surface of a second type of cytopathic vacuole (CPV-II). Expression of CHIKV non-structural proteins in *Aedes* mosquito cells (B) results in spherule formation directly on the surface of cytopathic vacuoles (dark blue) (CPV). Expression of structural proteins results in accumulation of viral glycoproteins on the plasma membrane and in arrays on the surface of the same type of cytopathic vacuole. Extensive budding of virions into the lumen of the CPV is observed in mosquito cells, a process which is much more limited in mammalian CPV-II. The figure uses the same key as Fig1.13, using data drawn from (179).

Negative strand synthesis is regulated by interactions between the N-terminal domain of the RdRp nsP4 and each of the components of P123. Mutations in the N-terminus of nsP4 which diminished negative strand synthesis were suppressed by mutations in nsP1-3 (181). Similarly, mutations in nsP2 and nsP3 can reduce or abolish negative strand synthesis (182,183). During replication, nsP1 catalyses the addition of a cap to nascent RNA, nsP2 acts a helicase to unwind RNA secondary structures and nsP3 plays an ill-defined role, potentially via zinc-binding or adenosine monophosphate (AMP) binding. P123-associated nsP4 is restricted to negative strand synthesis. Expression of uncleavable P123 results in negative strand synthesis only, without subsequent synthesis of genomic and sub-genomic positive RNAs. Extraction of nsP4 from uncleavable P123 complexes by fractionation demonstrated that nsP4 is capable of synthesising genomic RNA of both polarities in isolation and thus P123 locks nsP4 into a restricted conformation. In this context, nsP4 binds efficiently to the 3′ negative strand promoter in the positive genomic RNA, as demonstrated by electrophoretic mobility shift assay (EMSA) and RNA cross-linking assays, and produces negative strand genomic RNA (184,185). The negative strand intermediate does not contain a 3′ polyA tail but contains a 5′ polyU transcribed from the 3′ polyA of the positive-sense genomic RNA. The addition of a 5′ methylguanylyl cap to the negative-sense intermediate has not been investigated (186).

Sequential cleavage of P123 to nsP1/2/3 regulates the switch from negative to positive strand synthesis (187) (Fig 1.15). Mutational analyses demonstrated that nsP2 must cleave junctions 1/2 and 3/4 sequentially at conserved residues (188,189). *In vitro* analysis using uncleavable polyproteins and deletion mutants revealed that junction 1/2 was cleaved in *cis,* within 20-30 minutes, exposing an activator sequence which promotes cleavage of 2/3. Junction 2/3 is subsequently cleaved in *trans,* dependent on prior release of nsP1 (172,190). Thus, negative strand synthesis is time-limited by spontaneous cleavage of nsP1 *in cis* and sensitive to increasing concentrations of nsP2 capable of cleavage *in trans.* The transition from the P123 to nsP1/2/3 replicase is irreversible and permanently prevents negative strand synthesis, indicating extensive rearrangement within the complex (191). The structure of uncleaved P23 displays a highly charged interface, suggesting an electrostatic interaction post-cleavage (192). nsP1/2/3-associated nsP4 is restricted to positive-strand synthesis and recognises the genomic and sub-genomic promoters in the negative strand via two distinct RNA-binding sites on nsP4 (184,193). Isolated nsP4 is

incapable of synthesising sub-genomic RNA in isolation and requires cleaved nsP1/2/3 (185). In particular, mutations in nsP2 often result in defects in sub-genomic RNA synthesis, indicating that it plays a key role as a transcription factor for nsP4 (194,195). Similar phenotypes have been observed for nsP3 mutants, although genomic RNA synthesis is also affected (196,197). The nsP1/2/3/4 complex synthesises full length positive-sense RNA and sub-genomic 26S RNA encoding the structural proteins of ORF2. As replication progresses, cleaved nsP1 and nsP3 direct internalisation of spherules from the plasma membrane onto the surface of cytopathic vacuoles similar to those utilised in mosquito cells (198).

As the concentration of non-structural polyproteins in the cytosol increases, rapid cleavage occurs *in trans* which releases nsP1, nsP2 and nsP3 for non-replicative functions. nsP1 remains membrane-associated and induces formation of short extensions from the membrane involved in pathogenesis. Release of nsP2 is essential for nuclear localisation, transcriptional and translation shut-off and interference with innate immunity (199). nsP3 localises to cytoplasmic foci where it sequesters stress granule components such as GTPase Activating Protein (SH3 Domain) Binding Protein (G3BP) proteins and the mosquito homologue Rasputin (199,200).

Late during infection, viral genome replication initiates translational shutoff via activation of protein kinase R (PKR) and subsequent phosphorylation of eukaryotic translation initiation factor 2α (eIF2α) (201). Translation of the 49S genomic RNA is inhibited but translation of the sub-genomic 26S messenger RNA (mRNA) continues efficiently via a cap-independent mechanism, favouring production of structural proteins (202,203). During translation of the 26S RNA, autoproteolysis of the nascent capsid protein exposes a signal sequence in E3, directing the ribosome to the ER membrane. Signal peptidase cleaves 6K from pE2 and E1 upon translocation into the lumen (204). The proteins are trafficked through the secretory pathway where they undergo a series of complex folding events and acquire post-translational modifications such as palmitoylation and glycosylation (205). Furin cleavage of pE2 occurs in a post-Golgi compartment to form non-covalently associated E2 and E3 (206). E3 stabilises the interaction of E2 with the fusion loop of E1, preventing premature exposure in acidic pH (152).

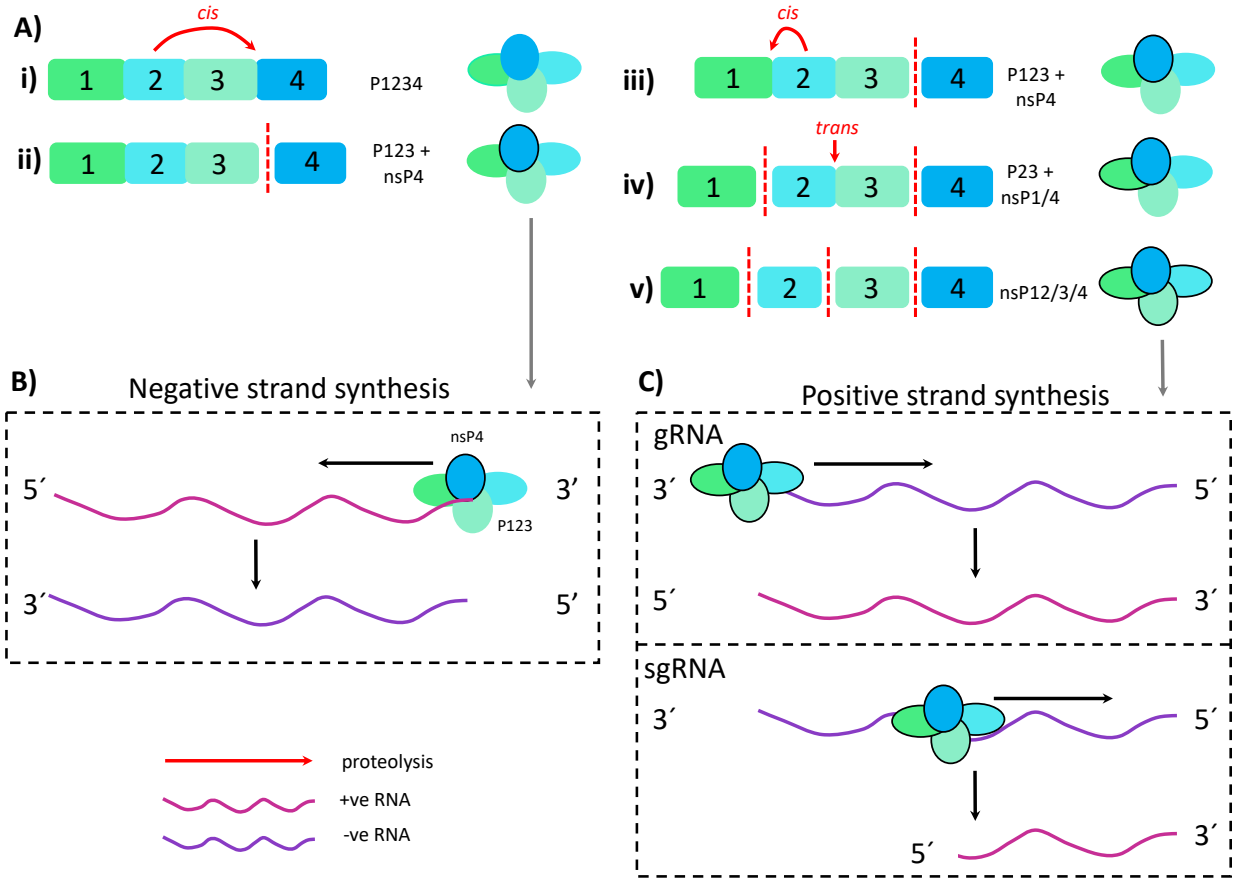


Figure 1.15: Regulation of CHIKV genome replication. (A) Processing of the replicase complex over the viral replication cycle and its relation to **(B)** negative strand synthesis and **(C)** positive strand synthesis. **(i)** P1234 is initially cleaved into P123 and nsP4 by nsp2 *in cis* (red arrow and dashed line), forming **(ii)** a negative strand-competent replicase. The P123/nsP4 complex binds (grey arrow) to the 3′of positive-sense 49S genomic RNA (gRNA) (pink line) to produce a full length negative-sense strand (purple line). **(iii)** Within ~30 minutes of translation, P123 is cleaved into nsP1 and P23 by nsp2 *in cis*. **(iv)** A final cleavage between nsP2 and nsP3 occurs *in trans* and depends upon increasing non-structural protein expression in the cytosol. **(v)** Fully cleaved, mature nsP1-4 forms the positive strand-competent replicase which binds (grey arrow) to the of the negative strand at two distinct sites: the 3′ genomic promoter and intergenic sub-genomic promoter. Recognition of the intergenic promoter requires nsP2. Positive strand synthesis produces genomic and sub-genomic RNA species (pink lines). Black arrows indicate direction of RNA synthesis. Black outline around nsPs indicates maturation and cleavage from the polyprotein. Data drawn from: (172,184,185,190).

Interaction between the capsid protein and an SFV-like packaging signal in *nsp2* nucleates assembly of the CHIKV nucleocapsid, followed by rapid dimerisation of bound capsid via N-terminal α-helices (207,208). A hydrophobic pocket of the capsid protein binds the phosphorylated cytoplasmic tail of E2 and mediates acquisition of the lipid envelope (209,210). Despite the ability of the nucleocapsid to assemble *in vitro* in the absence of E2, disruption of E2-C interactions resulted in aberrant capsid formation and non-budding phenotypes *in vivo* (211,212). An early interaction between C and E2 may direct formation of the nucleocapsid (NC) and tight interactions with E2 necessary for budding (213). 6K and TF proteins also influence virion stability, association of C-E2 and budding (214,215).

Arrays of glycoprotein spikes accumulate on the plasma membrane and on the internal side of cytopathic vacuoles (216). In mammalian cells, virions bud from the plasma membrane. However, in mosquito cells, budding also occurs into the lumen of cytopathic vacuoles. Intra-luminal budding is associated with cell-to-cell spread through the secretory pathway, evading immune recognition and aiding dissemination (178). Persistently infected mosquito cells appear to increase lysosomal degradation of vacuole-associated spherules and intra-luminal virions (179).

1.2.4 Structural proteins

The structures of the CHIKV capsid and envelope proteins have been recently resolved by X-ray crystallography (149,210). There have been no structural studies of CHIKV 6K or TF proteins and very few direct studies of CHIKV structural protein functions. Parallels must be drawn with related alphaviruses which have been studied in detail, particularly SFV and SINV. CHIKV structural proteins contain a high degree of protein sequence homology with known functional regions of SFV. Cryo-EM reconstruction of the CHIKV virion demonstrates high concordance between the structures of CHIKV, SFV, SINV and VEEV (151,217–219).

1.2.4.1 Capsid

The CHIKV capsid protein (CP) is a multi-functional 29 kDa protein which forms pentameric and hexameric capsomeres during nucleocapsid assembly. The CP consists of an N-terminal RNA-binding domain and a C-terminal serine protease domain (Fig 1.16) (220). The N-terminal domain

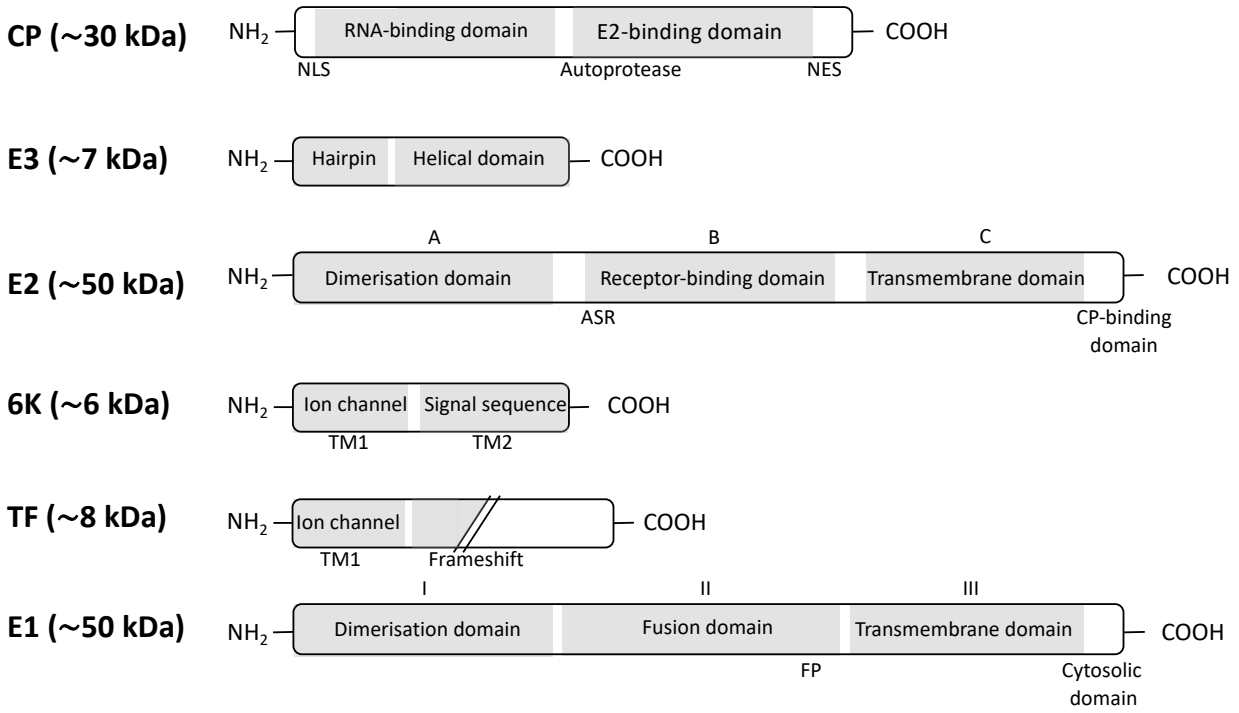


Figure 1.16: Structural protein domain organisation

Domain organisation of the alphaviral structural proteins: capsid (CP), envelope proteins 1-3 (E1-3), 6K and trans-frame (TF) protein. Recognised domains, in terms of either structural or genetic evidence, are indicated as grey boxes. Specific features, as indicated on their respective proteins, are briefly listed below. CP: NLS and NES denote nuclear localization and nuclear export sequences. E2: immunoglobulin domains A,B and C are labelled; ASR indicates the acid sensitive region of domain B. 6K: TM1 and TM2 indicate transmembrane domains 1 and 2; TM2 contains the signal sequence for co-translational translocation of E2. TF: TM1 indicates transmembrane domain 1; diagonal lines depict the +1 frameshift during translation, leading to a unique C-terminus for TF compared to 6K. E1: β-barrel domains I-III are labelled; FP denotes the fusion peptide.

is highly positively charged and intrinsically disordered (221,222). Deletion of the N-terminal domain reduces the specificity of genomic RNA encapsidation in VEEV and enables promiscuous capsomere dimerisation between the heterologous alphaviruses SINV, WEEV and RRV (223,224). The C-terminal domain of the CP is autoproteolytic, acting in *cis* to release itself from the N-terminus of the nascent structural polyprotein during translation of the 26S RNA (225,226). The crystal structure of the CHIKV C-terminal protease domain of CP has been resolved at 2.2 Å, revealing a chymotrypsin-like protease and a hydrophobic pocket capable of interacting with the cytoplasmic domain of E2 (210). In addition, the CHIKV CP contains sequences for nuclear import and export at the N- and C- termini respectively (227). Nuclear-cytoplasmic shuttling of the CHIKV CP relies on karyopherin-α-mediated import and chromosomal maintenance 1 protein (CRM1)-mediated export, but the function of nuclear localisation is unknown.

1.2.4.2 Envelope (E1-3) proteins

The mature CHIKV virion contains 240 heterodimers of envelope glycoproteins E1 and E2, arranged in 80 trimeric spikes (149,228). Cryo-EM reconstruction suggests that monomers of E1 form the base of each trimeric spike, creating an icosahedral lattice across the virion surface (229,230). Crystallographic studies of the E1-E2 heterodimer reveal that the extended stem and apex of each glycoprotein spike is composed of E2 (149). E2 interacts with host receptors to initiate endocytosis and E1 contains the hydrophobic peptide necessary for fusion of viral and endosomal membranes (230,231). E1 and E2 are both ~430 aa in length, glycosylated and palmitoylated (232). E1 contains three β-barrel domains; domain II contains the fusion peptide and domain III terminates in a transmembrane helix which extends through the lipid bilayer by 6 aa (Fig 1.16) (218). E2 is composed entirely of β sheet, with three immunoglobulin domains; domain B contains the highly-exposed apical structure suitable for receptor binding and domain C contains the transmembrane (TM) helix (149). The E2 TM helix extends 33 aa into the inner cavity of the virion, interacting with the hydrophobic pocket of the capsid protein. Each E2 cytosplasmic tail contacts a capsomere, forming 240 1:1 interactions which enable tight pairing of the envelope and the nucleocapsid (210). Dissociation of the E1-E2 heterodimers is induced by low pH, resulting in homotrimers of E1 with exposed fusogenic peptides (233,234).

Cryo-EM reconstruction of immature CHIKV virions reveals envelope protein E3 in complex with E2 (152). The 64 aa E3 protein contains an N-terminal β-hairpin and three α-helices (149). E2 and E3 are translated as the precursor polyprotein pE2, sometimes called p62. The precursor is necessary for correct folding of the pE2-E1 heterodimer during co-translational insertion into the endoplasmic reticulum membrane (235). The E3-E2 junction of pE2 is an exposed loop which is cleaved by furin during traffic through the secretory pathway (149). E3 is lost during maturation of the CHIKV virion post-egress and can be found in extracellular fluid but not purified mature virus (150,151).

1.2.4.3 Protein 6K/TF protein

6K is a small, hydrophobic, cysteine-rich protein of 61 aa amino acids in length, consisting of two transmembrane domains joined by an 8 aa cytoplasmic linker (Fig 1.16) (236). The first domain is thought to act as an ion channel via oligomerisation within the membrane. Insertion of bacterially expressed 6K proteins into planar lipid bilayers has demonstrated cation-selective ion channel activity of 6K, indicating that it may function as a viroporin (155). The second transmembrane domain of 6K is the signal sequence for E1 (236). Deletion or mutation of 6K is associated with decreased virion stability, release of multi-cored virions and accumulation of nucleocapsids at the plasma membrane with impaired budding (237–239).

A second accessory protein, the 8 kDa transframe (TF) protein, is encoded within the *6K* gene. The TF protein represents the N-terminal 70-80% of 6K with a C-terminal elongation of ~30 codons in the -1 reading frame. Frameshifting frequency is ~2% (240). The TF protein was identified in purified CHIKV virions using mass spectrometry. Phenotypic studies in SINV indicate functionality of TF as an ion channel and in virion budding in a similar manner to 6K (215). This is not surprising given that TF retains ~80% of the 6K primary sequence, but the unique functions of TF remain to be established.

1.2.5 Non-structural proteins

1.2.5.1 Non-structural protein 1 (nsP1)

Non-structural protein 1 (nsP1) is a 58 kDa protein with two primary functions (Fig 1.17). The N-terminal domain of nsP1 directs capping of nascent positive-sense viral RNA, which is essential for viral mRNA translation and successful replication (241) (Fig 1.18A). The guanine-7-methyltransferase (MTase) and guanylyltransferase (GTase) activities of nsP1 were first identified by biochemical assays (242,243). Subsequent studies using [α-^{32}P]GTP revealed cross-linking of nsP1 to the methyl donor S-adenosylmethionine and m^7Gp (244,245). Reverse genetic analysis revealed that successful MTase activity is a prerequisite for GTase activity, thus methylation of GTP precedes guanosine monophosphate (GMP)-transfer to the disphosphate mRNA (245). This is in contrast to canonical capping mechanisms where methylation occurs after transfer of the guanylyl moiety to diphosphate RNA (246). The MTase motif transfers a methyl group from S-adenosylmethionine to GTP, producing m^7Gppp. Subsequently, the GTase motif covalently binds m^7Gppp, releasing pyrophosphate (PPi) (244). The m^7Gp moiety is then transferred from the catalytic GTase residue to the 5′ pp-RNA generated by nsP2 triphosphatase activity (247). Reverse genetic analysis revealed that both MTase and GTase activities are necessary for successful SINV replication (241).

nsP1 is also responsible for anchoring the replication complex to cellular membranes, a process which is essential for productive SFV infection (248) (Fig 1.18B). The enzymatic activity of SFV nsP1 depends on association with anionic phospholipids such as phosphatidylserine and can be disrupted by the addition of detergents (171). nsP1 membrane-binding maps to a 20 aa sequence which is highly conserved across the *Alphavirus* genus (171). Nuclear magnetic resonance (NMR) studies of this peptide revealed a structural transition in the presence of anionic phospholipids, from a random coil to an amphipathic alpha-helix (248). Green fluorescent protein (GFP)-SFV localisation studies demonstrated that two tandem copies of the amphipathic peptide sequence are necessary for membrane-association (249). SFV nsP1 localises to the plasma membrane even in mutants deficient in spherule formation, suggesting that nsP1 anchoring is not sufficient for spherule biogenesis (250).

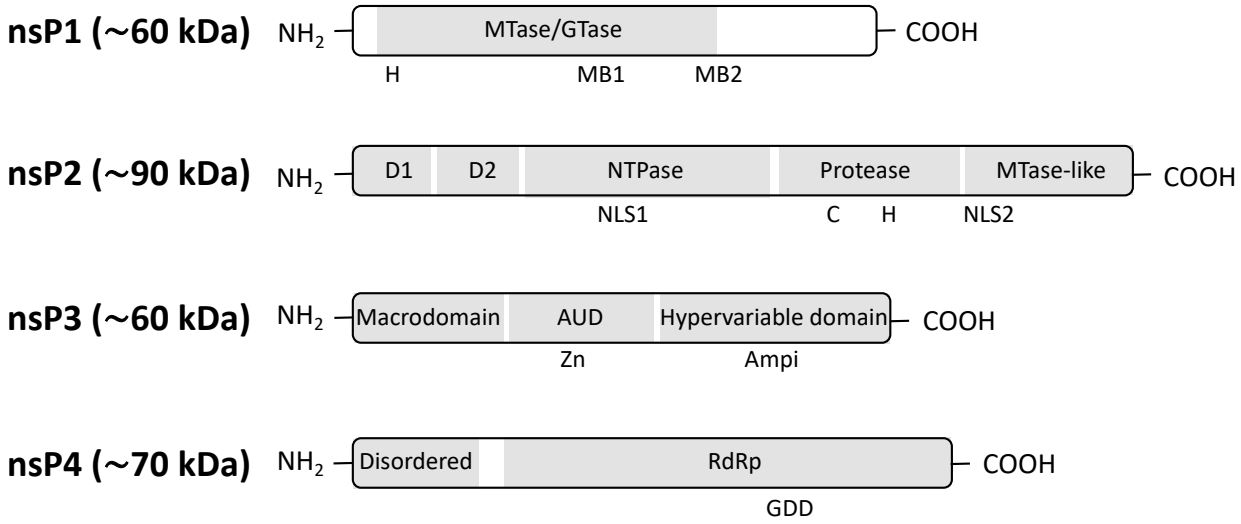


Figure 1.17: Non-structural protein domain organisation (adapted from (251)**)**

Domain organisation of the alphaviral non-structural proteins nsP1-4. Recognised domains, in terms of either structural or genetic evidence, are indicated as grey boxes. Specific features, as indicated on their respective proteins, are briefly listed below. nsP1: H designates the catalytic histidine residue of the GTase domain; MB1 and MB2 indicate the sites of the membrane-binding amphipathic helix and the palmitoylation site of nsP1, respectively. MTase, methyltransferase; GTase, guanylyltransferase. nsP2: NLS1 and NLS2 indicate nuclear localisation sequences; the catalytic residues of the protease active site are indicated with a C and H, respectively. NTPase, nucleoside triphosphatase. nsP3: the location of the zinc ion coordination site is denoted with Zn; the presence of ampiphysin interaction sites is denoted with Ampi. AUD, alphavirus unique domain. nsP4: the functional GDD catalytic triad of RdRp is shown.

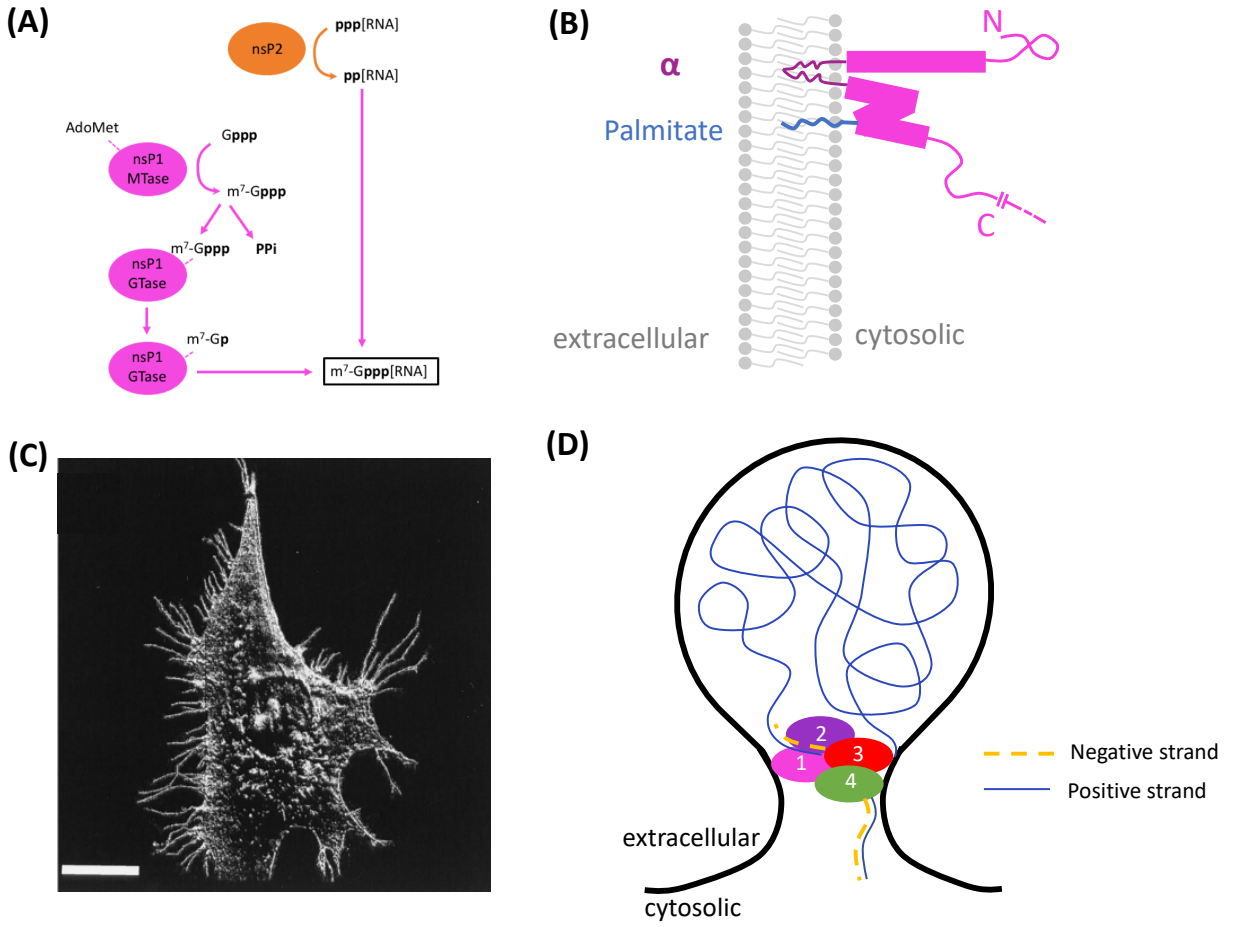


Figure 1.18: Functions of nsP1. (A) The alphaviral mechanism of capping requires nsP1 (pink) and nsP2 (orange). The methyl donor S-adenosylmethionine is labelled AdoMet and phosphate denoted by **p**. Guanine-7-methyltransferase (MTase) and guanylyltransferase (GTase) activities of nsP1 are labelled. Data from: (244,247). **(B)** Membrane-anchoring ability of nsP1 requires two amphipathic membrane-binding helices (purple) labelled α, which intercalate with the lipid bilayer (grey). Palmitate modification is shown in blue. The N- and C- terminals and cytosolic domains of nsP1 (pink) are displayed, with a dotted line representing the rest of the replicase complex. Data from: (249). **C)** Scanning electron micrograph of an SFV-infected baby hamster kidney (BHK-21) cell with short extensions of the membrane (bar represents 10 μm) (252). **(D)** Interaction of nsP1 (in the context of P123) with nsP4 within the spherule (black line), directing negative strand synthesis. nsPs are labelled 1-4. Positive-strand (blue) and nascent negative-strand (yellow, dashed) genomic RNA is displayed. Data from: (241,253,254).

In addition to the well-studied functions above, nsP1 has been implicated two further processes. A palmitoylation site in nsP1 is associated with the induction of short projections from the plasma membrane (255,256) (Fig 1.18C). Real-time live imaging of fluorescent protein-tagged SINV demonstrated the formation of nsP1-positive filopodia in SINV-infected cells – in both mosquito and mammalian cells (179). These filopodia can be induced by expression of nsP1 alone and removal of the palmitoylation site C418-C420 residues of nsP1 in SINV and SFV prevented induction of filopodial extensions (257). Palmitoylation-defective SFV and SINV nsP1 exhibited wild-type enzymatic activity, spherule biogenesis and membrane association (252,258). The function of nsP1-induced filopodia is unknown but inhibition of SFV nsP1 palmitoylation decreased pathogenesis and neurovirulence in mice (258).

The nsP1 palmitoylation site may be the site of nsP1-nsP4 contact, independent of the presence of palmitate. Removal of the 3 aa palmitoylation site resulted in attenuated negative strand synthesis but longer deletions of 7 aa replicate to wild-type levels, suggesting the existence of structural contacts necessary for nsP1-nsP4 association (259). Palmitoylation-defective nsP1 mutants which attenuated replication were associated with loss of nsP1-nsP4 binding and were restored by compensatory mutation in nsP1 which did not restore the palmitoylation site. Thus, palmitoylation is not thought to be necessary for viral replication of SFV (259). Disruption of the 3 aa site resulted in a serious functional defect in replication in mammalian and insect cells, with particularly severe reduction in sgRNA synthesis in insect cells (260). Interaction between nsp1 and nsp4 has been previously documented in genetic analysis of SINV and SFV and found to be related to negative strand synthesis (241,253,254) (Fig 1.18D).

1.2.5.2 Non-structural protein 2 (nsP2)

Nsp2 is a ~90 kDa protein with three essential catalytic functions. The N-terminal domain acts a helicase to unwind secondary RNA structures during replication. SFV nsP2 was capable of unwinding [α-^{32}P]-labelled dsRNA in the presence of Mg^{2+} and NTPs. Mutation of the NTP-binding domain prevented this activity (261). The N-terminal domain also possesses 5′ triphosphatase activity necessary for capping of viral mRNA (Fig 1.18A). Biochemical assays of CHIKV nsP2 using [γ-^{32}P]RNA and [α-^{32}P]NTPs demonstrated Mg^{2+}-dependent 5′ triphosphatase activity and NTPase activity (262).

The C-terminal domain possesses protease activity essential for cleavage of the non-structural polyprotein (263). The C-terminal domain contains a novel α/β cysteine protease fold and an enzymatically inactive MTAse-like subdomain (264,265). Mutation of the nsP2 protease active site abolished polyprotein cleavage and productive replication in both SFV and SINV. Complementation in *trans* with protease-competent nsP2 demonstrated that nsP2 can cleave P123 in *cis* and *trans* (266–268). As discussed in 1.2.3, sequential proteolytic processing of the replicase complex regulates the switch between negative and positive RNA synthesis. The requirement for sequential processing of junction 3/4, 1/2 and 2/3 was demonstrated by mutational analysis in SINV and SFV (173,189). *In vitro* analysis using uncleavable polyproteins and deletion mutants revealed that junction 3/4 was most efficiently cleaved and junction 1/2 was cleaved in *cis* within 20-30 minutes, exposing an activator sequence responsible for promotion of 2/3 cleavage. Junction 2/3 was cleaved in *trans* and was dependent on prior release of nsP1 (190,268)

In addition to catalytic functions, nsP2 acts as a transcription factor for nsP4, aiding recognition of the sub-genomic promoter. Mutants defective in sub-genomic promoter recognition and sub-genomic RNA synthesis map to nsP2 in SINV and SFV (183,194,195). nsP2 also modulates anti-viral host responses via independent processes of transcriptional and translational shut-off (269–271). CHIKV nsP2 induces degradation of a catalytic subunit of RNA polymerase II, resulting in transcriptional shut-off. Mutational analysis demonstrated that this activity depends on the helicase and MTase-like domains but not the protease function of nsP2 (272). In addition, nsP2 induces translational shut-off, contingent on the localisation of ~50% of nsP2 in the nucleoli (273,274). nsP2 translational shut-off depends on the MTase-like subdomain but the mechanisms behind this process are currently not well understood (183).

1.2.5.3 Non-structural protein 3 (nsP3)

nsP3 is a 60 kDa phosphoprotein with three distinct domains (Fig 1.15). The macrodomain of CHIKV nsP3 is an RNA-binding domain capable of recognising and removing adenosine diphosphate-ribose (ADP-ribose) from proteins (275). The crystal structure of the macrodomain in complex with either ADP-ribose or RNA revealed a binding pocket specific for ADP in ADP-ribose and AMP in RNA (276). Mono- and poly-ADP-ribosylation is an indicator of cellular stress;

proteins in stress granules are often ADP-ribosylated (277). The nsP3 macrodomain may downregulate host cell stress responses via hydrolysis of ADP-ribose. The enzymatic activity of nsP3 is also critical for initiation of genome replication, although the mechanism of initiation is unknown. Catalytically inactive CHIKV nsP3 macrodomains cannot form functional replicase complexes and cannot replicate in mammalian or mosquito cells (174,278). Similarly, SINV macrodomain mutants exhibited severely reduced genomic and sub-genomic RNA synthesis, as measured by incorporation of [^{3}H]uridine (182,196). nsP3 may regulate nsP4 during genome replication; mutants of nsP4 with defective negative strand synthesis were rescued by suppressor mutations in nsP3 (181).

The function of the alphavirus unique domain (AUD) of nsP3 has not been well characterised. Crystallographic studies of the SINV AUD identified a zinc-binding domain. Substitution of any of the four cysteine residues which co-ordinate zinc binding was lethal (192). Recent mutational analysis of the CHIKV AUD suggested that the AUD plays a role in genome replication (197). This study demonstrated that mutations in the CHIKV AUD impaired sub-genomic RNA production or abolished viral genome replication in mammalian and mosquito cell lines. One AUD mutant (P247A/V248A) exhibited significantly decreased binding affinity for the sub-genomic promoter and the 3′ termini of positive and negative genomic RNAs *in vitro*.

The hypervariable domain (HVD) is a hyperphosphorylated hub involved in multiple protein interactions, recruiting essential factors such as Four and a half LIM domain protein 1 (FHL1) to the replicase complex (279,280). The HVD binds to the membrane-associated adaptor proteins amphiphysin-1/2 and an abundance of cytoskeletal proteins, most strongly to CD2-associated protein (CD2AP) (175,281,282). The HVD also sequesters the components of stress granules in mammalian (G3BP1 and 3) and insect cells (Rasputin) into cytoplasmic foci (199,283). Several proteins including FHL1, G3BP1/3 and amphiphysin-1/2 have been demonstrated to co-localise with nsP3 in spherules at the plasma membrane and on cytopathic vacuoles, forming interactions that are essential to CHIKV genome replication (175,280,284). Although the mechanism of replication enhancement by HVD-interactors is currently unknown, loss of HVD:protein interactions prevents CHIKV replication in mammals and mosquitoes (175,177,284). Hosts and tissues which do not express homologues of certain HVD-interactors cannot support replication.

Consequently, the HVD is a major determinant of cellular permissivity to CHIKV infection (280,285).

1.2.5.4 Non-structural protein 4 (nsP4)

CHIKV nsP4 is a 70 kDa protein containing the RNA-dependent RNA polymerase (RdRp) GDD motif. Homology modelling and hydrogen-to-deuterium exchange (HDX)-mass spectrometry of the CHIKV nsP4 catalytic core predicted a classical RdRp structure with fingers, thumb and palm sub-domains (286). The C-terminal domain of nsP4 is solely responsible for RNA synthesis in the replicase; addition of purified SINV nsP4 to cell extracts containing uncleaved P123 resulted in *de novo* synthesis of negative strand RNA (287). In addition to *de novo* RNA synthesis, the GDD motif confers terminal adenosyl transferase (TATase) activity. GDD>GAA mutation prevented catalysis of both processes (287,288). SINV nsP4 catalyses the addition of non-templated adenosine residues to the 3′ terminus of positive-sense RNAs to create a polyA tail, dependent on the presence of divalent cations (287). Negative strand synthesis by SINV nsP4 requires the 19 nt 3′ CSE negative strand genomic promoter and a polyA tail of at least 11-12 nt in length (289). Electrophoretic mobility shift assay demonstrated that nsP4 binds directly to the 3′ CSE (185). RNA cross-linking studies have also demonstrated direct binding of nsP4 to the negative strand sub-genomic promoter, dependent on the presence of the other nsPs. Mutational analysis revealed distinct binding sites in nsP4 for the negative strand sub-genomic promoter and genomic promoters, at 329-334 aa and 531-538 respectively (184,193).

The N-terminus of nsP4 interacts with nsP1-3. Mutation of the N-terminus of SINV nsP4 resulted in defective production of sub-genomic and genomic RNA of both polarities. Suppressor mutations were found in nsP1-3, indicating that the N-terminal domain regulates protein-protein interactions within the replicase complex (181). In particular, lethal mutation of the N-terminal tyrosine (Tyr) residue of nsP4 is non-lethal in the context of nsP1 suppressor mutations, suggesting an important nsP1-nsP4 interaction (253,290). Complete deletion of the N-terminus precluded *de novo* negative strand RNA synthesis, even in the presence of P123 (288).

1.3 CHIKV RNA structure

1.3.1 RNA structure

In addition to coding sequences, the genomes of RNA viruses often contain *cis*-acting sequences and structures which act as activators, enhancers and repressors of fundamental processes such as translation, replication and evasion of host innate immunity (291). RNA viruses maximise the coding capacity of the genome by formation of intramolecular secondary and tertiary structures within and between coding regions. The most common RNA structural motifs in viral genomes are stem-loops, also known as hairpins, and pseudoknots (292).

Stem-loops consist of a double-stranded RNA helix formed by base-pairing of adjacent regions on a single RNA molecule, with an intervening single-stranded loop (Fig 1.19A). Unpaired bulges along the stem may also form part of the structure. Pairing within the stem can exhibit canonical Watson-Crick or non-Watson-Crick pairing; G:U pairs are particularly common (293). The structure and sequence of the stem as well as the sequence of the exposed loop are important components of stem-loops which may be recognised by different interactors and via different mechanisms (294). Pseudoknots are formed when the single-stranded terminal loop of a hairpin or stem-loop base-pairs with a complementary single-stranded region elsewhere in the RNA, usually in proximal flanking regions (295) (Fig 1.19B). Pseudoknots are particularly well studied in the context of translational recoding (296,297). Besides local secondary structure formation, long-range base-pairing is common in viral genomes, enabling critical conformational switching during viral replication and providing the striking flexibility characteristic of RNA viral genomes (298,299). Circularisation often depends on long-distance interactions between sequences such as cyclisation sequences which base-pair, or structures such as stem-loops which form kissing-loop interactions (300) (Fig 1.19C).

1.3.2 RNA structure determination

RNA structure can be determined experimentally using spectroscopic techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (301,302). These methods are based on the response of a sample to electromagnetic radiation. Fluorescence methods such as single-molecule spectroscopy can be used to measure dynamic RNA structures via the signal

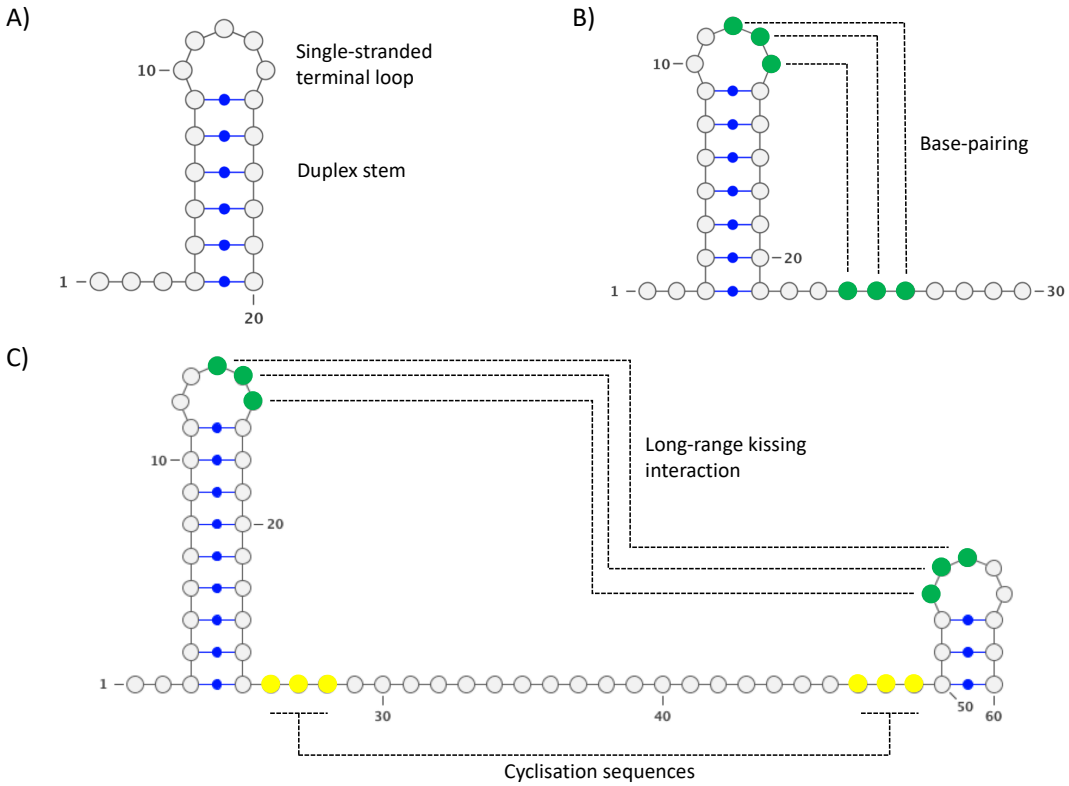


Figure 1.19: Schematic representation of RNA secondary structural elements (adapted from (303)**).** Schematic representation of common RNA secondary structures. White circles represent nucleotides, blue circles represent base-pairing, Watson-Crick or otherwise, between nucleotides. Green and yellow circles denote interacting regions, interactions are represented by dotted lines **(A)** An RNA stem-loop consisting of a heteroduplex stem with an unpaired terminal loop. **(B)** A pseudoknot whereby the terminal loop of the hairpin base-pairs to the downstream single-stranded region. **(C)** Long-range interactions between complementary unstructured sequences (yellow) or transient kissing-loop interactions between stem-loops (green).

emission of fluorophores (304). Alternatively, RNA structure can be determined biochemically using such methods as dimethylsulphate (DMS) footprinting, hydroxy radical footprinting and enzymatic analysis (305–307). Biochemical methods typically involve structure-specific modification of RNA followed by the identification of modification sites, either by fragment analysis of cleaved end-labelled RNA directly or following modification-sensitive primer extension. Hydroxyl radical footprinting involves cleavage of the sugar-phosphate backbone (308). Hydroxyl radical-induced cleavage occurs within milliseconds in a base-independent manner. Enzymatic analysis methods, whereby structure-specific nucleases cleave single- or double-stranded RNA, are similarly base-independent (307). In contrast, DMS methylates the base-pairing faces of nucleotides at N1 of adenosine, N7 of guanosine and N3 of cytidine (309). Thus, DMS footprinting is sensitive to changes in both secondary and tertiary structure.

Another such biochemical method of RNA structure determination, SHAPE mapping, can determine the likelihood of base-pairing at single nucleotide resolution. The reactivity of each nucleotide with the reagent NMIA is dependent on the availability of the 2′OH for acylation; only unpaired nucleotides become acylated. RNA is treated at sufficient concentration to achieve a single modification per RNA and reverse transcribed, producing a pool of cDNA fragments where reverse transcriptase cannot traverse the acyl adduct. Fragment analysis reveals the sites of NMIA modification. The proportion of fragments of each size generates a reactivity score for each nucleotide, giving the likelihood of base-pairing at that site in the average molecule (310). Reactivity data from biochemical methods such as SHAPE can be combined with computational approaches by assigning each nucleotide a relative reactivity score. *In silico* approaches to RNA structure determination typically rely on standard biochemical data obtained by melting oligonucleotides of known sequence. Reactivity scores for a specific RNA, such as those obtained via SHAPE, provide further constraints which can be introduced as a parameter during modelling, improving the prediction of parsimonious structures.

1.3.2 Functions of viral RNA structures

RNA structures regulate processes as diverse as packaging, translational initiation and recoding, genome replication and circularisation to regulate binding of *trans*-activating factors. Stem-loops impact RNA virus replication via sequence and/or structure-dependent mechanisms in a wide

range of virus families including *Retroviridae* (311), *Picornaviridae* (312), *Flaviviridae* (313), *Bunyaviriridae* (314) and *Orthomyxoviridae* (315).

Many RNA viruses utilise secondary structures in the genome to ensure specific packaging of unspliced, full length genomes and where appropriate, the correct array of genome segments. Packaging signals are usually discrete and can confer packaging specificity to non-viral reporters such as GFP-RNA (316). One widely studied packaging signal is the Ψ site in the human immunodeficiency virus 1 (HIV-1) genome, which consists of four stem-loops (SL1-4) and is sufficient for packaging of a reporter RNA into virus-like particles (VLPs) (317,318). NMR studies of the Ψ site have revealed a tandem three-way junction between the stem-loops which sequesters the splice donor site and the translation initiation site by long-range interactions (319) (Fig 1.20A). The Ψ site binds to the viral Gag protein with high affinity via specific tetraloop GGGG sequences in SL2 and SL3 (320–322). Recent studies have demonstrated that stabilisation of SL3 by the drug NSC disrupts Gag binding and thus packaging of viral genomes. This study provides proof of principle that RNA structure can provide direct targets for anti-viral interventions (323).

The function of viral RNA structures in translation is well studied, particularly for internal ribosome entry site (IRES) structures which enable cap-independent translation by direct recruitment of ribosomal subunits (324). There are four classes of IRES, each with decreasing dependence on cellular factors for initiating translation. Class I IRESs such as that of Poliovirus require a wide range of host initiation factors including eIF1-3/eIF4F/eIF5, while Class 4 IRESs such as that of Cricket paralysis virus requires none (325–327). Class III IRESs such as the hepatitis C virus (HCV) IRES have a complex architecture involving multiple stem-loops and a pseudoknot with helical junctions (306) (Fig 1.20B). The structure of the HCV IRES in complex with the human 40S and 80S ribosome has been resolved at 3.9 Å, revealing that the HCV IRES binds directly to the 18S ribosomal RNA (rRNA) to recruit the 40S ribosomal subunit; an apical loop of the IRES forms a kissing-loop interaction, pairing viral GGG with rRNA CCC (328,329). Direct recruitment of translation machinery enables evasion of the translational shut-off induced by phosphorylation of eIF2α (330,331).

5′ RNA elements are essential for genome replication during the lifecycle of many positive-sense RNA viruses including flaviviruses (332,333) and picornaviruses (334–336). The cloverleaf conserved replication element (CRE) of poliovirus (PV) is a well-known example, consisting of four stem-loops: a, b, c and d at the 5′ terminus of the genome (337,338) (Fig 1.20C). The cloverleaf was shown by foot-printing and EMSA to bind directly to the host protein PCBP2 via unpaired C nucleotides in stem-loop B and to the viral polymerase $3CD^{pol}$ and protease $3C^{pro}$ via stem-loop D. The protein complex nucleated around the 5′ cloverleaf interacts with PABP bound to the 3′ polyA tail, directing circularisation of the genome, transfer of the polymerase to the 3′ and subsequent negative strand synthesis (339). Mutations in the cloverleaf designed to selectively disrupt base-pairing in the positive strand were lethal to poliovirus replication. Compensatory mutations restored wild-type replication in most cases, indicating that secondary structure rather than sequence is the determining factor in cloverleaf function. However, one well- conserved sequence within a stem could not be compensated for, indicating that it contains sequences essential for replication (340).

In addition to local secondary structure formation, RNA structures can form between distant sequences and structures e.g. by kissing-loop interactions (341). Conformational switching is a common strategy during viral genome replication, whereby the global structure of the genome is regulated by competing local and remote interactions. In particular, long-range interactions can direct circularisation of the viral genome, bringing 5′ and 3′ elements into contact and enhancing translation and replication. The requirements for long-range interactions within the dengue virus (DENV) genome have been well-studied. Following interaction between the 5′ RNA stem-loop SLA and the viral RNA-dependent RNA polymerase (RdRp), long-range interactions and base-pairing between the 5′ and 3′ termini stabilises circularisation of the virus genome and RdRp transfer to a 3′ promoter (298,299). Analysis of RNA binding by EMSA identified 5′ and 3′ inverted complementary sequences necessary for cyclisation of the DENV genome. Mutation of either 5′ or 3′ cyclisation sequences resulted in nonviable virus, whereas compensatory mutations restored base-pairing and wild-type replication. The cyclisation sequences can be separated into three distinct duplex regions; the circularised form contains duplexes of 5′-3′ upstream of AUG (UAR), 5′-3′ downstream of AUG (DAR) and 5′-3′ cyclisation sequences (342) (Fig 1.20D). Both genomic conformations are necessary for productive infection. A structural element, the small hairpin (sHP), exclusive to the linear form is essential for viral replication and altered stability of this element relative to the UAR and DAR duplexes is detrimental (343,344).

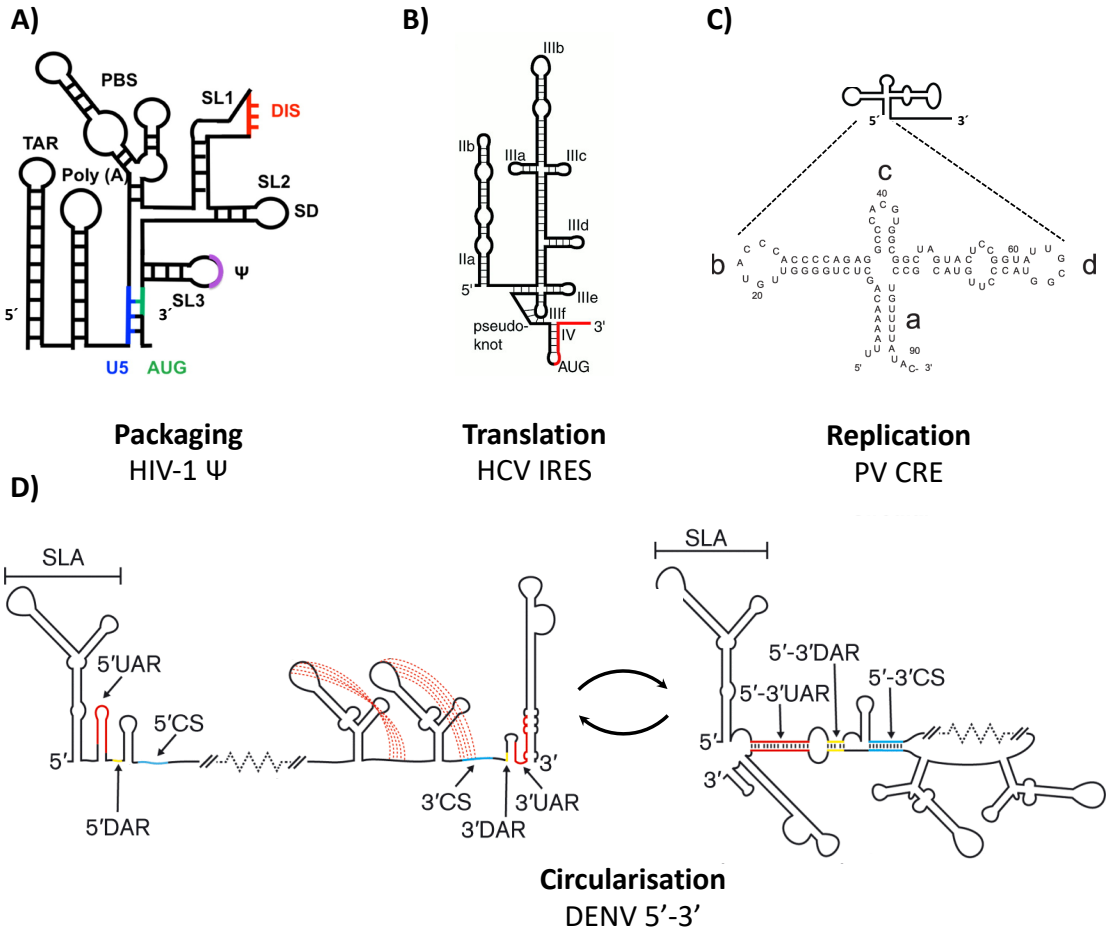


Figure 1.20: Examples of essential RNA structures in positive-sense viral genomes. RNA structures involved in **(A)** viral genome packaging, **(B)** cap-independent translation, **(C)** replication of the genome and **(D)** circularisation via long-range interactions. Figures adapted from: HIV-1 psi (345), HCV IRES (346), PV cloverleaf CRE (340) and DENV UAR/DAR/CS (347).

1.3.3 CHIKV RNA structures

RNA structural elements investigated in alphaviruses include the packaging signal mapped in nsP2 (208), translational regulation elements (348,349), replication elements (289,350,351) and an RNA stem–loop at the 5′ terminus of the 5′ UTR (SL3) which masks the alphavirus type-0 cap structure (352,353) (Fig 1.21).

1.3.3.1 Translational regulation elements

The mechanism of CHIKV genomic RNA translation is cap-dependent and does not require any RNA structures for initiation. However, RNA structures do contribute to translational recoding during expression of CHIKV proteins, specifically the nsP4 protein via readthrough and the TF protein via frameshifting. An opal stop codon (UGA) between *nsp3* and *nsp4* is responsible for termination codon readthrough, in combination with an elongated stem-loop (154,240). SHAPE analysis of the termination codon readthrough (TCR) element revealed a terminally forked stem-loop at the 3′ of the UGA (240). Variation/polymorphism between CHIKV lineages was found to maintain the structure by co-variation. TCR occurred at ~10% and could be abolished by disruption of the secondary structure. Translation of the TF protein occurs by -1 frameshift during translation of the 6K gene and requires the programmed ribosomal frameshifting (PRF) stem-loop (240,354). SHAPE analysis determined that the PRF is a small, bulged stem-loop with homology to that observed in EEV (240,348). A slippery heptanucleotide sequence upstream of the PRF element is also necessary for frameshifting, which occurs at ~2% (355). Mutational analysis of the PRF element utilised dual luciferase replicons and western blot to show disruption of TF expression (240).

Although the sub-genomic 26S RNA can undergo cap-dependent translation, this process is inhibited late during infection by phosphorylation of eIF2α during translational shut-off. eIF2α-independent translation of the sub-genomic RNA requires a highly stable hairpin predicted to form downstream of the AUG codon, termed the downstream stable hairpin (DSH) (203). The DSH is thought to stall the ribosome 43S scanning complex at the AUG and induce the closed ribosomal complex, possibly by binding of the DSH in the P site to initiate translation. The DSH confers independence from eIF4G and eIF2 and may interact with eIF3 as an alternative cap-binding mechanism (202,356). This structure is only required when eIF2α is activated by PKR (349).

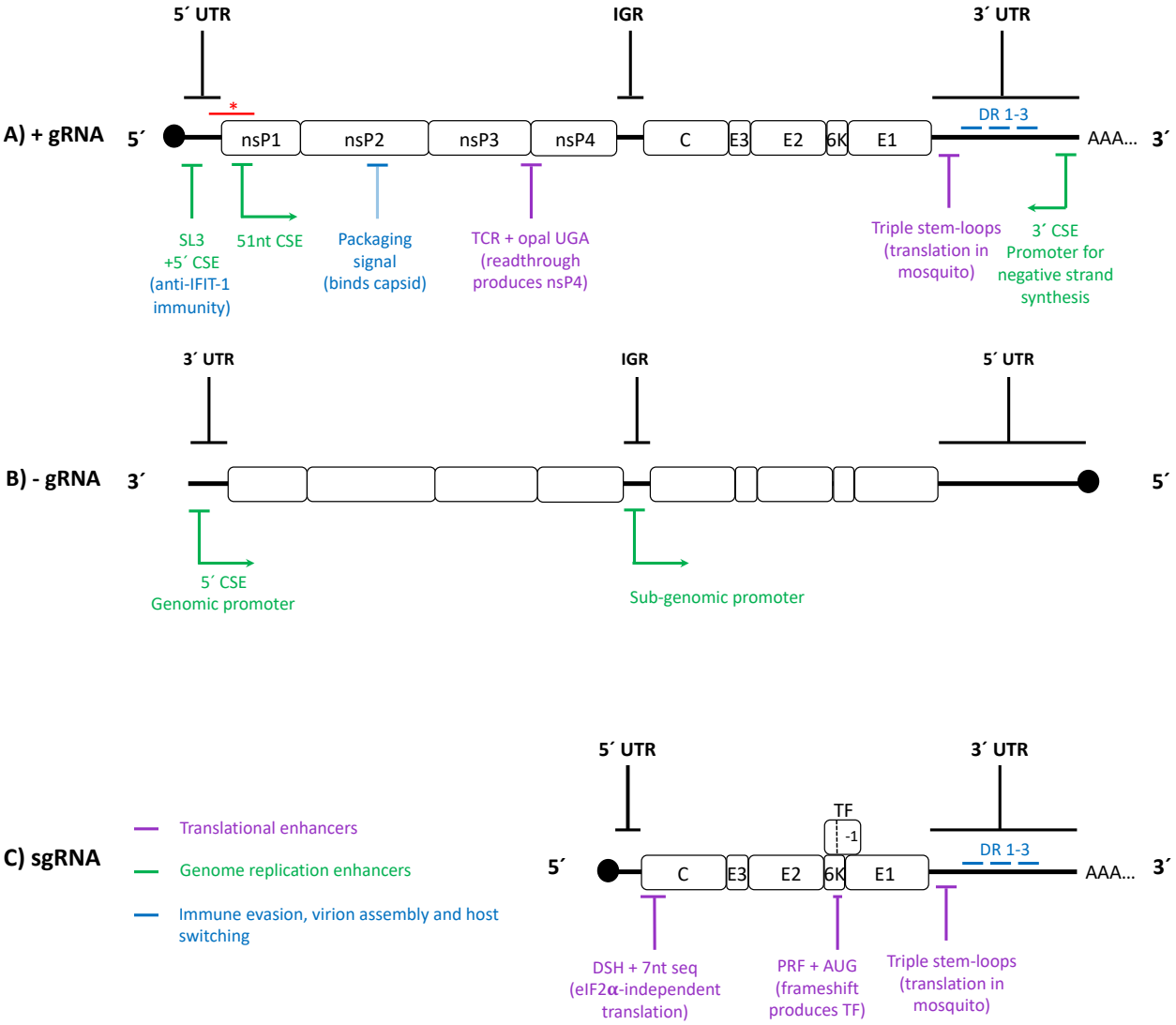


Figure 1.21: CHIKV RNA structural elements and related sequence requirements. Conserved RNA structural elements within the **(A)** positive-sense genome (+gRNA), **(B)** negative-sense replication intermediate (-gRNA) and **(C)** sub-genomic (sgRNA) RNA. Translational enhancers (purple), replication elements (green) and other functional elements (blue) are labelled and the functions described briefly. The untranslated regions (UTR) and intergenic region (IGR) are denoted, along with the 5′ cap (black circle), polyA tail (AAA…) and the 5′-3′ polarity of each strand. The conserved sequence elements (CSE), termination codon readthrough (TCR) element, the downstream stable hairpin (DSH) element, programmed ribosomal frameshift (PRF) and direct repeats (DR) are shown. Red line and (*) indicate the region examined during this project. Data drawn from: (13,154,203,208,240,289,298,353,357–360).

1.3.3.2 Replication elements

The information which follows was almost exclusively obtained from studies of VEEV, SINV and SFV. Much of what is known about CHIKV replication is assumed based on these related alphaviruses.

1.3.3.2.1 Sub-genomic promoter

The promoter for production of sub-genomic 26S RNA occurs in the intergenic region between ORF1 and ORF2 and is recognised by nsP4 in the negative strand. nsP4 is unable to bind the sub-genomic promoter in isolation and requires nsP1-3 (361). Truncation analysis revealed that 19 nt upstream of the start site and 5nt downstream are necessary and sufficient to drive sub-genomic RNA synthesis (360). The complete promoter spans 98 nt upstream and 14 nt downstream of the start site, containing flanking elements which act synergistically. The essential promoter region and flanking enhancer elements were conserved during passaging studies in mosquito and mammalian hosts (362). SINV replicase can recognise all known alphaviral sub-genomic promoters with the exception of SFV, indicating that the sub-genomic promoter is functionally and structurally conserved despite some sequence variation (363).

1.3.3.2.2 3′ UTR

The 3′ UTR of the CHIKV genome is 500-900 nt in length and contains three non-essential direct repeat sequence elements (DR1-3) and a conserved sequence element (CSE) adjacent to the polyA tail. DR1 and DR2 are unstructured but DR3 is predicted to form a stable Y-shaped secondary structure (298). The number of direct repeat duplications in the CHIKV 3′ UTR is lineage-specific, with the Asian CHIKV genotype containing the most duplications. A greater number of duplications is associated with increased replication in mosquito cells and a corresponding fitness loss in mammalian cells. The DRs are utilised as sites of recombination during host switching, mitigating the high cost of adaptation to different hosts by increasing genetic flexibility (298). The DRs have also been demonstrated via riboprobes to bind to the La autoantigen of mosquitoes and the human-antigen R (HuR) protein. These interactions stabilise viral RNA, protect against deadenylation and sequester HuR in the cytoplasm which prevents post-transcriptional regulation and stabilisation of host mRNA (364,365).

The 3′ CSE is the core negative strand promoter, consisting of the terminal 19 nt of the 3′ UTR and a minimum of 11-12 adjacent adenosines in the polyA tail. Deletion or substitution of the terminal 13 nt severely reduced or abolished negative strand synthesis (289,366). The 19 nt element terminates in a conserved cytosine, which determines the initiation site for negative strand synthesis. Mutation of the minimal promoter resulted in initiation within the polyA tail, although the cytosine could be substituted with three uracils (351). The conserved cytosine can also be displaced by up to 7 nt in either direction without abolishing recognition of the promoter. In addition, replication of certain 3′ CSE mutants was inhibited in mosquito cells and others in avian cells, indicating host-specific elements within the 3′ UTR (367).

Interestingly, SINV mutants with deletions of the 3′ CSE and polyA tail were found to be capable of genome replication when transfected into BHK-21 cells. Sequence analysis of eight isolates revealed insertion of AU rich sequences and restoration of a polyA tail, suggesting a repair pathway which likely involves the TATase activity of nsP4 (368). AU-rich sequences at the 3′ end conferred recognition and no novel compensatory mutations were detected in the rest of the genome during passaging (369). Furthermore, these repaired isolates can replicate to wild-type titres in neonatal mice, producing comparable viraemia and pathology. One isolate with UAUUU within the polyA tail maintained the motif for 9 passages in animals, indicating that it is relatively stable. The 3′ CSE is therefore not absolutely essential for SINV replication, stability and pathogenesis in animals (370).

1.3.3.2.3 5′ UTR

The 5′ UTR of the CHIKV genome is 76 nt in length with a 5′ AU dinucleotide, followed by a short GC-rich stem-loop (SL3 in CHIKV, labelled as SL1 in other alphaviruses). mFold-predicted structures of this stem-loop in alphaviruses suggest dramatic differences in the apical region, with conservation of the GC-rich base. For example, the terminal loop in SINV forms a pseudoknot with the upstream single-stranded sequence which is not seen in other alphaviruses (305,371). SL3 has been shown to prevent translational inhibition by evading interferon-induced protein with tetratricopeptide repeats 1 (IFIT-1) recognition of the type-0 cap. Mutation of SL3 resulted in restriction by IFIT-1 *in vivo* and *in vitro,* with no effect in IFIT-1 knockout MEFS. Compensatory mutations in SL3 of a VEEV/SINV chimera allowed IFIT-1 evasion (353,371).

SL3 in the genomic RNA functions as part of the core promoter for negative strand synthesis, and its complement in the negative-sense genomic RNA acts as the core promotor for positive strand synthesis (357,366,372). Deletion of any of the first 5 nt of the SINV genome was lethal *in vivo* and mutations in the first 15 nt were either lethal or severely inhibited viral replication. The severity of SL3 mutations was host-dependent; deletions in the stem had a stronger negative effect on replication in mosquito cells compared to avian cells (372).

SL3 is thought to interact with the 3′ CSE during negative strand synthesis. Panhandle structures have been observed between the 5′ and 3′ ends of the SINV genome by electron microscopy but there are no perfectly complementary regions which base-pair (373). SINV and SFV genomic RNA extracted from virions was shown to cyclise via biophysical assays. The linear and circular forms can be distinguished in a sucrose gradient and cyclisation was predicted to depend on a 10-20 nt region (374). SINV constructs lacking the 3′ CSE inhibit negative strand replication in a dose-dependent manner during template competition assays, indicating that the 5′ of these templates can bind factors involved in negative strand synthesis. During studies of chimeric alphaviruses, homologous 5′ and 3′ UTRs could be used as a template by a variety of alphaviral replicases. One exception was the SFV 5′ UTR, which can only be utilised by the SFV replicase. The incompatibility of the SFV 5′ UTR and SINV replicase was used to design a series of chimeric 5′ UTRs which identified the terminal dinucleotide and SL3 as the core elements necessary for recognition by SINV replicase (357). Mutations which disrupted the structure of SINV SL1 were found to improve translatability of the genome but decrease negative strand replication in BHK-21 and C6/36 cells (305).

Studies of the SL3 homologue in VEEV demonstrated that the stability and sequence of the stem rather than the loop sequence was necessary for promoter function. Compensatory mutations substituting a GC with a CG pair had a deleterious effect on replication. Passaging of deletion mutations yielded compensatory mutants with a 5′ addition of AU or AUG repeats similar to those seen for 3′ CSE deletion. Substitution of either a 5′ A or U was extremely deleterious and resulted in multiple AU additions at the 5′ terminus, indicating that the AU dinucleotide is critical for viral genome replication. Disruption of the short stem-loop also resulted in the addition to heterologous

sequences capable of forming secondary structure. Pseudorevertant mutations were detected in the non-structural proteins at the C-terminus of nsP1, the N-terminus of nsP2 and in nsP4 (358). Similarly, passaging of SINV 5′ mutants lacking the 5′ AU dinucleotide invariably resulted in regeneration of the 5′ AU. In one case, a deletion of 16nt restored a 5′ terminal AU. Some revertants with multiple AU repeats could replicate to higher than wild-type titres in mammalian cells but were inhibited in mosquito cells, indicating host-specific adaptive mutations (373).

1.3.3.2.4 51 nt CSE

Two stem-loops within the *nsp1*-coding region, collectively termed the 51 nt CSE, have been demonstrated to enhance SINV and VEEV replication. The sequence and structure of the 51 nt CSE are highly conserved; SL165 is identical in many alphaviruses and while there is a greater degree of sequence covariation within SL194, the structure is retained. SHAPE-constrained thermodynamic folding predictions for CHIKV, SINV and VEEV reveal remarkable conservation of the structures within the 51 nt CSE (13) (Fig 1.22). The 51 nt CSE is not essential for SINV replication in mammalian cells. Early work suggested that the SINV 51 nt CSE acts as a replication enhancer in vertebrate cells, whereby disruption reduced RNA replication 5-10 fold in BHK-21 cells and resulted in lower viral titres during infection of mice (269,357). Enhancement of replication by the 51 nt CSE in mammalian cells was not recapitulated in later studies. Viral replication in BHK-21 cells was not affected by the introduction of 43 mutations in the nsP1-coding region, producing similar viral titres to wild-type in a one-step growth assay and similar infectivity in an infectious centre assay (350).

Mutation of the 51 nt CSE often results in host-specific phenotypes in SINV. One double mutant of SL165 which retained base-pairing in the stem replicated to wild-type titres in chicken embryonic fibroblasts but exhibited a 90% reduction in replication in mosquito cells. Another SL165 stem mutant significantly reduced replication in both host species. Compensatory mutants designed to retain base-pairing in SL194 grew poorly in chicken cells but slightly better than the single mutants. Single mutants in mosquito cells grew ten-fold better than double mutants, suggesting that the sequence of SL194 is required. Substitutions in the terminal loop regions of the 51 nt CSE resulted in a 100-fold decrease in replication in avian and mosquito cells, suggesting

sequence-specific binding during replication (375). Generally, disruption of the 51 nt in SINV produces a more severe replication defect in mosquito cells than in vertebrate cells (350).

The ability of SINV 51 nt CSE mutants to act as templates was analysed by addition of the replicase *in trans* and detection of via [α-^{32}P]-labelled positive strands in infected cell lysates. Complementing this approach, an uncleavable P123/nsP4 replicase was used to assess the ability of the constructs to template negative strand synthesis *in vitro*. Deletion of SL165 or SL194 resulted in defective negative strand synthesis. Individual deletions of SL165, SL194 and an elongated upstream stem-loop (SL85/102) did not decrease RNA replication *in vivo* but deletion of both stem-loops of the 51 nt CSE reduced replication 3-10 fold. Synonymous mutation of the 51 nt CSE and SL85/102 in combination resulted in 50-fold lower replication *in vivo* and 4-6 fold lower negative strand synthesis *in vitro*. Thus, structures within *nsp1* are necessary for enhancing production of both positive and negative strands in SINV and the complement of this region in the negative strand likely influences the production of positive strand genomic RNA (357).

During one study of the 51 nt CSE in VEEV, the presence of secondary structure and the expression of nsP1 were uncoupled by extensive synonymous mutation to remove secondary structure in the coding region, followed by incremental re-addition of structural elements into the 5′ UTR. Deletion of the entire 51 nt CSE reduced viral replication to a barely detectable threshold. None of the individual deletions prevented replication, although the greatest reduction in viral titre was caused by deletion of SL165. The same phenotypes were observed in BHK-21 and C7/10 mosquito cells. Two passages in BHK-21 cells produced several pseudorevertants: in SL85/102 downstream of the initiating AUG, in the N-terminal domain of nsP2 and in nsP3. When reverse genetically reincorporated into VEEV, mutations in nsP2 and SL2 were additive and not synergistic, possibly affecting different processes (376). During similar passaging studies in SINV, suppressor mutations also arose in nsP2 and nsP3, indicating that these regions or proteins interact. Each of the pseudorevertants contained the same mutation at the N-terminus of nsP2 (E118K), one had a mutation in nsP3 and two had mutations in the 5′ UTR whereby the terminal 9nt were replaced with ~50nt from the sub-genomic promoter. Each of the revertants also contained mutations in E2. Reincorporation of the nsP2 mutant in isolation did not restore wild-type replication but double mutants of nsP2-nsP3 or nsP2-UTR had a strong synergistic effect restoring

SINV replication, in the absence of the E2 mutations. The suppressor mutations which restored replication in C7/10 cells resulted in replication defects in BHK-21 cells, again suggesting host-specific interactions (350).

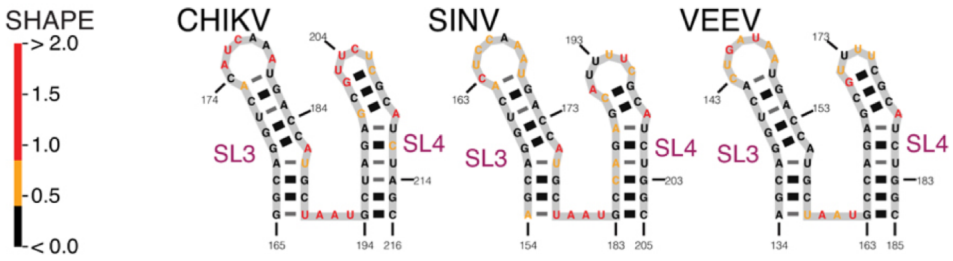


Figure 1.22: SHAPE-constrained thermodynamic folding prediction of the 51 nt CSE in CHIKV, SINV and VEEV (13). SHAPE reactivities are shown as a heat map: black SHAPE reactivities between 0–0.5 and increasing intensities from orange to red indicate increasing SHAPE reactivities, as denoted by the key. High reactivity (red) denotes unpaired nucleotides whereas low reactivity (black) denotes base-paired nucleotides. Stem-loop 3 (SL3) and stem-loop 4 (SL4), as labelled by Kutchko and colleagues based on the number of stem-loops in the 5′ region, correspond to SL165 and SL194 during later labelling schemes based on the first nucleotide of the structure.

1.4 Aims of the project

As discussed, there are no currently available vaccines or antiviral therapies for CHIKV; infection cannot be treated or prevented except by control of vector populations. A greater understanding of the viral replication cycle in both human and mosquito hosts is essential for the development of prophylactics and treatments. Recent work by A. Tuplin (University of Leeds) provided a preliminary SHAPE map of the 5′ terminus of the CHIKV genome, demonstrating the existence of several distinct RNA secondary structures, some novel and some with homology to previously described replication elements. These RNA structures may be necessary for productive replication of the virus and represent direct drug targets for intervention, as has been demonstrated for NSC-mediated disruption of the Ψ-site which inhibits HIV-1 replication.

Therefore, the aim of this PhD project is to investigate the importance of each of these structures during the CHIKV lifecycle in human and mosquito hosts. More specifically, the objectives of the project are:

1. Determine the phenotypic importance of each stem-loop for the CHIKV lifecycle in human and mosquito cells at multiple stages of the viral replication cycle

2. Investigate the sequence and structure requirements for functional interactions of the stem-loops and explore the routes of reversion taken during prolonged passaging

3. Identify potential tertiary interactions including pseudoknots

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Plasmids

All CHIKV cDNA constructs are shown in Fig 2.1. The full-length virus ECSA strain infectious clone (CHIKV_IC) and CHIKV monoluciferase sub-genomic replicon construct (CHIKV_Rep) were kind gifts from Prof A. Merits, University of Tartu. The translation reporter replicon (CHIKV_Rep(GDD>GAA)) was adapted from the dual luciferase sub-genomic replicon by S. Bradley (University of Leeds). Mutations were designed in the active site of the RdRp (GDD>GAA) which abrogate replication, allowing measurement of translation of the first ORF.

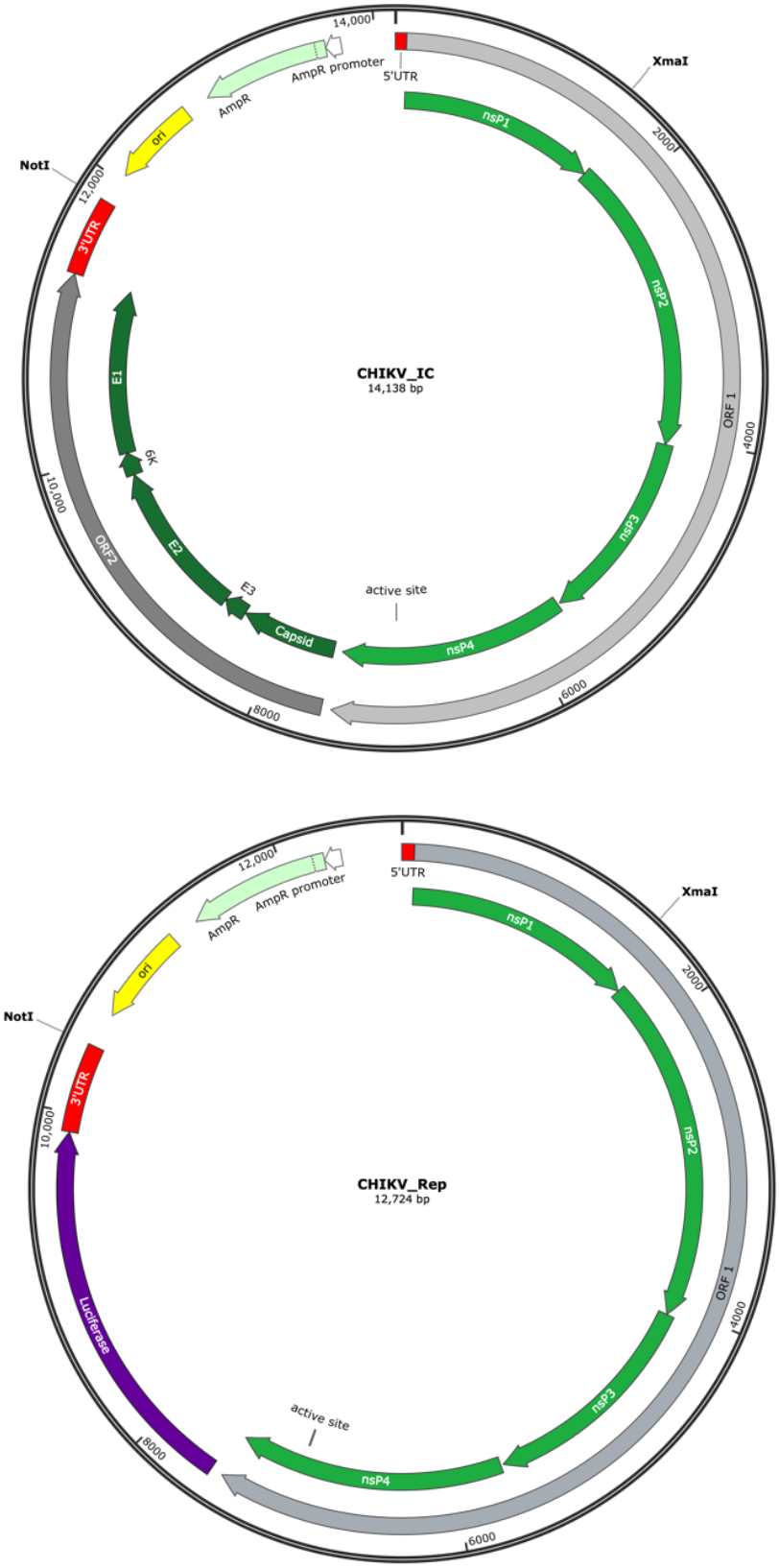
2.1.2 Primers

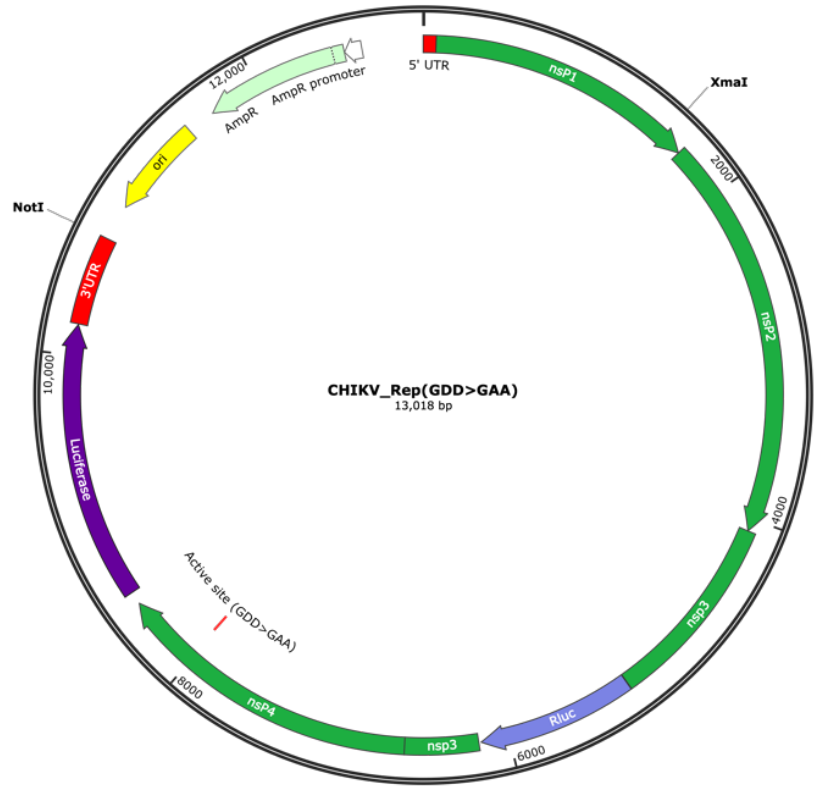
All oligonucleotide primers are shown in the appendix (Table 1). Oligonucleotide primers were synthesised for use in sequencing, mutagenesis and polymerase chain reaction (PCR) (Integrated DNA Technologies). Primers were resuspended in nuclease-free H_2O and stored at -20°C.

2.1.3 Bacterial strains

Supercompetent XL-1 Blue *E. Coli* (Agilent Technologies) were used for transformation of intact plasmids. Ultracompetent XL-10 Gold *E. Coli* (Agilent Technologies) were used for transformation of ligation products and mutagenized plasmids digested with *DpnI*. Mix and Go! *E. Coli* (Zymo Research) were adopted for all transformations from 2017 onwards. Glycerol stocks were made as follows: single colonies were shaken at 37°C in 5.5 mL LB containing amplicillin (80mg/L) overnight. 250 μL of the culture was added to 250 μL glycerol/LB 1:1 mixture in a sterile cryovial and stored at -80°C.

A)


CHIKV_IC
14,138 bp
14,000
AmpR promoter
5'UTR
XmaI
2000
nsP1
nsP2
nsP3
nsP4
ORF 1
4000
6000
active site
8000
Capsid
E3
E2
6K
E1
ORF2
10,000
12,000
3'UTR
NotI
ori
AmpR
CHIKV_Rep
12,724 bp
Luciferase



B)

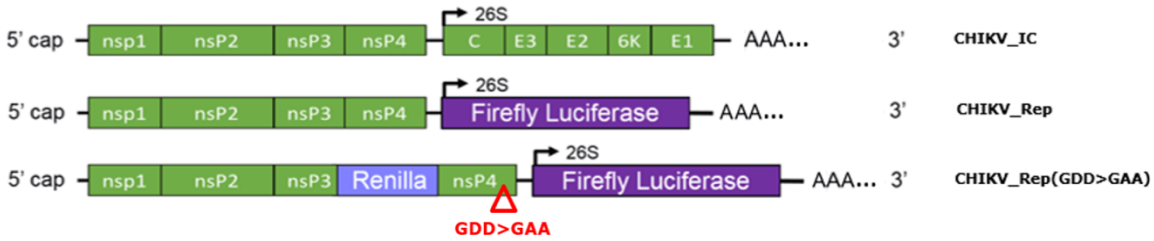


Figure 2.1: CHIKV cDNA constructs. A) cDNA plasmid constructs for full length virus (CHIKV_IC), monoluciferase replicon (CHIKV_Rep) and dual luciferase translation reporter replicon (CHIKV_Rep(GDD>GAA)). NotI was used for linearization in all cases, *NotI* and *XmaI* were used for excision of mutated regions of interest for cloning**. B)** RNA synthesised from each plasmid.

2.1.4 Cell culture

Continuous mammalian and mosquito cell lines, along with specific culture conditions, are listed in Fig 2.2. Mammalian cells were passaged by treatment with trypsin (Sigma Aldrich) at 37°C for 5-8 minutes. Insect cells were passaged by gentle scraping into 10 mL complete Leibovitz L-15

media (Gibco). Cells were centrifuged at 250 x *g* for 3 minutes, resuspended and diluted before seeding in T175 flasks in a total of 25 mL complete media.

Cell stocks were produced by resuspending $1x10^7$ cells in 900 µL foetal bovine serum (FBS) and 100 µL dimethyl sulfoxide (DMSO) before cooling -1°C/minute using Mr Frosty (Thermo Fisher Scientific). Cells were stored at -80°C overnight and transferred to liquid nitrogen for long term storage.

2.2 Methods in molecular biology

2.2.1 Agarose gel electrophoresis

The size and purity of plasmid DNA was determined by agarose gel electrophoresis, carried out using 1% agarose gels (1% analytical grade agarose (Sigma Aldrich), 1x TAE buffer (40 mM Tris-Acetate, 1 mM EDTA), 1:10,000 dilution SYBR SAFE DNA stain (Life Technologies)). Samples were mixed with 6x Purple Gel Loading Dye (NEB) and loaded alongside 1 KB Plus DNA ladder (NEB). Gels were run in 1x TAE buffer at 100 V for 60 minutes before visualisation under UV transillumination.

Larger gels were used for extraction of restriction fragments during cloning, these were run at 80 V for a minimum of 2.5 hours for clear separation of fragments. Appropriate bands were excised from the gel and DNA purified using Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions, eluting in 20 µL nuclease-free H_2O.

2.2.2 Denaturing MOPS agarose gel electrophoresis

The purity, integrity and size of RNA products of *in vitro* transcription was determined by denaturing formaldehyde agarose gel electrophoresis in MOPS buffer. 1% agarose MOPS gels

Cell line	Species/Features	Media	Maintenance	Seeding Dilutions
C2C12	Murine embryonic myoblasts (muscle progenitor) (377)	Dulbecco's modified eagle media (DMEM, Sigma Aldrich) + 20% (v/v) foetal bovine serum (FBS, Life technologies) + 1% penicillin/streptomycin (P/S)	37°C 5% CO_2 Passaged by trypsinisation	1:5 to 1:20
BHK-21	Baby Hamster Kidney fibroblast cell origin (378)	DMEM + 10% FBS + 1% P/S		
Huh7	Human, hepatocellular carcinoma (379)	DMEM + 10% FBS + 1% P/S		1:4 to 1:8
A549	Human lung carcinoma (380)	DMEM + 10% FBS + 1% P/S		
U4.4	Mosquito embryonic cells, *Ae. albopictus* (381)	Leibovitz's L-15 media (Gibco) + 10% FBS + 10% Tryptose phosphate broth (TPB)(Gibco) + 1% P/S	28°C Passaged by mechanical scraping	1:5 to 1:10
C6/36	Mosquito embryonic cells, *Ae. albopictus,* impaired RNA interference (RNAi) antiviral response (382)	Leibovitz's L-15 media + 10% FBS + 10% TBP + 1% P/S		

Figure 2.2: Cell lines and culture conditions

were prepared using 1% (w/v) analytical grade agarose (Sigma Aldrich), 1x MOPS buffer (40 mM MOPS pH 7.0, 10 mM NaAc, 1 mM EDTA pH 8.0), ddH_2O and 1:10,000 dilution SYBR SAFE DNA stain (Life Technologies). After melting of the agarose, formaldehyde was added to a final concentration of 1.7% v/v. Samples were mixed with 2x denaturing RNA loading dye (NEB), heated to 65°C for 10 minutes and cooled on ice. Samples were loaded alongside Millenium RNA ladder (Ambion). Gels were run in 1x MOPS buffer at 100 V for 60 minutes before visualisation under UV transillumination.

2.2.3 Bacterial transformation

For plasmid DNA amplification, 500 ng DNA was mixed with 25 µL super-competent XL-1 Blue *E. coli* (Agilent Technologies) and cooled on ice for 30 minutes. Cells underwent heatshock at 42°C for 30 seconds before recovery on ice for 2 minutes. 500 µL SOC media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) was added to the transformation mix and shaken at 37°C for 1 hour. 150µL transformation mix was then plated onto agar plates containing 1.5% bacteriological agar, Luria broth (LB) (10 g tryptone, 5 g yeast extract, 10 g NaCl, autoclaved in 1 L ddH_2O) and antibiotic to a final concentration of 100 µg/mL. Plates were incubated at 37°C for 16 hours until colony formation.

For transformation of ligation products and mutagenized plasmids digested with *DpnI* following PCR, ultra-competent XL-10 Gold *E. coli* (Agilent Technologies) were used. 5 µL of ligation or PCR product was mixed with 50 µL ultra-competent cells and heatshock carried out as described above. Mix and Go! *E. coli* (Zymo Research) were adopted for all transformations from 2017 onwards and used according to the manufacturer's instructions. Briefly, the appropriate amount of DNA is mixed with 50 µL Mix and Go bacteria, incubated for 10 minutes on ice and plated directly onto pre-warmed 37°C LB-agar plates containing the relevant antibiotic.

2.2.4 Plasmid DNA amplification

Single colonies were picked from agar plates following transformation and cultured in an appropriate volume of LB. For small volume amplifications, single colonies were grown overnight at 37°C in 5.5 mL LB containing antibiotic to a final concentration of 100 µg/mL. Cultures were

pelleted at 4000 x *g* for 20 minutes at 4°C and plasmid DNA isolated and purified using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific).

For larger volume amplifications, single colonies were grown in 5.5 mL LB containing the appropriate antibiotic for 8 hours before transfer to a 250 mL culture overnight. Cultures were pelleted at 5000 x *g* for 10 minutes at 4°C and plasmid DNA isolated and purified using the GeneJET Plasmid Maxiprep kit (Thermo Fisher Scientific). Purified DNA concentration and quality was quantified using a Nanodrop1000 spectrophotometer (Thermo Fisher Scientific). Integrity, purity and concentration of plasmid DNA was then assessed via agarose gel electrophoresis.

2.2.5 Restriction endonuclease digestion

Restriction endonuclease digestion was used for linearization of plasmid DNA for diagnostic purposes, for use as a template during *in vitro* transcription and for cloning. Diagnostic digests were carried out in a total volume of 10 µL ddH_2O, containing: 1 µg DNA, 1x Cutsmart buffer (NEB) and 0.1 µL (1-2 U) each enzyme (NEB). Diagnostic digests were incubated for 1 hour at the appropriate temperature and analysed by agarose gel electrophoresis without purification. Linearisation of plasmid DNA for use as a template for *in vitro* transcription was carried out in a total volume of 150 µL nuclease-free H_2O, containing: 25 µg DNA, 25 µL 1x Cutsmart buffer (NEB) and 2.5 µL (50 U) NotI-HF (NEB). Linearisation reactions were incubated for 1 hour at 37°C. Representative gel shown in Fig 2.3. Double digests to produce components necessary for cloning were carried out in a total volume of 150 µL nuclease-free H_2O, containing: 25 µg DNA, 1x Cutsmart buffer (NEB), 2.5 µL (50 U) *NotI-HF* (NEB) and 2.5 µL (25 U) *XmaI* (NEB). Double digests were incubated for 1 hour at 37°C.

2.2.6 DNA purification

Linearised plasmid DNA was purified using Wizard SV Gel and PCR Clean-Up System (Promega), eluting in 40 µL nuclease-free H_2O. After 2017, ethanol precipitation was used for purification of linearised DNA. 2-3 volumes 100% EtOH and 1/10 volume 3 M NaAc pH 5.2 were added to each reaction and incubated at -20°C overnight. The reaction was centrifuged at 17,000

x *g* for 30 minutes at 4°C, washed 1x with 70% EtOH and dried for 10 minutes before resuspension in 40 μL nuclease-free H_2O.

Double digests for use in cloning were subjected to separation by agarose gel electrophoresis as described above and the appropriate bands excised with minimal agarose. Gel slices were melted and the DNA purified using Wizard SV Gel and PCR Clean-Up System (Promega), eluting in 20 μL nuclease-free H_2O.

2.2.7 Site-directed mutagenesis

Site-directed mutagenesis was carried out using the QuikChange XL Site-Directed Mutagenesis kit (Agilent). Primers were designed according to the manufacturer's instructions. The protocol was amended as follows: reactions contained 10x reaction buffer (Agilent), 300 ng DNA, 125 ng forward primer, 125 ng reverse primer, 1 μL dNTP mix (Agilent), 3 μL QuikSolution (Agilent) and 1 μL (2.5 U) *PfuTurbo* DNA polymerase (Agilent). The cycling parameters were set to: 95°C for 2 minutes; 18 cycles of 95°C for 1 minute, 60°C for 50 seconds, 68°C for 25 minutes; 68°C

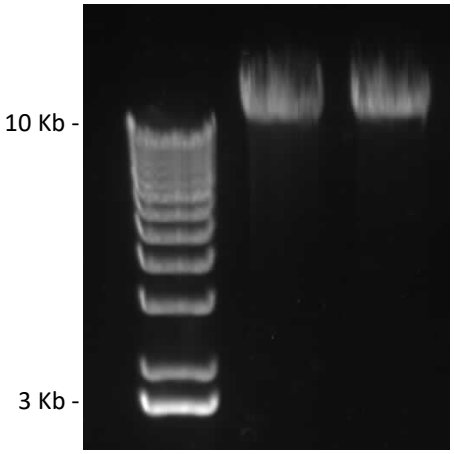


Figure 2.3: Linearised DNA. 1μg NotI-linearised CHIKV_Rep DNA on 1% agarose gel alongside 1 μL 1 KB Plus DNA ladder (NEB).

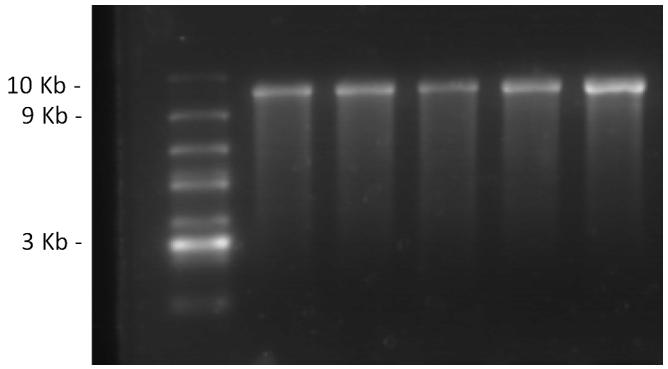


Figure 2.4: *in vitro* transcribed RNA. 500 ng *in vitro* transcribed CHIKV_Rep RNA on 1% MOPS gel alongside 1 μL Millenium RNA ladder (NEB).

for 13 minutes. Each reaction was then treated with 1 μL (10 U) *DpnI* to remove input DNA before transformation as described above. Several colonies were analysed via plasmid amplification, purification and sequencing to determine whether desired mutagenesis had occurred. Due to the error-prone nature of the *PfuTurbo* enzyme, correctly mutated regions were excised by double digest and ligated into the appropriate backbone which had not undergone mutagenesis.

2.2.8 DNA ligation

Ligation reactions were carried out at a 5:1 molar ratio of insert:vector. Ligation reactions contained 100 ng DNA backbone, 10x T4 DNA ligase buffer (NEB), 2 μL (800 U) T4 DNA ligase (NEB) and the appropriate amount of insert, made up to 20 μL total volume with nuclease-free H_2O. Ligation reactions were incubated at room temperature overnight before transformation as described above.

2.2.9 DNA Sequencing

500ng DNA was sequenced with 25 pmol primer in a total volume of 10 μL using Eurofin Genomics (formerly GATC) Sanger sequencing service.

2.3 Replicon production and transfection

2.3.1 *In vitro* transcription

Capped RNA was synthesised for use in replicon transfections and generation of infectious virus. CHIKV replicon RNA was synthesised using the mMessage mMachine SP6 Transcription kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Full-length infectious CHIKV RNA was synthesised using the Amplicap SP6 High Yield Message Maker kit (CamBio) according to the manufacturer's protocol.

Uncapped RNA was synthesised for use in SHAPE mapping. 2 μg DNA template was mixed with 10x RNA pol buffer (NEB), 3.2 μL ATP (100 mM) (NEB), 3.2 μL UTP (100 mM) (NEB), 3.2 μL CTP (100 mM) (NEB), 3.2 μL GTP (100 mM) (NEB), 2 μL (80 U) RNasin RNase inhibitor (Promega), 3.2 μL (0.32 U) Yeast Inorganic Pyrophosphatase (YIPP)(NEB), 1.12 μL 1 M $MgCl_2$

and 4 μL (80 U) SP6 polymerase (NEB). The mixture was made up to 80 μL with nuclease-free H_2O and incubated at 40°C for 3.5 hours. 1 μL (40 U) RNasin RNase inhibitor (Promega) and 2 μL (4 U) DNAseI (NEB) were added and incubated at 37°C for 30 minutes.

2.3.2 RNA purification

RNA was purified using the RNeasy Mini kit (QIAGEN) according to the manufacturer's protocol, eluting in 40 μL nuclease-free H_2O. From 2017, RNA was purified by Lithium Chloride (Thermo Fisher Scientific) precipitation according to the manufacturer's instructions. Purified RNA concentration and quality was quantified using a Nanodrop1000 spectrophotometer (Thermo Fisher Scientific). Integrity, purity and concentration of all transcripts was then assessed via denaturing formaldehyde gel electrophoresis.

2.3.3 Replicon transfection of mammalian cells

Lipofectamine transfection was used for replicon assays to allow for standardisation of mammalian and insect cell transfection methods. Mammalian cells were seeded at $1x10^5$ cells per well in 24 well plates in 1 mL complete media until 70-80% confluent. For experiments longer than 12 hours duration, cells were seeded at $5x10^4$ cells until 60-70% confluent. 500 ng CHIKV replicon RNA per well was mixed with 50 μL Optimem per well. 1 μL Lipofectamine2000 per well was mixed with 50μL Optimem per well and incubated at room temperature for 5 minutes. The RNA mix was gently mixed with the lipofectamine mix and incubated at room temperature for 20 minutes. Cells were washed 1x in PBS and media replaced with 500 μL Optimem per well. 100 μL per well of the RNA-liposome mix was added dropwise to each well and mixed by gentle swirling. Cells were incubated at 37°C for the appropriate duration. Monolayers were washed 1x in PBS, 100 μL passive lysis buffer (PLB, Promega) was added and rocked for 20 minutes at room temperature before harvest. Samples were stored at -80°C prior to analysis.

2.3.4 Replicon transfection of insect cells

Insect cells were seeded at $1x10^5$ cells per well in 24 well plates in 1 mL complete media until 70-80% confluent. 250ng CHIKV replicon RNA per well was mixed with 50 μL Optimem per well. 1 μL Lipofectamine2000 per well was mixed with 50 μL Optimem per well and incubated at room

temperature for 5 minutes. The RNA mix was gently mixed with the lipofectamine mix and incubated at room temperature for 20 minutes. Cells were washed 1x in PBS and media replaced with 500 μL Optimem per well. 100 μL per well of the RNA-liposome mix was added dropwise to each well and mixed by gentle swirling. Cells were incubated at 28°C for the appropriate duration. Monolayers were washed 1x in PBS, 100 μL PLB was added and rocked for 20 minutes at room temperature before harvest. Samples were stored at -80°C prior to analysis.]

2.3.5 Luciferase assay

Lysates resulting from replicon transfections were analysed for luciferase activity using a FLUOstar microplate reader (BMG LABTECH). 10 μL lysate was analysed for each replication-competent replicon. 30 μL lysate was analysed for the replication-deficient GAA translation reporter replicon due to low signal. Lysates were transferred into a white-bottomed 96 well plate and luciferase activity measured by auto-injection of LARII and Stop & Glo reagents (Promega) at 50 μL per well.

2.4 Virus production and infection

All virus work was carried out under BSL3 containment conditions, in accordance with the classification of CHIKV as a Category 3 virus. Permission was obtained from the Health and Safety Executive.

2.4.1 Electroporation with viral RNA

Electroporation transfection was used for production of viral stocks in BHK-21 cells and in other cell types where viral stocks could not be produced in BHK-21 cells. Mammalian cells were trypsinised, centrifuged for 3 min at 1000 x *g* and washed 2x in ice cold DEPC PBS before resuspension at $3x10^6$ cells/mL in DEPC PBS. $1.2x10^6$ cells were added to 1 μg purified RNA in a pre-cooled 4mm cuvette and electroporated using a Gene Pulser Xcell (BioRad) with a square-wave protocol: 260 V for 25 Ms. Electroporated cells were resuspended in 10 mL complete media and incubated in a T75 flask at 37°C for a minimum of 24 hours.

Insect cells were harvested via scraping, washed 1x in ice-cold DEPC PBS and resuspended at $1x10^7$ cells/mL in cold cytomix (2 mM EGTA pH 7.6, 120 mM KCl, 0.15 mM $CaCl_2$, 10mM K_2HPO_4/KH_2PO_4 pH 7.6, 25 mM HEPES pH 7.6, 5 mM $MgCl_2$.6H2O, 0.5% Dextrose, 200 μg/mL BSA, 2 mM ATP; pH 7.6 and filter sterilised). $1x10^7$ cells were added to 8 μg purified RNA in a pre-cooled 4mm cuvette and electroporated three times using an exponential protocol: 850 V, 25 uF capacitance, with 3 second pause between pulses. Electroporated cells were incubated at room temperature for 15 minutes, resuspended to 10 mL in complete media and incubated in a T75 flask at 28°C for a minimum of 24 hours. The protocol for electroporation of insect cells was adapted from (383).

2.4.2 Lipofectamine transfection with viral RNA

Lipofectamine transfection of insect cells with viral RNA was carried out where virus could not be produced by electroporation. Insect cells were seeded at $4x10^5$ cells per well in 6 well plates in 2 mL complete media until 70-80% confluent. 2 μg RNA per well was mixed with 250μL Optimem per well. 5 μL Lipofectamine2000 per well was mixed with 250 μL Optimem per well and incubated at room temperature for 5 minutes. The RNA mix was gently mixed with the lipofectamine mix and incubated at room temperature for 20 minutes. Cells were washed 1x in

PBS and media replaced with 2.5 mL Optimem per well. 500 µL per well of the RNA-liposome mix was added dropwise to each well and mixed by gentle swirling. Cells were incubated at 28°C until supernatant was harvested for plaque assay.

2.4.3 Plaque assay

BHK-21 cells were seeded at $3x10^5$ cells per well in 6 well plates and incubated in 2 mL complete media until 70-80% confluent. Virus stocks were thawed at room temperature and serial ten-fold dilutions were prepared in complete media. Cells were washed 1x in PBS before the addition of 200 µL diluted virus per well. Cells were rocked for 10 minutes at room temperature to ensure even coverage of the monolayer, incubated at 37°C for 50 minutes to allow virus entry and washed 1x in PBS to remove virus. 2 mL 0.8% MC overlay (50:50 1.6% MC in water to complete media) was added to each well. Plates were incubated for 48 hours until plaque formation. MC overlay was replaced with 10% formaldehyde and fixed at room temperature for 15 minutes. The fixative was removed and 2 mL per well crystal violet dye solution (0.25% crystal violet, 10% EtoH, 5 mM $CaCl_2$, 25 mM Tris in water) was used to stain each well for 15 minutes at room temperature before gentle washing in water. Plaques were counted in an appropriate dilution and titre calculated according to the following equation:

$$PFU/mL = \frac{Number\ of\ plaques}{(dilution\ factor\ 10^{-x}\ x\ volume\ added\ per\ well\ in\ mL)}$$

Multiplicity of infection, the number of infectious virions added per cell during an infection, can then be calculated using the following equation:

$$MOI = \frac{PFU\ per\ mL\ x\ mL\ used\ per\ well}{number\ of\ cells\ per\ well}$$

2.4.4 Virus infection

For one-step growth assays, mammalian and insect cells were seeded at $1x10^5$ cells per well in 24 well plates and incubated in 1 mL complete media until 70-80% confluent. Virus stocks were thawed at room temperature and dilutions were prepared in complete media to a final concentration of $5x10^5$ plaque-forming units (PFU)/mL. Cells were washed 1x in PBS before the addition of 200

μL diluted virus per well, giving a final multiplicity of infection (MOI) of 1. Cells were rocked for 10 minutes at room temperature to ensure even coverage of the monolayer, incubated at 37°C or 28°C for 50 minutes to allow virus entry and washed 1x in PBS to remove virus. 500 μL complete media was added per well and plates were incubated at 37°C or 28°C for the appropriate duration. The supernatant was harvested and titre calculated by plaque assay. Samples were stored at -80°C.

For passaging infectious virus, mammalian and insect cells were cultured in T75 flasks until 60% confluent. Monolayers were washed 1x in PBS. 1 mL passaged virus media was thawed at room temperature and added to each flask before rocking for 10 minutes at room temperature. The monolayer was incubated at 37°C or 28°C for 50 minutes before washing 1x in PBS. 10 mL complete media was added per flask and each monolayer was incubated for 48 hours at the appropriate temperature before supernatant was harvested and 1mL used for the subsequent passage. For growth of confirmed escape mutants, the incubation time was shortened to 24 hours and the monolayer was harvested in 6 mL Tri-Reagent (Sigma Aldrich). Samples were stored at -80°C.

For extraction of viral RNA for single-stranded qPCR, Huh7 cells were seeded at $1x10^5$ cells per well in 12 well plates and incubated in 1 mL complete media until 70-80% confluent. C6/36 cells were seeded at $6x10^5$ cells per well in 12 well plates and incubated in 2 mL complete media until 70-80% confluent. Cells were washed 1x in PBS before the addition of 200 μL diluted virus per well, giving a final MOI of 1. Cells were rocked for 10 minutes at room temperature to ensure even coverage of the monolayer, incubated at 37°C or 28°C for 50 minutes to allow virus entry and washed 1x in PBS to remove virus. 1mL complete media was added per well and plates were incubated at 37°C or 28°C for 24 hours. Supernatant was harvested after 24 hours. Monolayers were washed 1x in PBS before harvest in 500 μL Tri-Reagent (Sigma Aldrich) per well. RNA was extracted from these samples according to the manufacturer's instructions and resuspended in 20μL water, stored at -80°C. Positive and negative strand genome copy number was quantified via quantitative reverse transcription polymerase chain reaction (qRT-PCR) by M. Müller (University of Leeds).

2.5 Passaging methods

Passaging began with virus stocks generated via electroporation of BHK-21 cells with RNA transcribed *in vitro* from an infectious clone. Virus was harvested after 24 hours and titred in triplicate (P0). For each passage, a T75 flask of the appropriate cell type was cultured to 60% confluency before infection. The passage 1 (P1) monolayer was infected at MOI 1, passages thereafter were infected using 1 mL of the previous passage media. Following several passages, the supernatant was titred. If titres for WT and a mutant indicated possible reversion, infections were carried out at in 24 well plates at MOI 1 for 24 hours and the resulting supernatant titred. Where one-step growth curves indicated reversion of mutant phenotype, the genetic basis of reversion was investigated by RNA extraction and sequencing.

2.5.1 RNA extraction

Once phenotypic reversion was seen, confirmed escape mutants were used to infect a monolayer for 24 hours and harvested with 6 mL Tri-Reagent (Sigma Aldrich). RNA was extracted from these samples according to the manufacturer's instructions and resuspended in 20 μL water, stored at -80°C.

2.5.2 Reverse transcription

4 μL RNA sample extracted from each monolayer/9μL if extracted from infection supernatant was combined with 10μL RT buffer and 1μL enzyme mix from the High Capacity RNA-to-cDNA Kit (Life Technologies). Reaction volume was made up to 20 μL with H_2O, centrifuged briefly and incubated for 60 minutes at 37°C. Samples were then heated to 95°C for 5 min and stored at -20°C.

2.5.3 Polymerase Chain Reaction

PCR was used to amplify cDNA products for sequencing in order to identify mutations responsible for phenotypic reversion in passaged mutants. Reactions were performed in a thermocycler (ProFlex PCR system, Thermo Fisher Scientific).

For the first kb of the CHIKV genome, primers were designed for hemi-nested PCR. PCR reactions were performed in a total volume of 50 μL containing: 2 μL cDNA produced by reverse

transcription, 2 μL 10 mM forward primer, 2 μL 10 mM reverse primer, 0.6 μL 25 mM dNTPs, 5 μL 10x ThermoPol reaction buffer (NEB) and 0.5 μL Vent polymerase (NEB). Primary PCR was performed as follows: 95°C 5 minutes denaturation; 20 cycles of 95°C 30 seconds, 51°C 30 seconds, 72°C 1 minute elongation; 72°C 5 minutes extension. Secondary PCR reactions were made up as previously, containing 1 μl cDNA from the primary reaction. Secondary PCR was performed as follows: 95°C 5 minutes denaturation; 35 cycles of 95°C 30 seconds, 70°C 30 seconds, 72°C 1 minute elongation; 72°C 5 minutes extension.

For the entire non-structural region of the CHIKV genome, primers were designed to span the region in 1 kb overlapping sections. PCR reactions were performed in a total volume of 50 μL containing: 2 μL cDNA produced by reverse transcription, 2 μL 10 mM forward primer, 2 μL 10 mM reverse primer, 0.6 μL 25 mM dNTPs, 5 μL 10x ThermoPol reaction buffer (NEB) and 0.5 μL Vent polymerase (NEB). Primary PCR was performed as follows: 95°C 5 minutes denaturation; 30 cycles of 95°C 30 seconds, 55°C 30 seconds, 72°C 1 minute elongation; 72°C 5 minutes extension.

PCR products were stored at 4°C and amplification from the cDNA was confirmed by agarose gel electrophoresis. PCR products were purified using Wizard SV Gel and PCR Clean-Up System (Promega), eluting in 20 μL nuclease-free H_2O.

2.6 Statistical analysis

Statistical analysis was carried out using two-tailed Student's *t*-tests for unpaired samples of equal variance. *P* values of ≤0.05 (*), ≤0.01 (**), ≤0.001 (***), ≤0.0001 (****) were used to represent degrees of significance for each mutant compared to wild-type. A minimum of three independent biological repeats was performed for each experiment.

Chapter 3. RNA stem-loops within the 5′ region of the CHIKV genome enhance viral genome replication

3.1 Introduction

5′ RNA secondary structure in the untranslated regions and adjacent coding sequence of mRNA impacts the stability of the molecule and the efficiency of translation, for both host and viral transcripts (384,385). Furthermore, viral RNA secondary structures modulate packaging of viral genomes during infection (208,319) and may interfere with host immune mechanisms (353). Importantly, 5′ RNA structures play a role in the transition between translation and replication of RNA virus genomes and as a sampling mechanism, ensuring the integrity of the 5′ UTR in each template (339). Although replication of positive-sense RNA virus genomes initiates at the 3′ end of the molecule, 5′ RNA elements are demonstrably essential to this process during the lifecycle of several well-studied RNA viruses including various flaviviruses (332,333) and picornaviruses (334–336). For example, the RNA stem–loop SLA located in the 5′ UTR of flaviviruses such as DENV is an essential promoter for initiation of genome replication (386). Following interaction between SLA and the viral RNA-dependent RNA polymerase (RdRp), long-range interactions between the 5′ and 3′ ends stabilise circularisation of the virus genome and RdRp transfer to a 3′ promoter (298,299).

The stability of RNA secondary structure varies with several factors including temperature, salt concentration and protein binding (387,388). The temperature-sensitivity of RNA secondary structure is of particular relevance in the study of arthropod-borne RNA viruses. Arboviruses must maintain a productive replication cycle in mammalian and invertebrate hosts, each with vastly different temperatures, immune mechanisms and intracellular protein milieu. As a consequence, host-specific viral adaptations are highly constrained due to constant dichotomous selection pressure, as demonstrated in flaviviruses (389) and alphaviruses (390–392). Arboviral elements which enhance replication exclusively in a single host have been observed, including genome sequences (294) and RNA structures (350,393). For example, nsP3 of EEEV contains a stop codon which enhances viral replication in human cells, while being dispensable to replication in *Ae. albopictus* cells (394). Conversely, two stem-loops in the 3′ UTR of DENV, SL-I and SL-II, each

impact viral replication in mosquito cells but are functionally redundant in the human host (393). In addition to host-specific viral elements, multi-purpose pro-viral factors exist which utilise host-specific mechanisms. For example, a small hairpin structure (sHP) in the 3′ of the DENV genome is necessary for replication in both mosquito and human cells, enhancing replication via structure- and sequence-dependent mechanisms in human and mosquito cells respectively (294).

Relatively little is known about the structure and function of RNA secondary or tertiary structures within the CHIKV genome or their potential function as RNA replication elements. Therefore, parallels must be drawn with related alphaviruses which have been studied in greater detail. The 5′ terminal dinucleotide AU is highly conserved in alphaviruses including CHIKV (358) and is thought to play a role in binding of the RdRp to the 3′ end of the negative-sense genomic copies during positive-strand synthesis in conjunction with the adjacent stem-loop SL3. Other RNA structural elements investigated in alphaviruses include the packaging signal mapped in nsP2 (208), the 3′ replication element (289,351) and SL3 in the context of innate immune evasion (352,353).

A 51 nt conserved sequence element (CSE) consisting of two short stem–loops within the 5′ nsP1-encoding region was predicted to be highly conserved following multiple sequence alignment across a range of alphaviruses and SHAPE-constrained modelling in CHIKV, SINV and VEEV (13) (Fig 3.1). Reverse-genetic analysis has demonstrated that the 51 nt CSE enhances replication of SFV, SINV and VEEV (13,350,357,375,376). For SINV, disruption of the 51 nt CSE severely impaired replication in mosquito and avian cells (350,375), with a more moderate and variable effect on replication in mammalian cells (350,357). Mutation or deletion of the VEEV 51 nt CSE severely diminished or abolished RNA synthesis respectively in BHK-21 cells (352). However, another study indicated that deletion of either of the two stem–loops of the 51 nt CSE was well tolerated in BHK-21 or mosquito-derived cells, but deletion of both structures prevented VEEV replication (376).

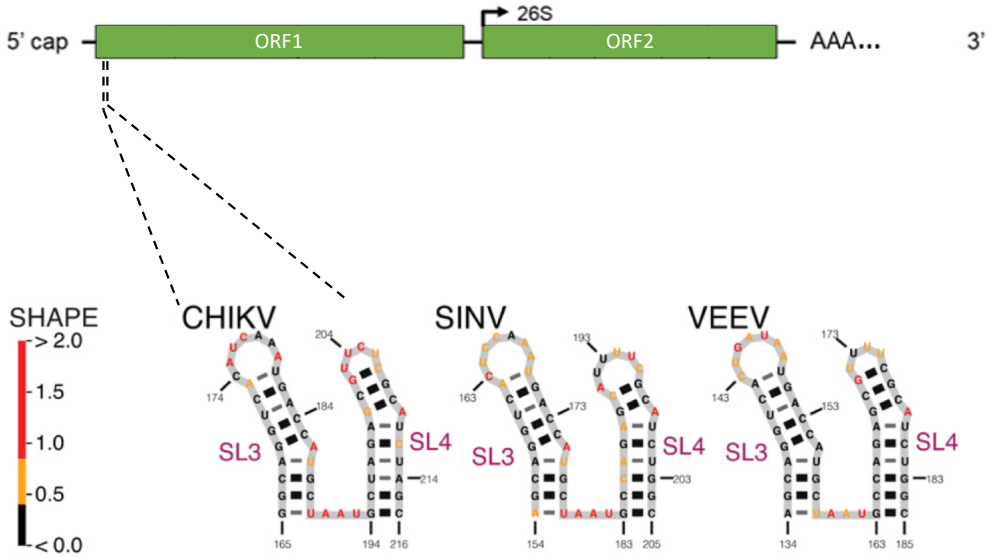


Figure 3.1: The 51 nt CSE (adapted from (13)). SHAPE map labelling as seen in Fig 1.22. A schematic of the CHIKV genome displays the position of the 51 nt CSE (dashed lines) within *nsp1*.

3.1.1 Preliminary data

In silico thermodynamic folding predictions and biochemical SHAPE mapping carried out by A. Tuplin (University of Leeds) suggested a highly ordered structured region at the 5′ of the CHIKV genome, spanning ~300 nt including the 5′ UTR and the 5′ coding sequence of *nsp1* (Fig 3.2). During recent outbreaks, synonymous site variability within this ~300 nt region was restricted in CHIKV isolates from the sera of infected individuals, indicating strong constraints on sequence variability, such as those imposed by functional RNA elements (395). In conjunction with phylogenetic analysis across the alphaviruses, which demonstrates the potential for structure conservation in this region via co-variation, these data suggest a function for the structured region during in the CHIKV lifecycle.

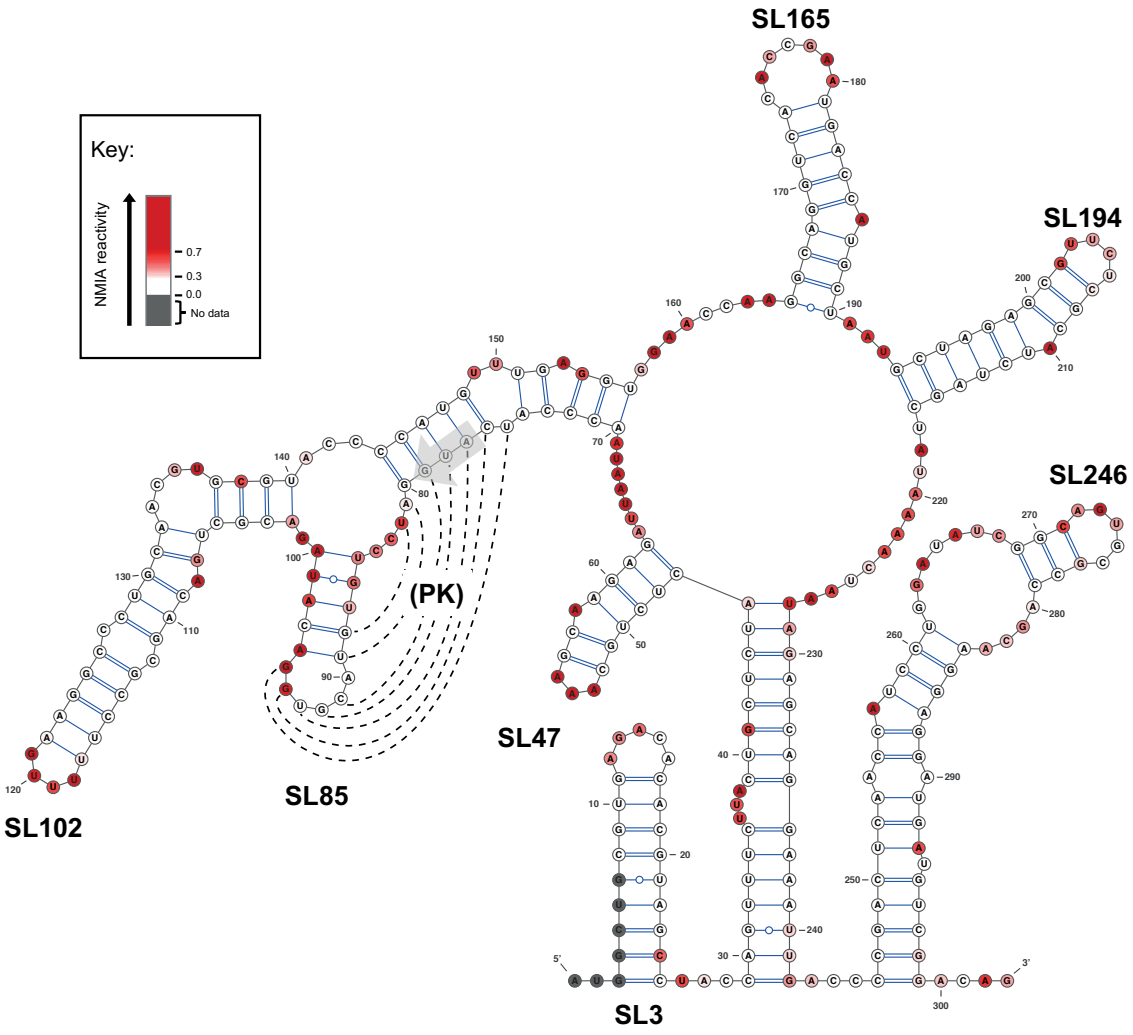


Figure 3.2: Thermodynamic folding model of CHIKV 5′ 300 nt overlaid with SHAPE reactivity data. 37°C SHAPE reactivities (n=3) for individual nucleotides overlaid onto a 37°C thermodynamically-derived model of RNA folding, generated using SHAPE-directed constraints. The AUG start codon of nsP1 is denoted by a grey arrow. SHAPE reactivities are shown as a heat map: grey indicates no data, white SHAPE reactivities between 0–0.3 and increasing intensities from light pink to dark red indicate increasing SHAPE reactivities, as denoted by the key. High reactivity (red) denotes unpaired nucleotides whereas low reactivity (white) denotes base-paired nucleotides. Predicted stem–loops are labelled SL3, SL47, SL85, SL102, SL165, SL194 and SL246. PK denotes a putative pseudoknot (PK) structure, where dotted lines represent potential base-pairing (A. Tuplin, University of Leeds).

SHAPE-constrained thermodynamic folding analysis predicted seven discrete stem-loop structures – two within the 5′ UTR (SL3 and SL47) and five within the adjacent nsP1-encoding region of ORF1 (SL85, SL102, SL165, SL194 and SL246) (Fig 3.2). SL3 has previously been predicted to form in CHIKV and demonstrated to mimic the methylated cap structure and function in viral immune evasion, avoiding recognition by IFIT-1 (371). Consequently, it was not investigated further. Whilst SL165 and SL194 correspond to the 51 nt nsP1 CSE investigated in VEEV and SINV, SL47, SL85, SL102 and SL246 are novel structures and their potential functions have not previously been investigated.

For six of the seven structures (SL3, SL47, SL102, SL165, SL194 and SL246) there was a high degree of concordance between the structural prediction and NMIA reactivity. Interestingly however, this was not the case for SL85 – located directly down-stream of the AUG start codon. SHAPE reactivities within SL85 indicated a high degree of exposure for nucleotides of the stem and low reactivity within nucleotides of the predicted terminal loop. Contradiction between the raw SHAPE reactivity values and SHAPE-constrained thermodynamic structure predictions may indicate that SL85 and its adjacent sequence are involved in dynamic secondary and higher-order RNA-RNA interactions. One such potential pseudoknot (PK) annotated in Figure 3.2 is explored further in Chapter 5. For initial reverse genetic analysis, SL85 was assumed to form according to the SHAPE-constrained prediction as shown in Figure 3.2.

The aim of the work described in this chapter was to investigate the phenotypic importance of RNA structures within the upstream region of the CHIKV genome; analysing stem-loop mutant phenotypes in human and mosquito cells and determining at which stage of the viral replication cycle these effects are exerted.

3.2 Results

3.2.1 Design of mutations for phenotypic analysis

In order to determine the phenotypic consequences of disrupting predicted RNA stem-loops mapped by SHAPE, a reverse genetic approach was taken. Individual stem-loops (SL47, SL85, SL102, SL165, SL194 and SL246) were disrupted via synonymous site mutagenesis (preliminary mutagenesis undertaken by D. Banda and A. Tuplin (University of Leeds)) (Fig 3.4A). Resulting phenotypic changes, to different stages of the virus replication cycle, were then assayed in human- and mosquito-derived cell lines.

A systematic and rational approach was taken to mutagenesis design, in which synonymous substitutions were designed to disrupt the predicted base-paired duplex stems of individual mutants in such a way that further synonymous mutations could then be incorporated to restore predicted base-pairing. The intention of this approach was that by phenotypic comparison of a stem-loop mutant with a disrupted duplex stem to one in which base-pairing was restored, phenotypic changes due to disruption in stem-loop structure could be distinguished from those due to alterations in primary synonymous nucleotide sequence. This approach precluded the use of codon shuffle algorithms which reshuffle codons to create mutant sequences of equivalent codon usage and dinucleotide frequency among other features. Although codon shuffling would disrupt the stem-loop structures while maintaining other characteristics of the region of interest, it would complicate the design of compensatory mutants which restore stem-loop base-pairing. Thus, mutations were designed to maintain synonymy and prioritise the possibility of structure reformation while preserving the degree of codon usage bias and dinucleotide frequencies as much as possible.

Furthermore, as compensatory mutations were designed based on a structural model of the positive-sense genomic copy of the CHIKV genome, rescue of wild-type (WT) phenotype by this approach would confirm the importance of a stem-loop in the positive-sense genomic strand - rather than due to unforeseen disruption of potential functional elements in the negative strand intermediate of the viral genome. The lack of structural similarity between the positive and negative strand can be seen in the UNAfold structural prediction of the 3′ 300nt of the negative

strand shown below, equivalent to the mutated region (Figs 3.2 & 3.3). The impact of mutations on this region is discussed further in later sections.

Stem-loop mutants were incorporated into the full-length CHIKV infectious clone (CHIKV_IC) and corresponding CHIKV sub-genomic replicon systems (Fig 3.4B) via site-directed mutagenesis (as previously described in 2.2.7). The disruption of targeted structures was corroborated by UNAfold prediction of each mutant, which can be observed in the appendix (Table 2). The outcome of CHIKV genome replication events was measured using a sub-genomic replicon, which encodes viral non-structural proteins nsP1-nsP4 and a Firefly luciferase reporter in place of ORF-2 (CHIKV_Rep).

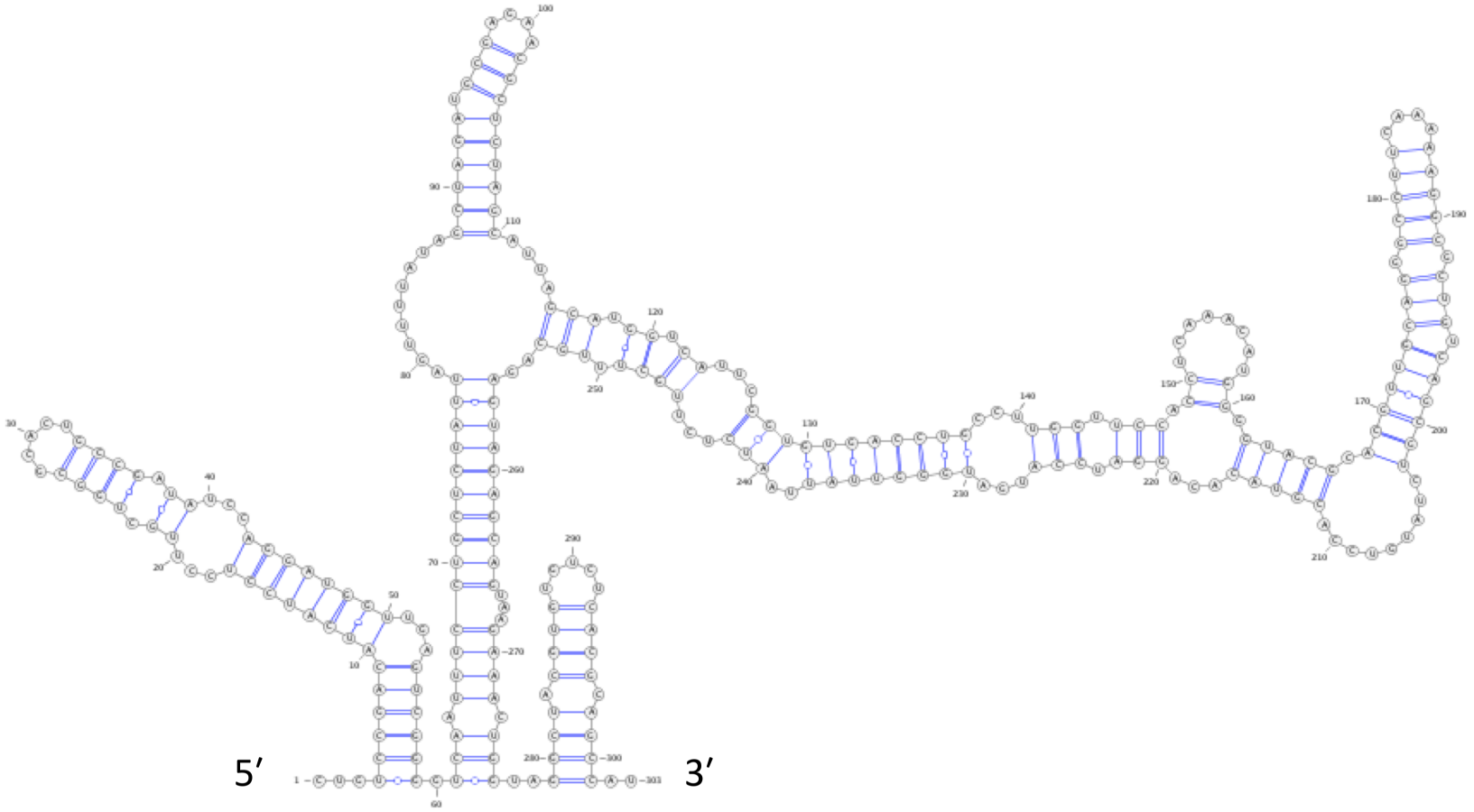


Figure 3.3: UNAfold prediction of RNA secondary structure for the 3′ 300nt of the CHIKV negative strand intermediate. Schematic representation of the secondary structure predicted to form in the negative strand of the CHIKV replication intermediate. Predictions were made using 37ºC parameters using UNAfold software.

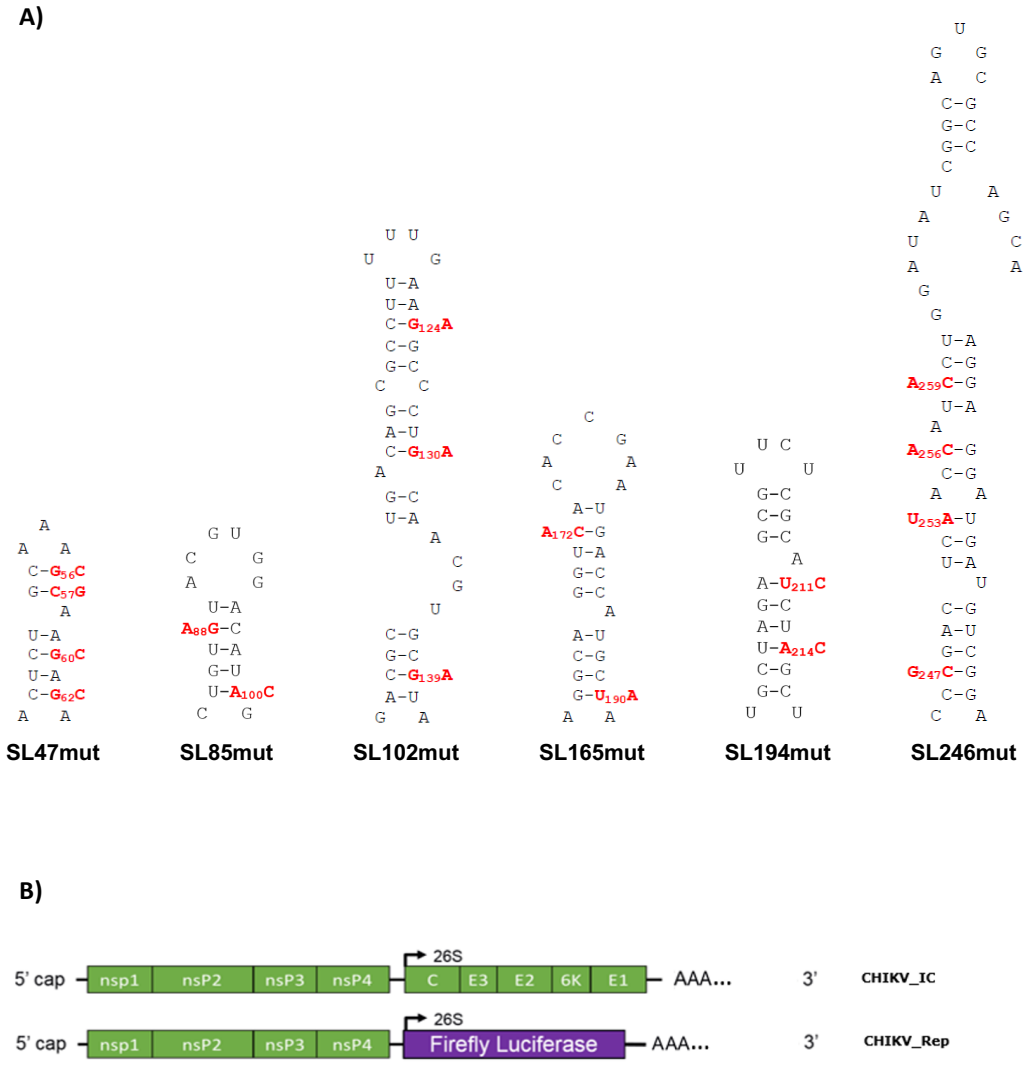


Figure 3.4: RNA secondary structure mutants and constructs. (**A**) Schematic representation of individual RNA stem–loops and associated mutations (red) designed to destabilise the base-pairing of the heteroduplex stem. All mutations in the nsp1-encoding region are synonymous. Mutations are labelled on the outside of the stem, while the wild-type sequence is displayed as part of the stem; for example, in SL85 $A_{88}G$ denotes a G>A mutation at nt 88 while $A_{100}C$ on the other side of the stem denotes an A>C mutation at nt 100. (**B**) Schematic representation of CHIKV infectious clone (CHIKV_IC) and sub-genomic replicon construct reporting genome replication through expression of Firefly luciferase (CHIKV_Rep). Non-structural proteins nsp1–4 are encoded by the first ORF. Structural proteins C, E1–3 and 6K are translated from a sub-genomic RNA, termed 26S RNA (black arrow), encoded by the second ORF.

3.2.2 SL47, SL85, SL102, SL165 and SL194 enhance CHIKV replication in human hepatocytes

To determine the importance of each mapped stem-loop during the CHIKV lifecycle, mutant viruses were generated by electroporation of BHK-21 cells with capped *in vitro* transcribed viral RNA (Fig 2.4). Huh7 cells were infected with either WT CHIKV or mutant virus with disrupted RNA structure at an MOI of 1 for 24 hours. In order to assess the impact of each set of mutations on CHIKV replication, the number of infectious virions released was determined by plaque assay (Fig 3.5A). CHIKV replication was significantly reduced upon disruption of SL47 in the 5′ UTR ($p<0.005$). Of the stem-loops in the coding sequence of nsP1, four had an impact on viral replication in human hepatocytes. SL85 and SL102 each significantly decreased viral titre when disrupted ($p<0.005$). SL165 and SL194, the equivalent of the 51 nt CSE investigated in VEEV and SINV, also significantly decreased CHIKV replication when disrupted ($p<0.005$ and $p=0.0001$ respectively). SL246 had no significant impact on CHIKV replication ($p=0.71$). This data suggests that SL47-SL194 enhance CHIKV replication in human hepatocytes. Interestingly, the reduction in viral replication was greater for *SL85mut, SL165mut* and *SL194mut* (~100-fold) than for *SL47mut* and *SL102mut* (~10-fold).

In order to confirm these results, intracellular viral genome copy number was determined by extraction of RNA from infected Huh7 cells. The qRT-PCR strategy, optimised and carried out by M. Müller (University of Leeds), measured the number of positive-sense genome copies and negative-sense replication intermediates in order to quantify viral genome replication. Disruption of SL47 in the 5′ UTR and SL85, SL102, SL165 and SL194 in the nsP1-encoding sequence significantly decreased the number of positive-sense CHIKV genome copies present in infected Huh7 cells ($p<0.005$) (Fig 3.5B). Similarly, disruption of SL47-SL194 significantly decreased the number of negative-sense CHIKV genome copies present in infected Huh7 cells ($p<0.005$) (Fig 3.5C). The greater reduction in viral titre for *SL165mut* and *SL194mut* (~100-fold) relative to *SL47mut* and *SL102mut* (~10-fold) was recapitulated at the level of positive sense genome copies. However, the number of positive sense genome copies for *SL85mut* (~25 fold) was closer to that of *SL102mut*. While disruption of SL47-SL194 significantly decreased the number of negative-sense CHIKV genome copies, differences between the mutants were not detected, possibly due to the smaller population of negative strand genome copies. Together, these results demonstrate that

SL47-SL194 enhance CHIKV replication in human hepatocytes; disruption of these stem-loops results in a decrease in viral genome replication and number of infectious virions released.

3.2.3 SL47 and SL246 enhance CHIKV replication in *Ae. albopictus* cells

The phenotypes presented in Figure 3.5 were observed in a human-derived cell line. As a mosquito-borne arbovirus, CHIKV must replicate in human and mosquito cells. Thus, it was important to investigate the impact of each set of mutations on CHIKV replication in *Ae. albopictus* cells. C6/36 cells were infected with either wild-type CHIKV or virus with disrupted RNA structure at an MOI of 1 for 24 hours. The resulting supernatant was titred by plaque assay (Fig 3.6A). CHIKV replication in C6/36 cells was significantly reduced by disruption of SL47 ($p<0.0001$). This suggests that SL47 enhances CHIKV replication in mosquitoes. Disruption of SL85, SL102, SL165 and SL194 had no effect on CHIKV replication in mosquito cells. Interestingly, CHIKV replication in C6/36 cells was significantly reduced by disruption of SL246 ($p<0.0005$), which had no impact in Huh7 cells, although the effect size was small. Strand-specific qRT-PCR was then carried out on RNA extracted from infected C6/36 cells to quantify the differences in viral genome replication between wild-type CHIKV and mutants exhibiting replication phenotypes. Disruption of SL47 in the 5′ UTR and SL246 in the nsP1-encoding sequence significantly decreased the number of both positive- and negative- strand CHIKV genome copies present in infected C6/36 cells ($p<0.002$) (Fig 3.6B & C). This data demonstrates that SL47 in the 5′ UTR enhances CHIKV replication in both human and mosquito hosts, while stem-loops SL85-SL246 in the coding sequence act in a host-specific manner.

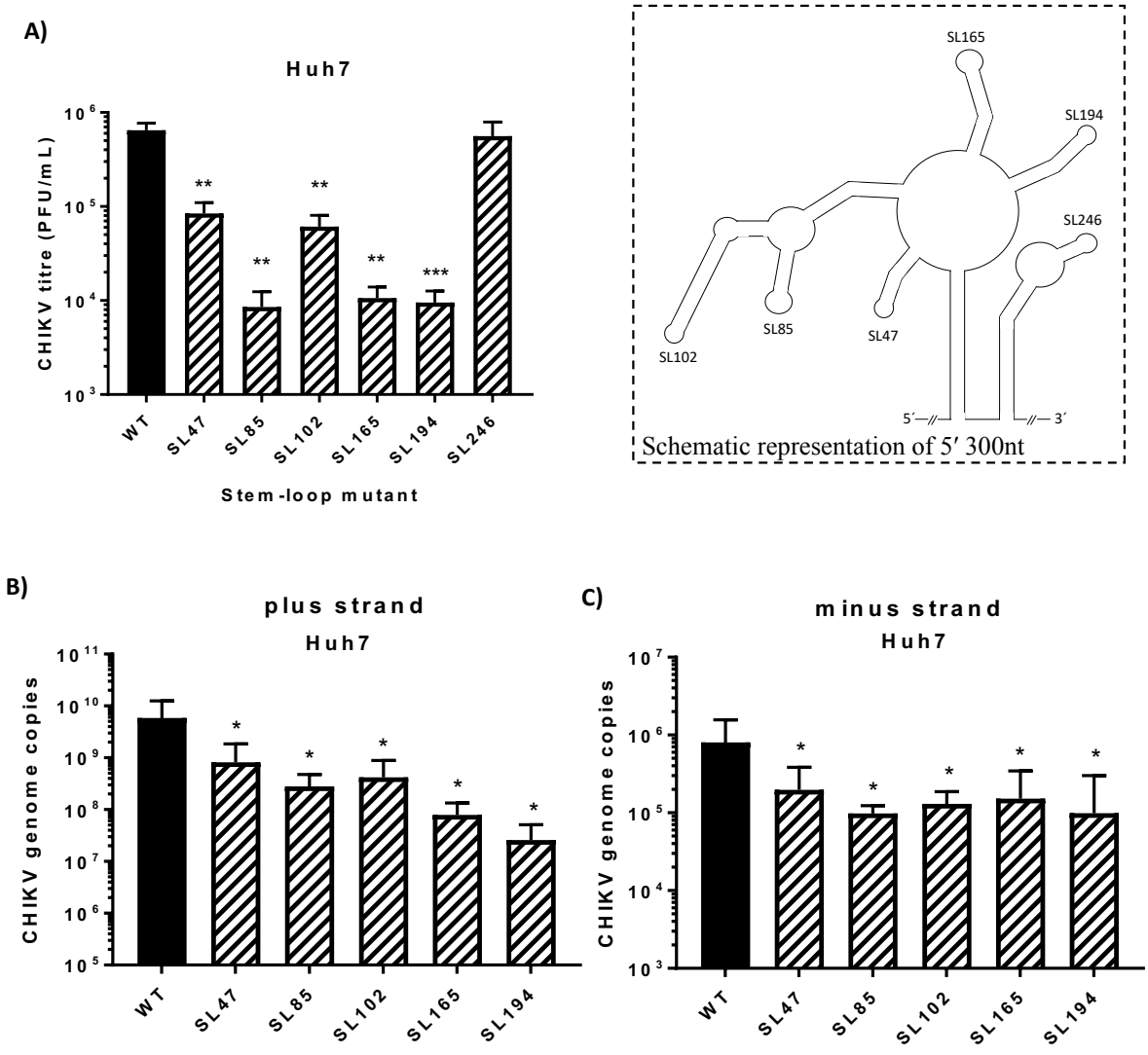


Figure 3.5: Stem-loops SL47-SL194 enhance viral replication in Huh7 cells. (**A**) Replication phenotype of WT CHIKV (black bar) compared to virus bearing mutations predicted to destabilise the heteroduplex stem RNA structures (hatched bars), in Huh7 human cells. Viral genome copy number is shown for (**B**) positive strands and (**C**) negative strand intermediates of the CHIKV genome following infection of Huh7 cells for 24 h with WT and destabilised mutants (n=3). qRT-PCR data from (B) and (C) obtained by M. Müller (University of Leeds). A schematic of the structured region (dashed box) is displayed for reference. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

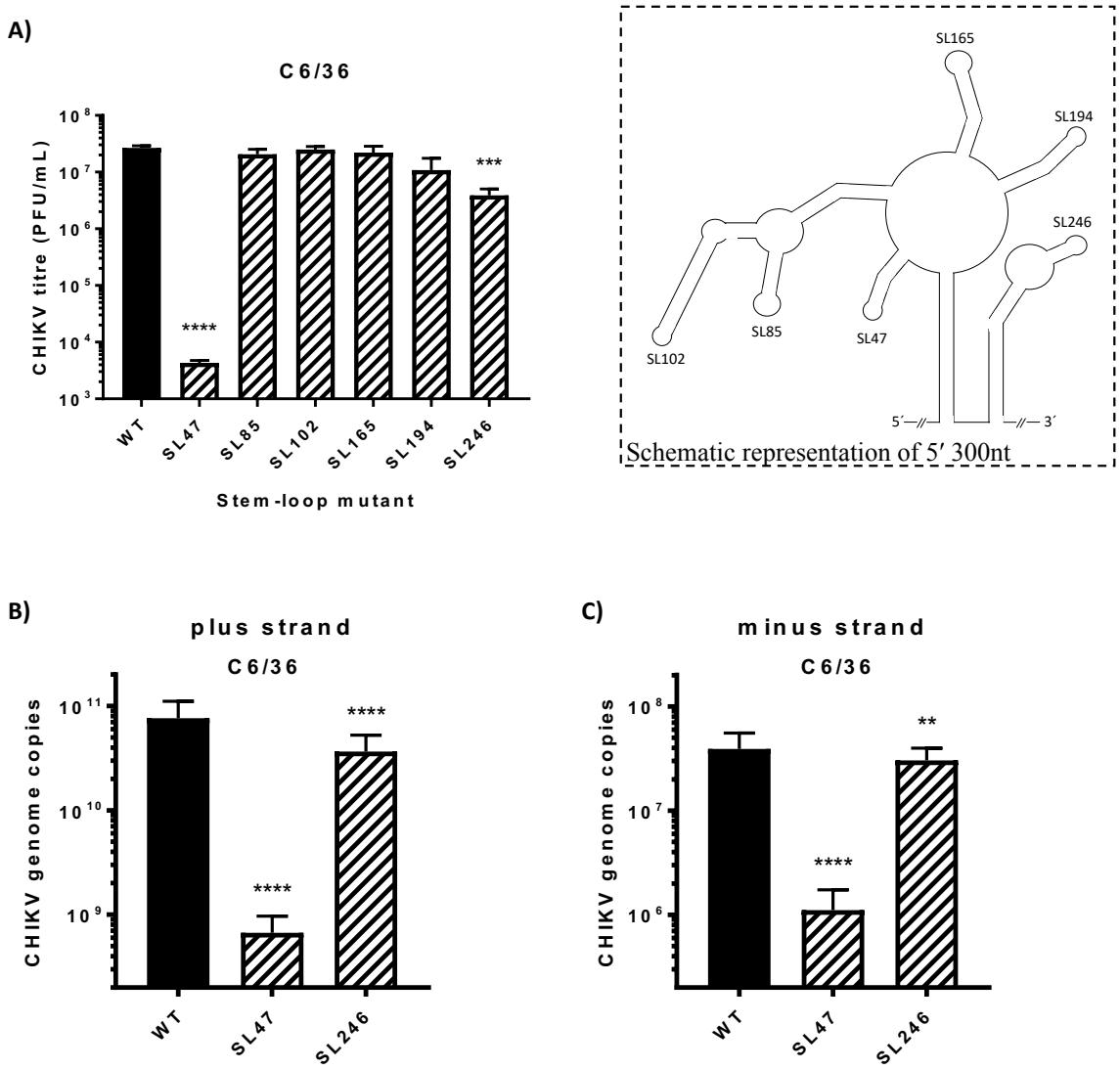


Figure 3.6: Stem-loops SL47 and SL246 enhance viral replication in *Ae. albopictus* cells. (**A**) Replication phenotype of WT CHIKV (black bar) compared to virus bearing mutations predicted to destabilise the heteroduplex stem RNA structures (hatched bars), in C6/36 *Ae. albopictus* cells. Viral genome copy number is shown for (**B**) positive strands and (**C**) negative strands of the CHIKV genome following infection of C6/36 cells for 24 h with WT and destabilised mutants (n=3). qRT-PCR data from (B) and (C) obtained by M. Müller (University of Leeds). A schematic of the structured region (dashed box) is displayed for reference. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

3.2.4 Host-specific replication phenotypes are not temperature-dependent

RNA structure is temperature-sensitive and structures may unfold at higher temperatures. Conversely, stable RNA secondary and tertiary structures may form at lower temperatures which would not be thermodynamically favourable at higher temperatures. Consequently, the effect of temperature on the stability of RNA secondary structure cannot be disregarded in the study of arbovirus replication, where host species vary in body temperature so widely. The different replication phenotypes observed for stem-loops in *nsp1* during infection of Huh7 and C6/36 cells may be a product of the temperature difference between these cells: 37 and 28ºC respectively. Temperature-dependent structural conformations could be exploited by CHIKV as a strategy to mediate viral replication in the two host cell types. Dynamic formation of two alternative structures may occur, with different functional elements suitable for each host. If SL85, SL102, SL165 or SL194 form differently at 28ºC, substitutions made to disrupt these structures as they are predicted to form at 37ºC would not target base-paired stems at the lower temperature. Alternatively, the structural mutants may be temperature-sensitive. Due to the relatively small number of substitutions per mutant, stabilisation of the heteroduplex stem at low temperatures is possible. The lack of phenotype for SL85, SL102, SL165 and SL194 in C6/36 cells could be due to the retention of wild-type structure at 28ºC.

Two experiments were conducted to investigate this. Firstly, SHAPE mapping was carried out by A. Tuplin (University of Leeds) on RNA folded at 28ºC to demonstrate that the wild-type RNA structured region exists at insect cell temperatures (396). The structured region shows remarkably little structural deviation at lower temperatures. Secondly, Huh7 cells were infected with mutant virus at the lower temperature, to determine whether the phenotypes seen at 37ºC were a result of the higher temperature (Fig 3.7A & B). At 24 hours post-infection, the phenotypes were consistent with those observed at 37ºC, although a significant improvement ($p<0.001$) was seen in SL85 replication. These results suggest that the effect of SL85 in Huh7 cells is partially temperature-dependent. The potential for dynamic structure formation around SL85 is explored in Chapter 5. The observed replication phenotypes at 28ºC argue against temperature-sensitivity of mutants in *nsp1*. Taken together, these experiments suggest that the effect on CHIKV replication seen in Huh7 cells for SL85, SL102, SL165 and SL194, which was not seen in C6/36 cells, is cell-type dependent rather than temperature-dependent, with the partial exception of SL85.

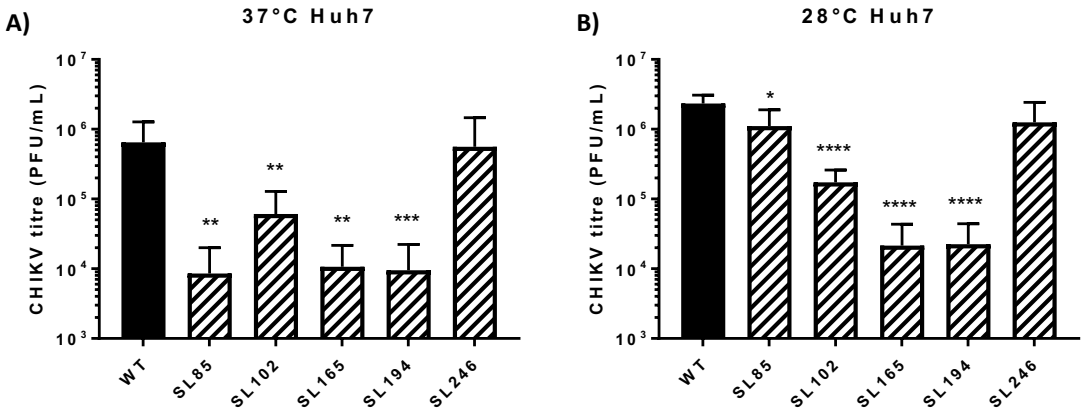


Figure 3.7: Stem-loops SL47-SL194 enhance viral replication in Huh7 cells in a temperature-independent manner. Replication phenotype of WT CHIKV (black bar) compared to virus bearing mutations predicted to destabilise the heteroduplex stem RNA structures (hatched bars), in Huh7 cells grown at (**A**) 37°C and (**B**) 28°C (n=3). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

3.2.5 SL47, SL85, SL102, SL165 and SL194 enhance CHIKV genome replication in human hepatocytes

Following the results of infection studies, whereby disruption of individual RNA structures in the genome were shown to inhibit CHIKV replication, further investigation was carried out to determine at which stage of the viral lifecycle they function. Strand-specific qRT-PCR suggested that CHIKV replication phenotypes were a result of inhibited genome replication. In order to examine the importance of each stem-loop during CHIKV genome replication, the effects of the stem-loop mutations on replication of a CHIKV sub-genomic replicon were measured over time. Mutations were cloned into a monoluciferase replicon CHIKV_Rep expressing the replicase polyprotein, with Firefly luciferase in place of the CHIKV structural proteins (Fig 3.4B). The sub-genomic replicon allows isolated study of genome replication in the absence of mechanisms of entry, packaging and egress; the replicon is transfected into cells and lacks capsid and envelope genes necessary for virion formation and egress.

Prior to the transfection of mutant sub-genomic replicon into cells, optimisation was carried out with WT replicon to determine the ideal time course for measurement of genome replication in several cell lines (n=1). The expansion of mammalian cell lines included those representative of human lung (A549), mouse muscle (C2C12) and human liver (Huh7 and Huh7.5) (Fig 3.8). As previously shown, A549 cells do not support CHIKV genome replication (397). C2C12 cells were permissive of WT CHIKV genome replication. However, high growth rates resulted in cell death from over-confluence by 48 hours post-transfection. Huh7 and Huh7.5 were permissive of CHIKV genome replication over 48 hours. Genome replication was therefore measured up to 24 hours post-transfection for mammalian cell lines: C2C12, Huh7 and Huh7.5.

Results from these sub-genomic replicon studies in Huh7 cells recapitulated the stem-loop mutant phenotypes observed during the infectious virus assays. Compared to wild-type, disruption of SL47 in the 5′ UTR significantly inhibited replication in Huh7 cells over 24 hours (Fig 3.9A). Likewise, disruption of SL85, SL102, SL165 and SL194 significantly inhibited replication in Huh7 cells (Fig 3.9B-E). Furthermore, in agreement with the infectious virus study, disruption of SL246 had no significant effect on sub-genomic replicon replication in Huh7 cells (Fig 3.9F). These results suggest that the enhancement of CHIKV replication observed during infectious virus studies occurs at the level of genome replication in Huh7 cells.

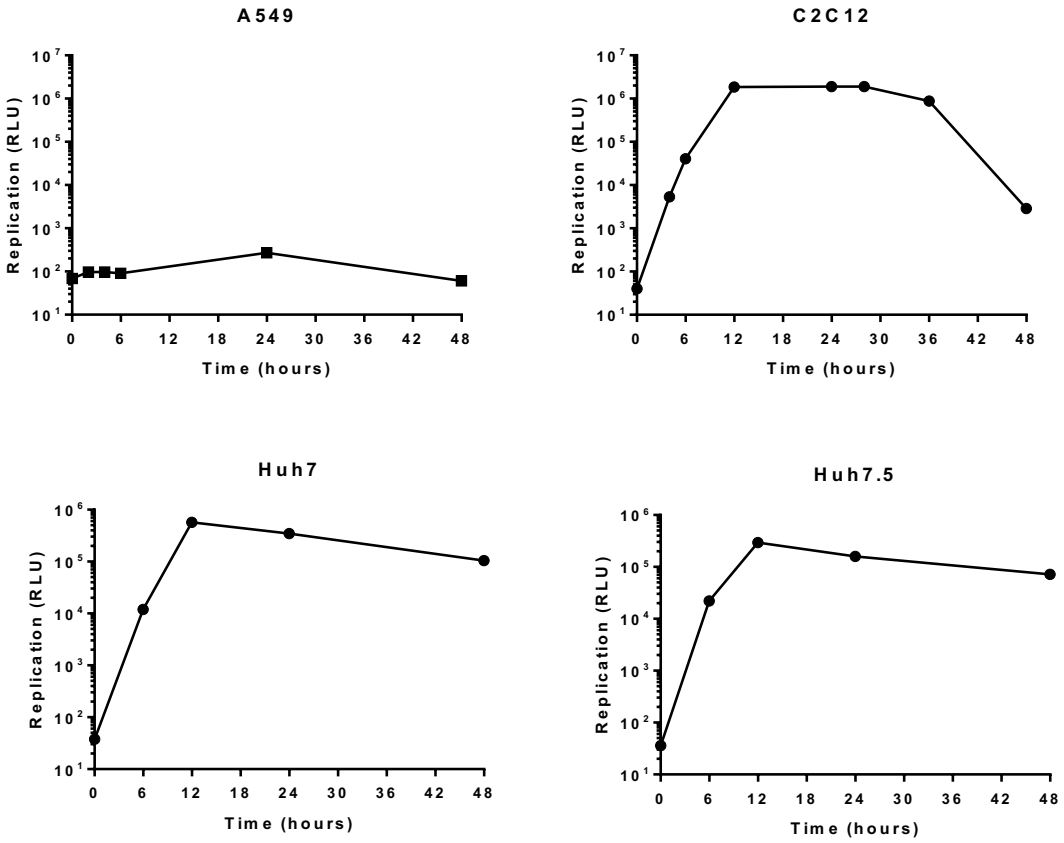


Figure 3.8: Optimisation of WT sub-genomic replicon replication in mammalian cells. Replication level of WT sub-genomic CHIKV replicon over 48 hours post-transfection in various mammalian cell lines: A549 human lung carcinoma, C2C12 murine myoblasts, Huh7 human hepatocellular carcinoma and Huh7.5 RIG-I knock-out Huh7 cell line (n=1).

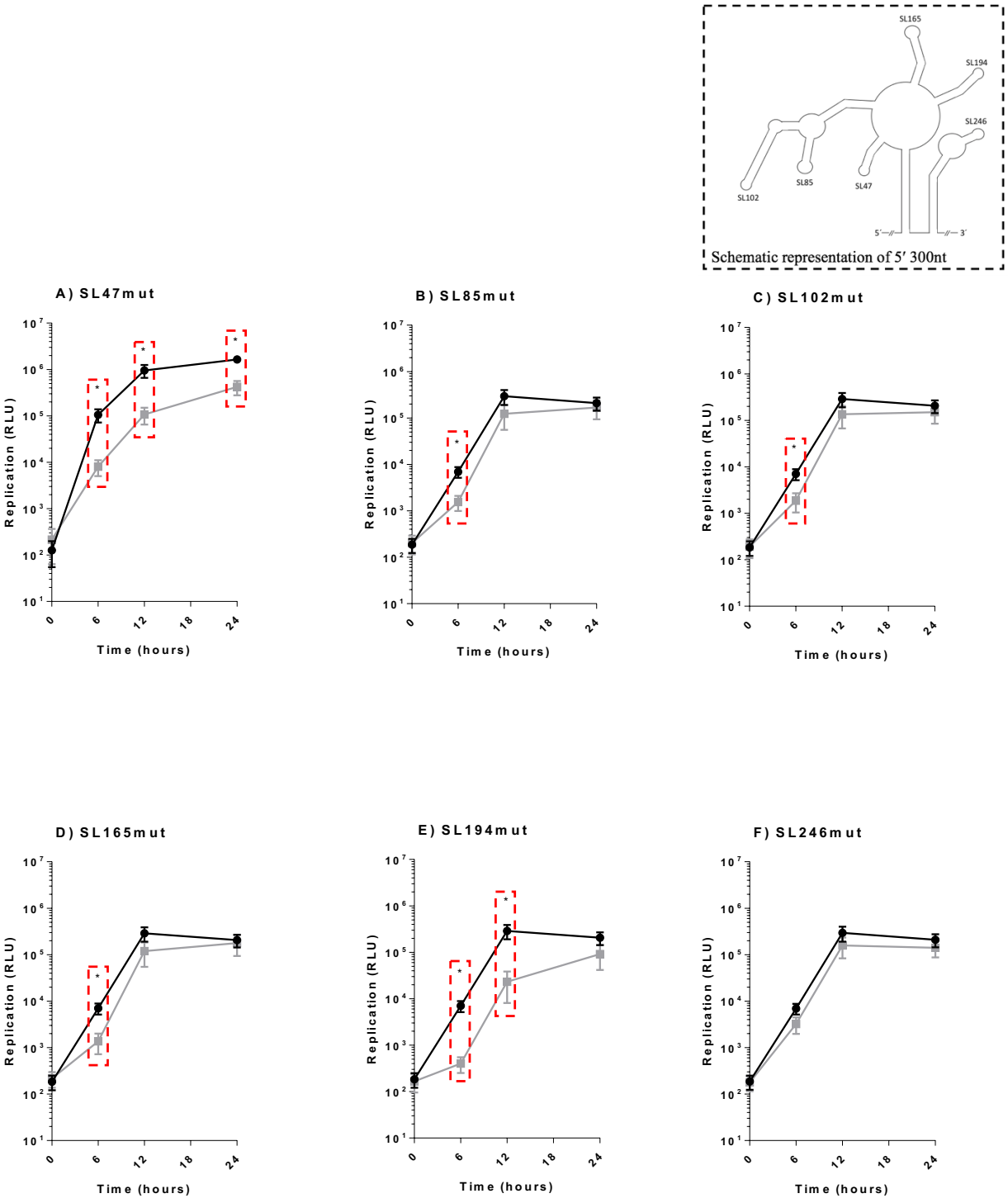


Figure 3.9: Stem-loops SL47-SL194 enhance viral genome replication in human hepatoma cells. Replication phenotype of sub-genomic CHIKV replicons in Huh7 human cells over 24 hours post-transfection. Replicons with wild-type RNA structure (black) compared to replicons bearing mutations predicted to destabilise the heteroduplex stem of RNA structures (grey) **(A–F).**

Significance ($p<0.05$) represented by red boxes for appropriate time-points (n=3). A schematic of the structured region (dashed box) is displayed for reference. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p\leq0.05$ (*). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

3.2.6 SL47, SL85, SL102, SL165 and SL194 enhance CHIKV genome replication in murine myoblasts

CHIKV replicates in several human tissues to high titre, including the liver. However, the pathology of infection is often associated with muscle and joint tissues. In order to investigate the function of the CHIKV structured region in muscle tissue, a mouse muscle progenitor cell line was used. C2C12 cells were transfected with sub-genomic replicon over 24 hours (Fig 3.10). The results re-capitulated those seen in hepatocytes; SL47, SL85, SL102, SL165 and SL194 enhanced CHIKV genome replication while SL246 had no significant role. These results suggest that the stem-loops function similarly in permissive mammalian tissues by enhancing CHIKV genome replication.

3.2.7 SL47 and SL246 enhance CHIKV genome replication in *Ae. albopictus* cells

As with mammalian cell lines, sub-genomic replicon assays were optimised with WT sub-genomic replicon in mosquito cell lines prior to mutant phenotyping (n=1). The *Ae. albopictus* cell lines C6/36 and U4.4 were transfected with WT replicon and measurements taken over 72 hours, longer than in mammalian cell lines due to the lifelong asymptomatic infection naturally acquired by mosquitoes. CHIKV genome replication did not peak until 24 hours post-transfection in *Ae. albopictus* cells, reaching a steady state in both cell lines (Fig 3.11). Measurements at 6 hours were below the threshold for accurate quantification. The luciferase signal was far higher in C6/36 cells than in U4.4 cells, likely due to the presence of Dicer-2 in U4.4 cells dampening CHIKV genome replication. Although the WT signal was above threshold in U4.4 cells, mutants displaying a replication phenotype may be indistinguishable from background signal. For these reasons, the C6/36 cell line was taken forward for measurement of genome replication up to 24 hours post-transfection. Results from sub-genomic replicon studies in C6/36 cells recapitulated the stem-loop mutant phenotypes observed in the infectious virus assays. Compared to wild-type, disruption of SL47 significantly inhibited replication in C6/36 cells (Fig 3.11A), while disruption of SL85, SL102, SL165 and SL194 had no significant effect (Fig 3.12B-E). As in the infectious virus study, disruption of SL246 significantly inhibited sub-genomic replicon replication in mosquito cells, although the effect size was minor (Fig 3.12F). These results confirm those of the infectious virus study, whereby SL47 in the 5′ UTR significantly impacts CHIKV genome replication in a host-independent manner, while stem-loops in the coding sequence of *nsp1* have host-specific roles.

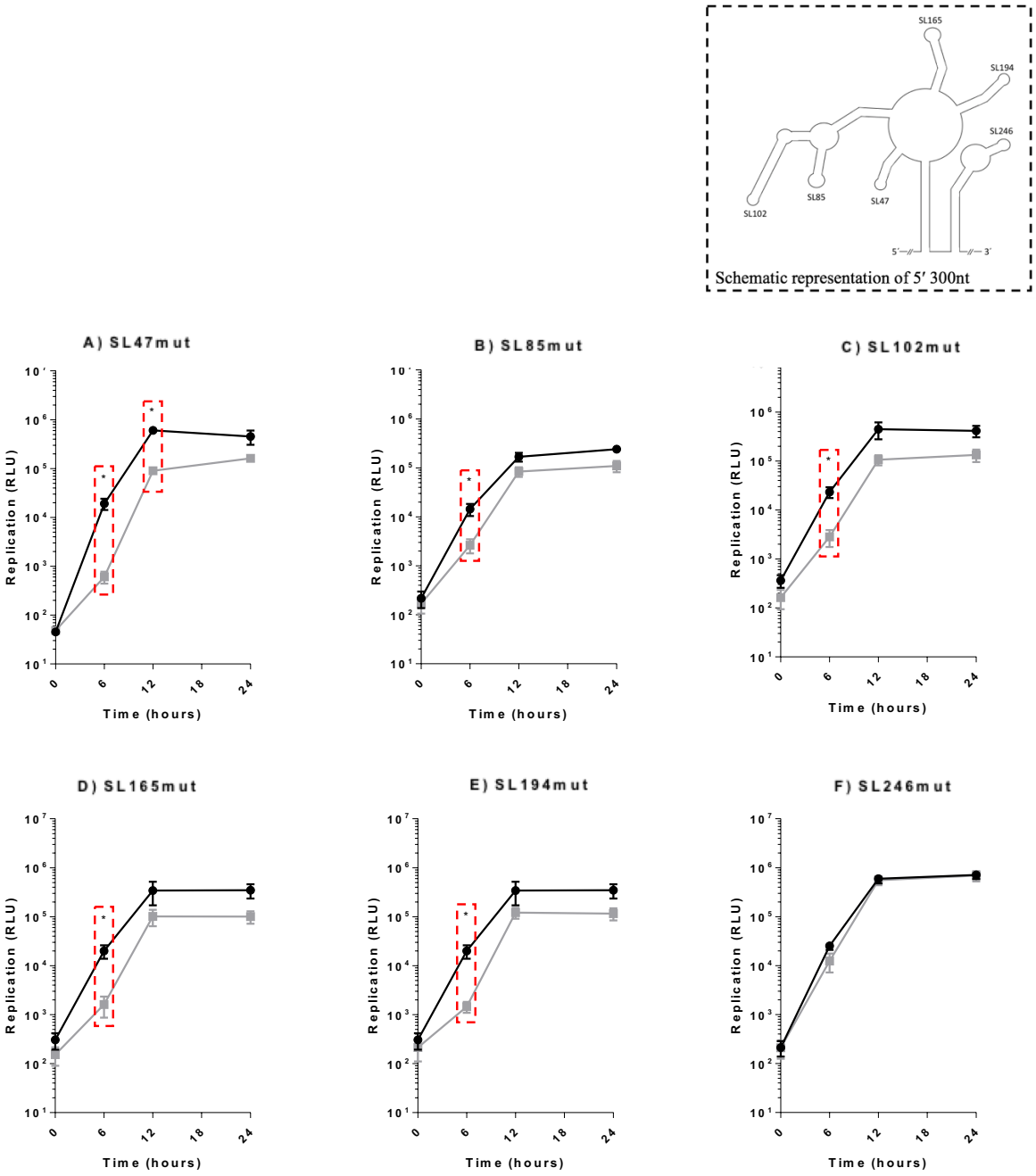


Figure 3.10: Stem-loops SL47-SL194 enhance viral genome replication in murine muscle progenitors. Replication phenotype of sub-genomic CHIKV replicons in C2C12 murine myoblast cells over 24 hours post-transfection. Replicons with wild-type RNA structure (black) compared to replicons bearing mutations predicted to destabilise the heteroduplex stem of RNA structures (grey) **(A–F).** Significance ($p<0.05$) represented by red boxes for appropriate time-points (n3). A schematic of the structured region (dashed box) is displayed for reference. * represents statistical

significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

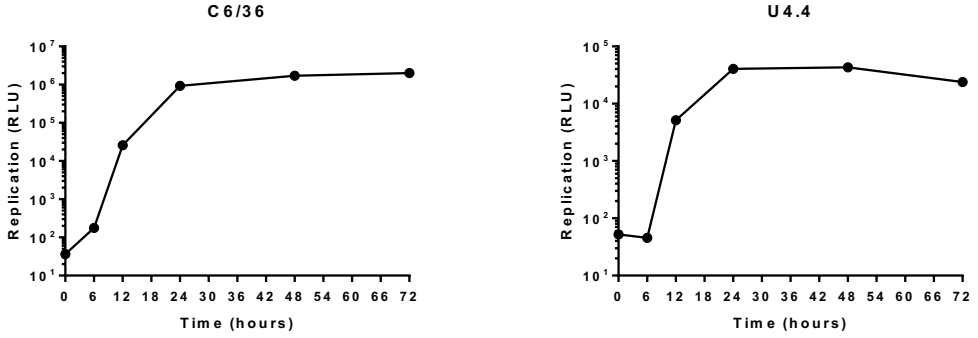


Figure 3.11: Optimisation of WT sub-genomic replicon replication in mosquito cells. Replication level of WT sub-genomic CHIKV replicon over 72 hours post-transfection in two *Ae. albopictus* cell lines: U4.4 mosquito embryonic cells and C6/36 cells, a Dicer-2 knock-out U4.4 cell line (n=1).

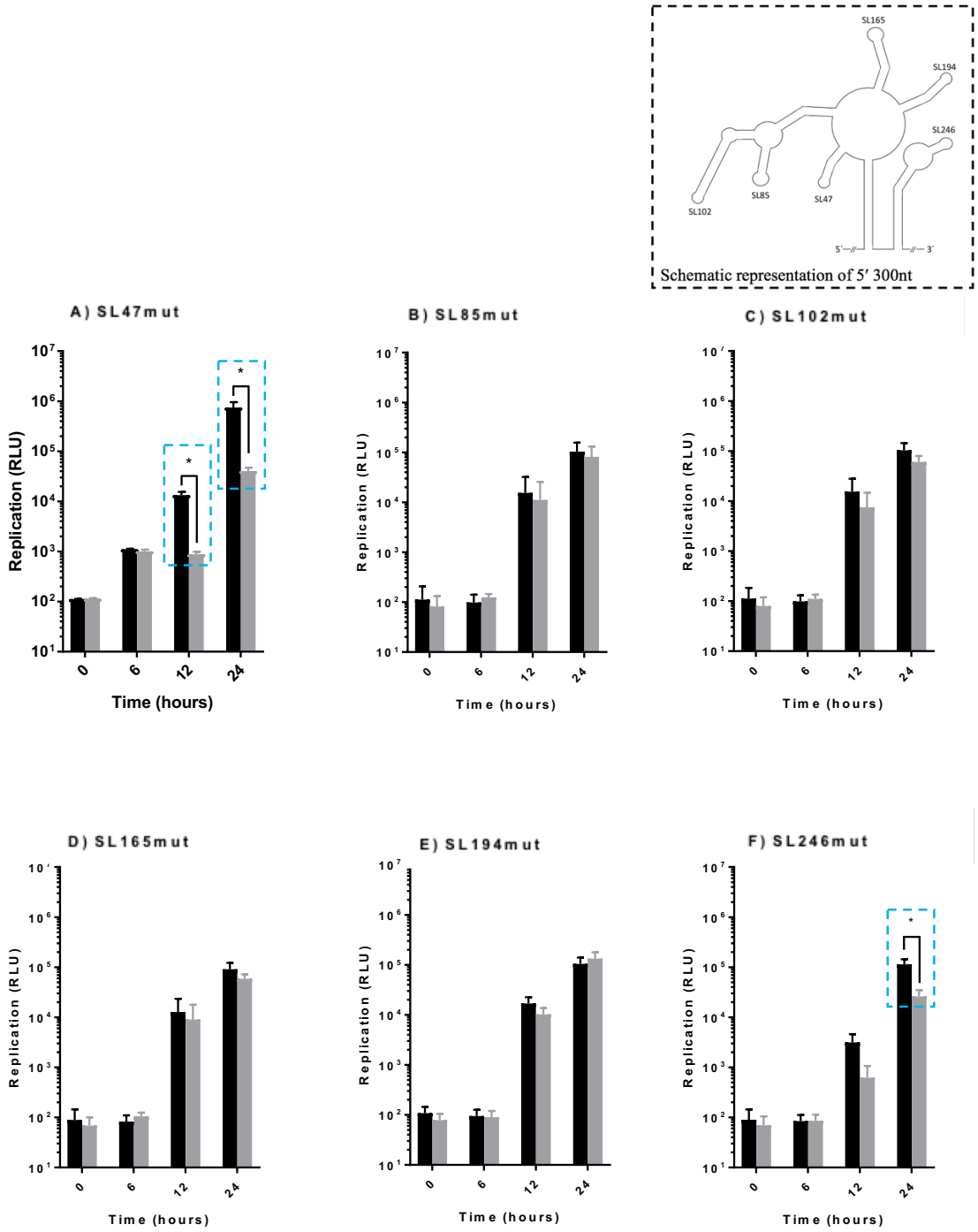


Figure 3.12: Stem-loops SL47 and SL246 enhance viral genome replication in *Ae. albopictus* cells. Replication phenotype of sub-genomic CHIKV replicons in C6/36 *Ae. albopictus* cells over 24 hours post-transfection. Replicons with wild-type RNA structure (black) compared to replicons bearing mutations predicted to destabilise the heteroduplex stem of RNA structures (grey) (**A–F**). Significance ($p<0.05$) represented by blue boxes for appropriate time-points (n3). A schematic of the structured region (dashed box) is displayed for reference. * represents statistical significance

for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

3.2.8 Disruption of CHIKV 5′ RNA structure does not impact translation

Whilst clearly demonstrating that the stem-loops function during CHIKV genome replication, the replication-competent sub-genomic replicon cannot distinguish between different stages of this process, such as initiation of transcription or translation of the ORF-1 non-structural polyprotein. Mutations inhibiting translation of ORF-1 would impair efficient production of replicase complexes, thereby inhibiting the down-stream process of genome replication. ORF-1 translation phenotypes were measured using a replication-deficient dual luciferase replicon (CHIKV_Rep(GDD>GAA)) produced by S. Bradley (University of Leeds), in which an additional reporter gene (Renilla luciferase) was incorporated into ORF-1 as a fusion within *nsp3* (Fig 3.13). CHIKV_Rep(GDD>GAA) undergoes initial translation to express the reporter in ORF-1 but subsequent transcription events cannot occur, allowing translation of the non-structural proteins to be studied in isolation from genome replication. In order to investigate the possibility that disruption of the stem-loops inhibited ORF-1 translation, the stem-loop mutations were incorporated into CHIKV_Rep(GDD>GAA). Using this construct, containing a mutation of the RdRp active site, the efficiency of ORF-1 translation alone was measured and compared between wild-type and stem-loop mutant transcripts, in the absence of subsequent genome replication. Translation of input sub-genomic replicon 5′-capped transcripts was measured at 6 hours for mammalian cell lines, Huh7 and C2C12, and 8 hours post-transfection into the mosquito cell line, C6/36. No significant differences in levels of translation were observed between the CHIKV_Rep(GDD>GAA) replication-deficient mutant encoding wild-type stem-loops to those encoding the mutant structures (Fig 3.14A-C). In combination with earlier results, from replication-competent sub-genomic replicon mutants, these results indicate that the individual stem-loops do not influence ORF-1 translation but rather function at the level of genome replication.

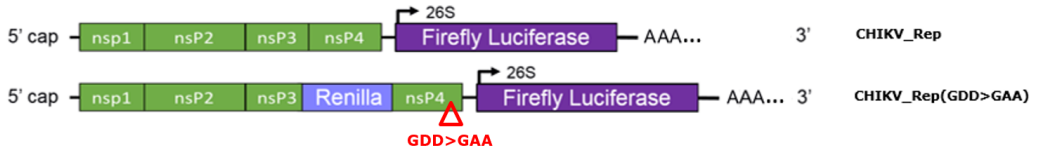


Figure 3.13: Translation reporter sub-genomic replicon

Schematic representation of CHIKV sub-genomic replicon constructs—reporting genome replication through expression of Firefly luciferase (CHIKV_Rep) or translation through expression of Renilla luciferase, in the context of a replication deficient RdRp GDD>GAA mutant (indicated by red triangle) (CHIKV_Rep(GDD>GAA)). CHIKV_Rep(GDD>GAA) was produced by S. Bradley (University of Leeds). Non-structural proteins nsp1–4 are encoded by the first ORF. Firefly luciferase is translated from a sub-genomic RNA, termed 26S RNA (black arrow), encoded by the second ORF.

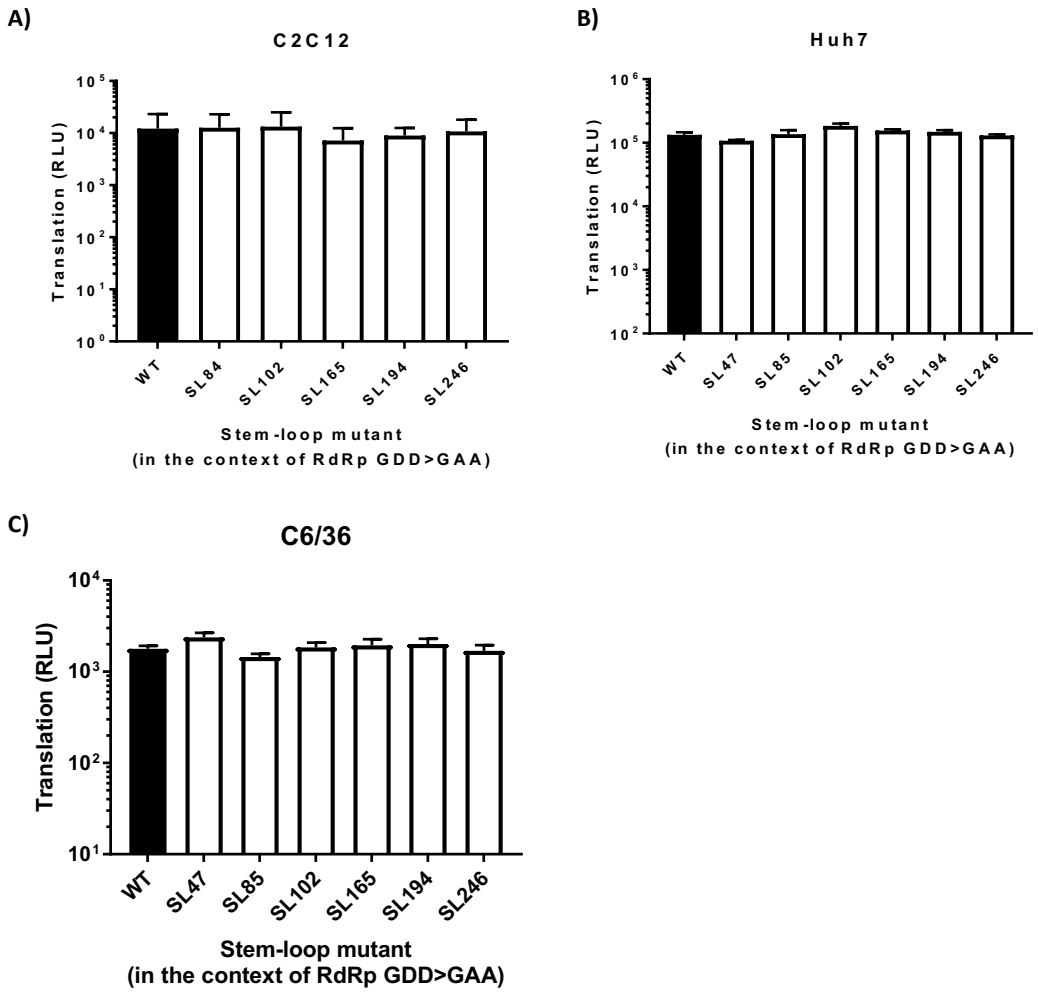


Figure 3.14: Translation is not impeded by disruption of CHIKV 5′ RNA structure. Translation phenotype in the context of replication-deficient sub-genomic CHIKV replicon system (CHIKV_Rep(GDD>GAA)) in (**A**) C2C12, (**B**) Huh7 and (**C**) C6/36 cells at 6 h post-transfection for mammalian cell lines and 8 h post-transfection for C6/36 cells (n=3)8. Replicons with wild-type RNA structure (black) are compared to replicons bearing mutations predicted to destabilise the heteroduplex stem of RNA structures (white). A schematic of the structured region (dashed box) is displayed for reference. Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

3.3 Discussion

Using reverse-genetic analysis in virus and replicon systems, this study demonstrates that the structured region of the CHIKV genome, comprising the 5′ UTR and adjacent nsP1-coding region (Fig 3.2), enhances CHIKV genome replication in human and *Ae. albopictus* cell lines. Novel RNA structural elements, within the positive-strand genomic copy of the CHIKV genome, were demonstrated to function by host-independent and dependent mechanisms (i.e. in either human- or mosquito-derived host cells or in both) during efficient replication of the virus genome. More specifically, RNA structures within the nsP1-encoding region are required for efficient CHIKV genome replication in a host-dependent manner, while SL47 in the 5′ UTR acts during genome replication in both human and mosquito host cells.

While the synonymous mutations incorporated into the structured region (Fig 3.2) were designed to disrupt base-pairing in each stem, the effect of substitution in the primary nucleotide sequence cannot be discounted. Mutations alter dinucleotide frequencies, codon usage and potential binding motifs. Increased frequency of the dinucleotides CpG and UpA in viral genomes has been demonstrated to inhibit replication of several single-stranded RNA viruses including influenza A virus (398) and the picornavirus, echovirus 7 (399). The impact of dinucleotide frequency is of particular interest in multi-host pathogens such as arboviruses, which infect highly divergent host organisms. *In silico* analysis across the *Flaviviridae* demonstrated that flaviviruses infecting vertebrate hosts exhibit underrepresentation of CpG and UpA, while insect-only viruses suppress UpA alone (400). The mutations incorporated by site-directed mutagenesis were designed to maintain CpG and UpA frequencies where possible (Fig 3.16). Thus, it is unlikely that the replication phenotypes observed were due to changes in dinucleotide frequency. Similarly, codon usage was taken into consideration during design of structural mutants (Fig 3.17). Substitution of codons with vastly different transfer RNA (tRNA) frequencies can impact translational efficiency (401). Extensive codon deoptimisation attenuated CHIKV in primate and mosquito cells, indicating that codon usage is a critical determinant in productive CHIKV infection (402). Although further study was necessary to formally disregard altered codon usage as responsible for the replication phenotypes observed in this study (Chapter 4), the absence of translation phenotypes for the stem-loop mutants in mammalian and mosquito cells strongly suggests that changes to codon usage had no significant impact (Fig 3.14). Finally, the primary RNA sequence

may contain binding motifs for trans-activating factors. The relative importance of primary nucleotide sequence and structure in CHIKV replication was explored in Chapter 4, via compensatory mutations in the stem which restored base-pairing and extensive mutagenesis of single-stranded terminal loops and bulges. Compensatory mutants also provide insight into the potential impact of dinucleotide frequency and codon usage bias on replication; if the phenotypes observed in this study result from either of these factors, phenotypic compensation would not occur in mutants which retain the original substitutions. An alternative approach to mutagenesis design utilizing codon shuffle algorithms could have been taken to ensure equivalent dinucleotide composition and codon usage frequencies in the disrupted mutants. While extremely useful for disrupting structures and binding motifs without impacting other sequence features, codon shuffle methods do not allow for restoration of structures in the same manner as the approach taken here.

Synonymous mutations were designed using the structural map of the 5′ of the positive-sense CHIKV genome but consideration must be given to the impact on structures in the corresponding region of the negative strand. The 3′ secondary structure of the negative strand intermediate differs greatly from that of the positive-sense genome due to determinants of structural stability such as stacking energy which vary by RNA context (403). UNAfold modelling predicts that the corresponding structure of the region of interest in the negative strand is very different to that of the positive strand (Fig 3.15). Several of the original mutations would not disrupt pairing in the negative strand intermediate. For example, the G>A mutations made in *SL102mut* which produce C:A pairs in the positive strand would represent G:U pairs in the corresponding negative strand. Given the replication defect exhibited by *SL102mut,* while the negative strand element could still form in the context of these mutations, it is clear that SL102 in the positive strand is responsible for enhancement of genome replication. Phylogenetic analysis demonstrates strong sequence and structure conservation in this region of the genome but does not distinguish between conservation of sequence elements in the positive or negative strands of the CHIKV genome. However, the frequency of G:U co-variation, which would produce disruptive C:A pairs in the negative strand, suggests structural conservation in the positive rather than the negative strand of the genomic RNA. Further consideration of the impact of mutations on negative strand secondary structure is given in Chapter 4.

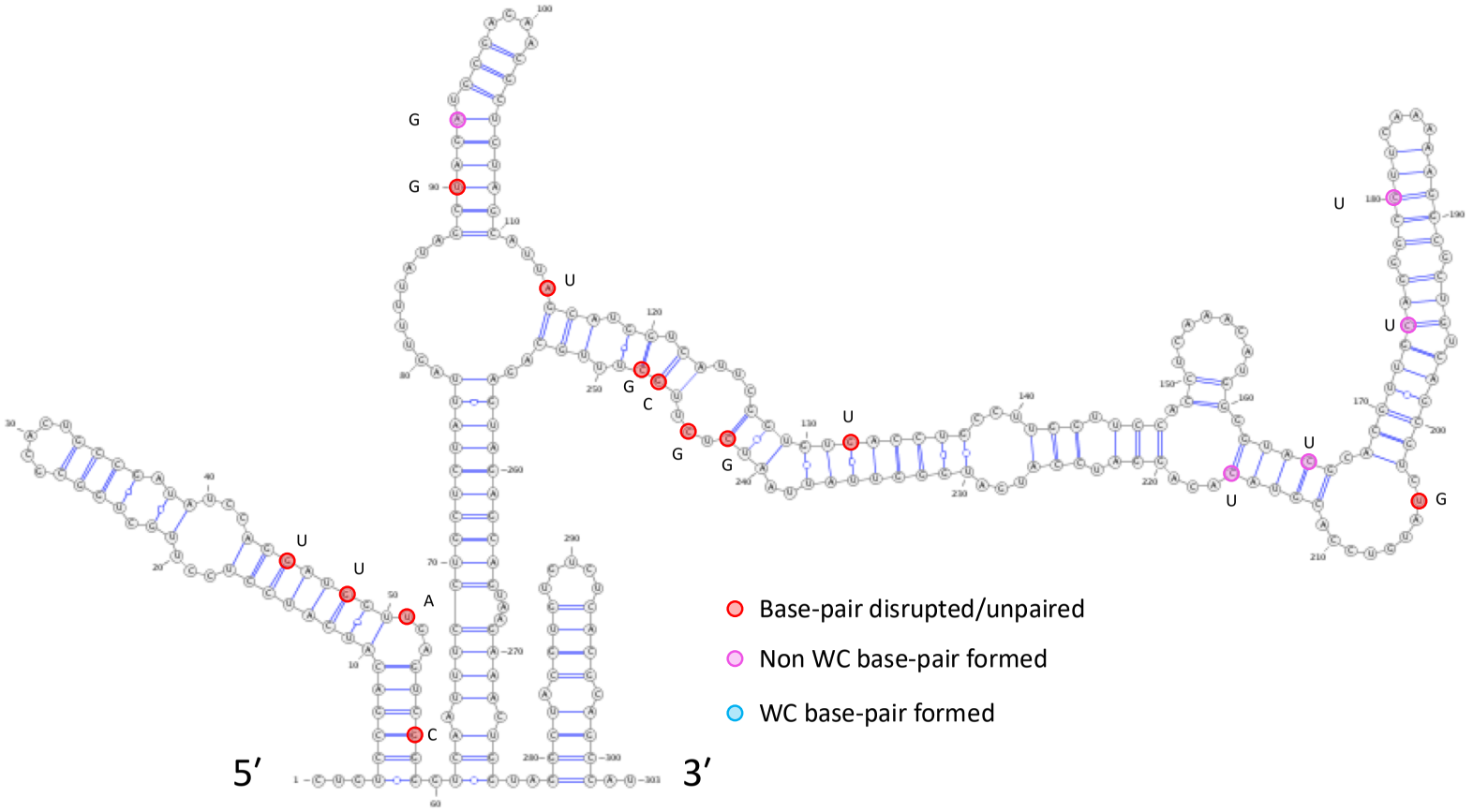


Figure 3.15: Types of base-pair formed by mutation in the negative strand. UNAfold prediction of secondary structure in the 3′ 300nt of the wild-type negative strand intermediate with mutations denoted by coloured circles which depict the type of alteration made in the negative strand structure: red circles indicate base-pairing absent or disrupted, pink circles indicate a non-Watson-Crick base-pair formed by an introduced mutation. The structure is displayed 5′ to 3′, whereby the 5′-most structure corresponds to SL246 (the 3′-most structure in the positive strand). Base-pairing in the predicted wild-type structure is denoted by blue lines: one line for A:U pairs, two lines for G:C pairs and a line with a white circle for G:U pairs.

Mutation of stem-loop 47 (SL47) in the 5′ UTR of the CHIKV genome reduced production of infectious virus, viral genome replication and replication of a sub-genomic replicon in human- and *Ae. albopictus*-derived cell lines. The dual-host replication phenotype exhibited by *SL47mut* may be explained by a host-independent mechanism of replication enhancement, utilising host factors present in both of the disparate cellular environments. Host-independent enhancers of arboviral genome replication have been demonstrated, for example, increasing the fidelity of the CHIKV nsP4 RdRp by a single mutation C483Y negatively impacts viral replication in mosquitoes and mice indiscriminately (404). Alternatively, SL47 may enhance replication in each host via a distinct pathway as has been demonstrated for sHP in the DENV 3′ UTR. The sequence of sHP is necessary for efficient replication of the DENV genome in mosquito cell lines, whereas the structure of the stem-loop alone is necessary to achieve this enhancement in human cells, indicating different mechanisms by which sHP impacts replication in each host (294). While clearly demonstrating a function for SL47 in CHIKV genome replication, this phenotypic study does not delve into the mechanism behind enhancement of replication. Further studies provide mechanistic insight by identifying the structure and sequence requirements for SL47-mediated enhancement of CHIKV genome replication in each host (Chapter 4).

Stem-loop mutant	CpG	UpA
SL47mut	0	0
SL85mut	1	1
SL102mut	-2	2
SL165mut	0	0
SL194mut	1	-2
SL246mut	0	2

Figure 3.16: Changes in CpG and UpA dinucleotide frequency in disrupted structure mutants. Net change in CpG and UpA dinucleotide frequencies in each of the stem-loop mutants, expressed as the number of each dinucleotide pair introduced (positive) or removed (negative) by the mutations displayed in Fig 3.4A.

	Percentage change in codon efficiency	
	Human	*D. melanogaster*
SL47mut	-	-
SL85mut	-2	5
SL102mut	-13	-2
SL165mut	-8	1
SL194mut	16	-4
SL246mut	-15	-3

Figure 3.17: Changes in codon efficiency in disrupted structure mutants. Net change in codon efficiency for each stem-loop mutant in human- and *D. melanogaster*-derived cells. Changes to codon efficiency are displayed as a relative percentage increase (positive) or decrease (negative) in codon efficiency in each organism upon mutation. Each mutant contains a minimum of two substitutions relative to wild-type and thus each percentage is an average of the change induced by all substitutions per mutant. Standard deviation for each average was equal to or less than 10%. Codon efficiency is measured as codon frequency in the genome of each organism, as drawn from the codon frequency tool from Genscript (405). SL47 is in the untranslated region, thus codon usage is null.

Figure 3.18: Nucleotide alignment of CHIKV ECSA, WA and Asian genotypes with other members of the *Alphavirus* genus - comparing the first 320 nt of the genome. Predicted stem-loops for CHIKV and ONNV (green boxes) are shown, along with the putative pseudoknot structure for Pk1 (PK - interacting nucleotides in red boxes). Alignment gaps are indicated by dashes.

SL47 represents a previously unreported RNA replication element which is highly conserved across CHIKV genotypes and closely related viruses such as O′nyong nyong virus and Mayaro virus, although variation is observed in the single-stranded terminal loop sequence (Fig 3.18). Within more divergent members of the Semliki Forest Complex, including Ross river virus (RRV) and SFV, covariant and semi-covariant substitutions (the latter involving non-canonical G:U interactions) are observed within the base-paired stem, maintaining broadly homologous structures (Fig 6.1). It has been detailed in previous studies that the length of 5′ UTR region varies widely across the alphavirus genus (353) and more divergent alphaviruses such as VEEV, SINV and Eilat virus, lack the sequence domain within the 5′ UTR responsible for SL47 formation. Previous studies have demonstrated that the 5′ UTR sequences of SFV and SINV are involved in controlling template specificity for initiation of negative-strand synthesis (357). Since SL47 is not conserved across the alphavirus genus but is required for efficient CHIKV genome replication, it can be speculated that SL47 plays a role in CHIKV replicase template specificity.

Disruption of SL85, SL102, SL165 and SL194 in the nsP1-encoding region of the CHIKV genome reduced viral replication, viral genome replication and replication of a sub-genomic replicon in mammalian- but not *Ae. albopictus*-derived cell lines. These results suggest host-specific mechanisms of action for stem-loops in *nsp1* during enhancement of CHIKV genome replication in mammalian cells. Immediately downstream of the AUG start codon, SHAPE-constrained thermodynamic analysis demonstrated that the RNA structure of CHIKV differs from that previously observed for other alphaviruses (13). For instance, in SINV this region forms a single, long RNA element – deletion of which does not inhibit virus RNA replication (350). In this study, the homologous region of the CHIKV genome was demonstrated to contain two distinct RNA replication elements (SL85 and SL102), both of which are involved in efficient genome replication in human-derived cells but not those from *Ae. albopictus*. Interestingly, while the *in silico* thermodynamically predicted structure, SHAPE mapping and reverse genetic analysis were in close agreement for SL102, this was not the case for SL85 (Fig 3.2). Phenotypic analysis demonstrated that SL85 plays a role in CHIKV genome replication in human-derived cells. However, SHAPE mapping of SL85 was inconsistent with the predicted structure, suggesting that this region of the genome may be structurally dynamic and able to form alternative interactions. One such potential interaction is a pseudoknot between complementary sequence motifs in the

apical region of SL85 and an adjacent upstream region of the genome overlapping the AUG start codon, which would destabilise base-pairing within the stem of SL85. Further work was needed to determine whether SL85 forms as predicted by SHAPE-constrained thermodynamic analysis or as may be expected from SHAPE reactivity (Chapter 5). Compensatory mutations which reformed the stem of SL85 as it would be predicted to form by constrained thermodynamic prediction, shed light on the formation of the stem-loop *in vivo* (Chapter 4). Further to this, investigation of the spatio-temporal dynamics of SL85 formation may reveal at which stage of the viral lifecycle SL85 exerts its effects and the location of its formation within the cell i.e in replicative complexes or in the cytosol.

Downstream of SL85 and SL102 lies the alphaviral 51 nt conserved sequence element (CSE), comprised of SL165 and SL194 in CHIKV. The primary nucleotide sequence and structure of the nsP1 51 nt CSE has previously been shown to be highly conserved between divergent alphaviruses (Fig 3.1) (13). Although studies investigating the 51 nt CSE have not been previously published for CHIKV, published studies in VEEV and SINV have demonstrated that this element enhances virus replication in both mosquito- and mammalian-derived cells, via initiation of negative strand replication (357,375). It should be noted that several studies have provided conflicting reports of the importance of the 51 nt CSE in mammalian cells, with individual or dual disruption being tolerated during some studies (350,358). By comparison, this study demonstrates that the 51 nt CSE of CHIKV acts in a host-dependent manner – enhancing replication in human cells, while having no significant effect on viral replication in *Ae. albopictus*-derived cells. Given the high level of conservation of the 51 nt CSE, it has been hypothesised that it functions through a conserved mechanism across divergent members of the *Alphavirus* genus. However, differences demonstrated in this study between the host specificity of the CSE in CHIKV and those in previously published studies in VEEV and SINV, suggest that the function or mechanism of action may be more divergent than previously recognised.

Within this study SL246 was the only RNA replication element, other than SL47, which was essential for efficient CHIKV genome replication in mosquito-derived cells. However, unlike SL47, it was not required for efficient replication in human cells. Furthermore, SL246 is much more highly conserved across divergent alphaviruses than SL47, with *in silico* UNAfold analysis

indicating that homologous structures are capable of forming within divergent viruses such as SINV, VEEV and Eilat (Fig 3.19). The conservation of this novel RNA replication element within the alphavirus genus suggests that SL246 functions within the positive-sense genomic RNA molecule during efficient genome replication in several alphaviruses.

Each of the stem-loop mutants *SL47mut-SL246mut* exhibited a replication defect during this study, but the relative reduction in CHIKV titre varied between mutants. A ten-fold greater reduction in CHIKV titre was observed for *SL85mut, SL165mut* and *SL194mut* compared to *SL47mut* and *SL102mut* in Huh7 cells, and a thousand-fold greater reduction for *SL47mut* compared to *SL246mut* in C6/36 cells (Figs 3.5 & 3.6). This variation in magnitude of replication defect could be the result of individual, non-redundant mechanisms of action for each stem-loop, for example stem-loop-specific *trans*-activating factors or scaffold proteins. Alternatively, stem-loops displaying similar replication phenotypes may be involved in a cooperative interaction. Another possibility is that the stem-loops form part of a larger region of secondary or tertiary structure which is disrupted to differing degrees in each stem-loop mutant. For example, disruption of SL165 in the centre of a larger structure may cause a greater degree of disruption than *SL47mut* at the 5′ terminus. Pairwise and combinatorial mutagenesis of the stem-loops may provide insight into interactions and redundancy between the stem-loops and the possibility of a larger structure. This work was begun in Chapter 5.

In addition to differing effect sizes between stem-loop mutants, luciferase assays produced a smaller effect size than that seen in viral one-step growth assays. There are several possible reasons for this, which can be broadly divided into biases of the two systems and true biological effects. Firstly, the sub-genomic replicon CHIKV_Rep does not contain the structural protein-coding ORF sequence. There may be RNA structures within this region of the genome which interact with the stem-loops directly or through recruitment of binding factors. Secondly, the products of CHIKV_Rep replication cannot egress and enter untransfected cells whereas CHIKV_IC can complete multiple rounds of reinfection. On average, CHIKV can complete three rounds of infection in 24 hours. Thus, it is likely that any difference in replication between the wild-type and mutant RNA would be amplified in viral assays compared to replicon assays. Thirdly, the reduction in initial RNA synthesis may disrupt the regulation of downstream events, amplifying

the initial effect by altering the progression of infection e.g by changing the ratio of non-structural proteins and genomic copies during early infection, prolonging the pre-26S RNA phase of infection. Finally, the stem-loops may play a direct role in other aspects of the viral lifecycle such as genome packaging, stability and transition between genomic and sub-genomic RNA synthesis and translation. This would result in a far greater effect on viral replication than the initial reduction in viral RNA synthesis.

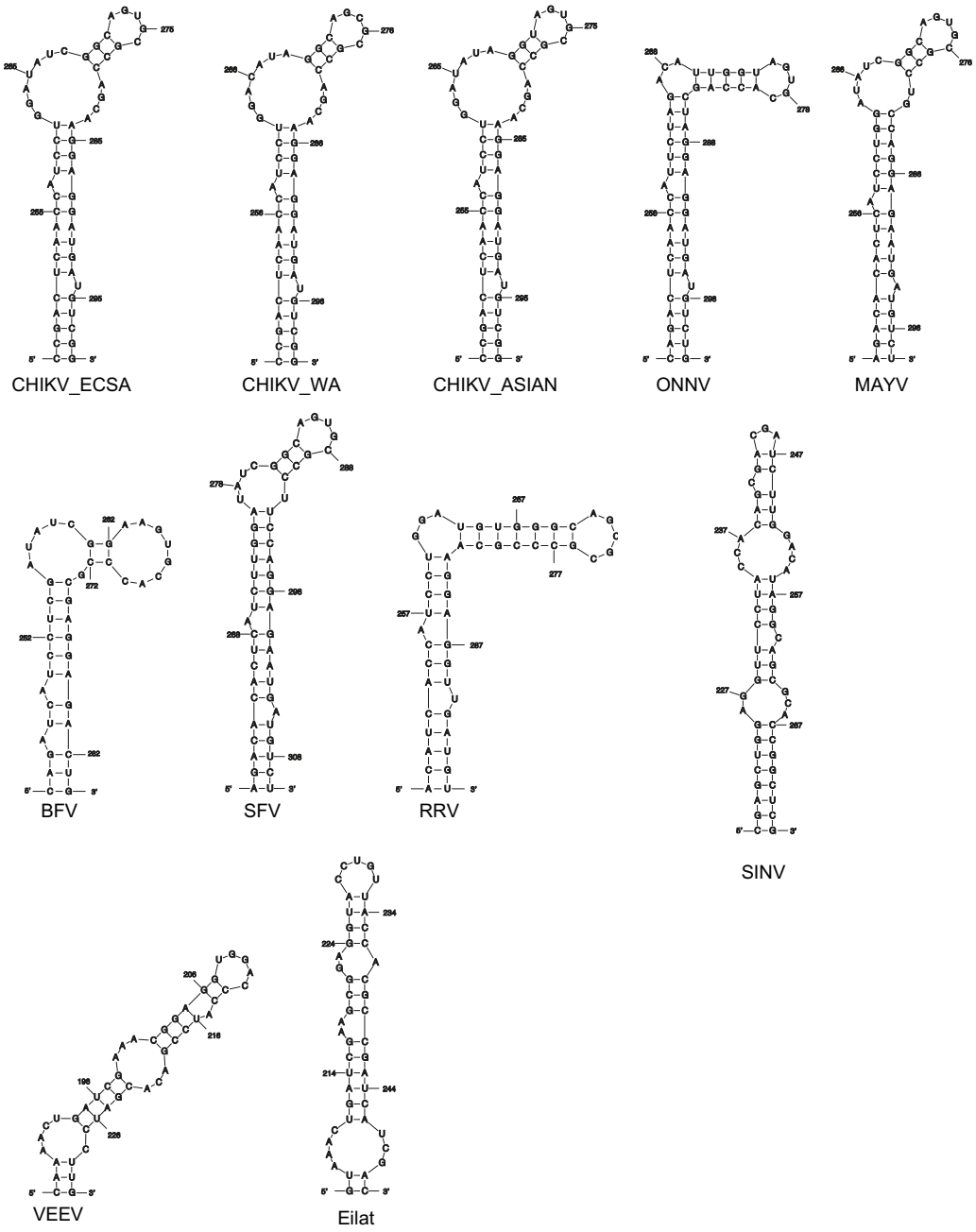


Figure 3.19: Homologues of SL246 across the *Alphavirus* genus. 28°C in silico predicted UNAfold thermodynamic predictions for CHIKV SL246 homologous structures from the ECSA, WA and Asian CHIKV genotypes, closely related alphaviruses (ONNV and MAYV), other members of the complex (BFV, SFV and RRV) and from more distantly related members of the genus (SINV, VEEV and Eilat virus). There is conservation of an elongated stem in this region, which is most conserved in structure and sequence between the Old World viruses most closely related to CHIKV.

As an arbovirus, CHIKV must maintain fitness in both vertebrate and invertebrate host-cell environments. A compelling hypothesis for modulation of CHIKV replication in such different environments was a temperature-dependent switch between alternate structures. Given that mosquitoes are ectothermic insects native to tropical climes, where diurnal temperature range is 0-40°C, temperature-sensitivity may be considered a major disadvantage for any arbovirus. It may be expected that structural variability with temperature would be disadvantageous and viral genomes would evolve to maintain stable structures across as great a temperature range as possible. However, mosquito innate immune responses vary greatly with temperature in a complex manner. For example, in anopheline mosquitoes, melanisation and defensin expression are highest at lower temperatures (18°C) while nitric oxide synthase expression peaks at higher temperatures (30°C) (406). In the case of ZIKV infection of *Ae. aegypti,* a reduction in temperature from 28°C to 20°C limits ZIKV replication via cold-induced changes to the midgut environment such as vitellogenin expression upregulation, altered ROS metabolism and other aspects of innate immunity such as Toll signalling pathway components (407). Thus, temperature-responsive elements may be beneficial in evading host innate immunity. In addition, the transmissible temperature for CHIKV is much narrower than the diurnal temperature range. CHIKV is transmissible between 16°C and 32°C in *Ae. albopictus* with an optimum temperature of 26°C, and between 18°C and 35°C in *Ae. aegypti* with an optimum temperature of 29°C. Above and below this range, transmission drops to zero (408). Therefore, temperature-dependent limits on the stability of the virus exist in mosquitoes. While it may not be surprising that CHIKV maintains a stable genomic RNA structure in the 5′ region between 28°C and 37°C, investigation of temperature-dependence as a mechanism of adaptation to disparate hosts was warranted.

In order to investigate whether the different permissive temperature conditions of mosquito and human cells influenced the thermodynamic stability and resulting folding structure of the CHIKV genome, SHAPE mapping of full-length CHIKV genomic transcripts was carried out by A. Tuplin (University of Leeds) at mosquito and human host-cell permissive temperatures (28 and 37°C respectively). SHAPE reactivity profiles were strikingly similar within the predicted stem–loop structures, indicating that different ambient growth temperatures are not likely to result in alternative folding structures for essential replication elements in this region of the CHIKV genome (396). This suggests that host specificity of RNA replication elements identified and

analysed in this study is more likely due to interaction with host-specific *trans*-activating factors, than from different invertebrate and vertebrate host-cell temperatures stabilising alternative RNA structure conformations. Furthermore, Huh7 cells infected with mutant virus at 28°C displayed similar replication deficiencies to those observed at 37°C, confirming that the temperature difference cannot account for the host-specific phenotypes observed for stem-loop mutants in *nsp1*. While each stem-loop mutant in *nsp1* replicated to a higher titre at 28°C than at 37°C, there remained a significant decrease in viral replication relative to virus of wild-type RNA structure. Interestingly, *SL85mut* replicated to a considerably higher titre at 28°C than at 37°C, around 1.5 logs higher, though viral titre was still significantly decreased compared to WT. This is in line with SHAPE reactivity data which suggests the potential for dynamic interactions around SL85; lower temperatures may preferentially stabilise an alternate conformation of the region, stabilise a tertiary interaction or increase the stability of the stem-loop conformation, allowing it to form at the appropriate stage of replication. Further study is necessary to determine the range of structures possible in this region and the dynamics of folding.

Surprisingly, translation of the viral genome was not significantly impacted by disruption of any of the stem-loops studied. Extensive 5′ RNA secondary structure has been demonstrated to reduce the efficiency of translation in cellular mRNAs (384). Conversely, specific RNA secondary structures have been demonstrated to regulate viral mRNA translation in several viruses (409–411). In this study, disruption of the stem-loops did not significantly affect translation either positively or negatively, suggesting that this region does not contain structures which impede the helicase activity associated with elongation or viral RNA elements which up-regulate translation of the viral genome. However, SHAPE-constrained thermodynamic analysis of the structure in this region (Fig 3.2) reveals base-pairing of the AUG start codon with a second in-frame potential start codon AUG. These codons are separated by SL85 and SL102 in a manner which brings them into close proximity. The mutations introduced into SL85 and SL102 (Fig 3.4A) maintain base-pairing of each AUG start codon, meaning that the lack of translation phenotype observed for *SL85mut* and *SL102mut* (Fig 3.13) does not preclude the possibility of a structure modulating translation in this region, for example through a mechanism alternatively blocking or stabilising AUG:AUG base pairing.

Following the demonstration that each of the nsP1-encoding region stem–loops function as RNA replication elements required for efficient genome replication, further studies were conducted to determine whether the phenotypes observed are due to changes to the primary nucleotide sequence of the duplex stem or the disruption of base-pairing. In addition, the importance of the primary nucleotide sequence within single stranded bulge or terminal-loop regions for replication was investigated, for example as recognition signals for host/viral *trans*-activating factors (Chapter 4). The potential contribution of, as yet uncharacterised, RNA-RNA interactions within the negative-strand RNA intermediate remain to be investigated.

3.3.1 Conclusion

This study indicates the existence of RNA stem-loops within the positive-strand genomic copy of the CHIKV genome, which function by host-dependent and independent mechanisms (i.e. in either human- or mosquito-derived host cells or in both) during efficient replication of the virus genome. Further studies will identify the sequence and structural requirements responsible for enhancement of CHIKV replication by these six stem-loops (Chapter 4) and the potential tertiary interactions within this region (Chapter 5).

Chapter 4. Sequence and structure requirements for CHIKV genome replication

4. 1 Introduction

The previous chapter demonstrated that stem-loop 47 (SL47) within the 5′ UTR of the CHIKV genome enhanced viral genome replication in human- and *Ae. albopictus*-derived cells. In contrast, stem-loops within the adjacent nsP1-encoding region (SL85, SL102, SL165, SL194 and SL246) enhanced CHIKV genome replication in a host-specific manner (Figs 3.5 & 3.6). The mechanism behind the replication phenotypes was unknown, as were the relative importance of primary nucleotide sequence versus folding structure for the functionality of each RNA element.

Stem-loops consist of a heteroduplex stem with a single-stranded terminal loop. Unpaired bulges along the stem may also form part of the structure. Stem-loops impact RNA virus replication via sequence and/or structure-dependent mechanisms in virus families as diverse as *Retroviridae* (311), *Picornaviridae* (312), *Flaviviridae* (313), *Bunyaviriridae* (314), and *Orthomyxoviridae* (315). Following the discovery in the previous chapter that synonymous substitutions in the 5′ region of the CHIKV genome inhibit virus genome replication, it was necessary to distinguish between nucleotide sequence- and structure-dependent effects of mutations in each stem. The original disruptive mutations (Fig 3.4A) were designed in such a manner that synonymous compensatory mutations could be made to restore base-pairing in the stem without returning to the wild-type RNA sequence - allowing the effects of RNA sequence and structure in the stem to be distinguished.

Aside from the importance of RNA sequence and base-pairing in the stem regions, single-stranded terminal loop and bulge sequences have also been shown to function in recognition of viral genomes by *trans*-activating factors. For example, co-crystallisation of the HIV-1 *trans*-activation response element (TAR) stem-loop bound to the viral *trans*-activator of transcription (Tat) protein indicated that the terminal loop directs the interaction (412). In addition, the single-stranded bulge region of TAR has been shown by mutational analysis to contribute to Tat binding (413). In such instances, heteroduplex stems may present recognition sequences for preferential binding with interacting protein or RNA partners. In the case of HIV-1 TAR, disruption of base-pairing would

decrease the availability of terminal loop sequences for binding and thus impair replication indirectly. The relative importance of primary nucleotide sequence and structure in single-stranded terminal loops and bulges for CHIKV stem-loop functionality can be determined by reverse genetic analysis.

Passaging phenotypic virus mutants can be used as a tool to investigate the importance of mutated elements and the interactions by which they function. For example, U_{57}>C_{57} mutants in a non-coding stem-loop in the DENV genome reverted directly to wild-type sequence within 10 passages in 10 of the 18 lineages passaged, indicating the importance of the region during replication of the virus (414). Indirect reversion, forming compensatory mutations which restore optimal base-pairing without reverting to wild-type sequence, can demonstrate the importance of structure over primary sequence in heteroduplex stems. This has occurred in studies of multiple RNA structural elements in cucumber mosaic virus (415) and for the splice donor hairpin of HIV-1 (416). Escape mutants may also identify interactions with other factors which, for example, overcome inhibition of viral replication via pseudoreversion in interacting proteins. Pseudoreversion was demonstrated during passage of VEEV, whereby changes to stem-loop 3 (SL3) were suppressed by non-synonymous mutations in nsP1, nsP2 and nsP4 (358). Similarly, recent passaging studies of the CHIKV 3′ UTR demonstrated that a 258 nt deletion was compensated for by non-synonymous mutation in *nsp1* and synonymous mutation in the capsid gene (417).

The aim of the work described in this chapter is to investigate the relative importance of RNA structure and primary sequence in the function of stem-loops within the 5′ region of the CHIKV genome. Escape mutants were generated by prolonged passaging in each host species to explore the potential genetic routes of reversion.

4.2 Results

4.2.1 Stem-loops in the 5′ region of the CHIKV genome enhance CHIKV replication in a structure-dependent manner

In order to confirm that previously observed mutant phenotypes (Figs 3.5 & 3.6) were due to synonymous-site disruption of base-pairing within predicted stem–loops, rather than due to alteration of the primary nucleotide sequence, compensatory synonymous mutations were incorporated into CHIKV_IC and CHIKV_Rep (Fig 3.4B). Compensatory mutations were designed to restore base-pairing within the duplex-stems without reverting to the wild-type nucleotide sequence (Fig 4.1).

The ability of the compensatory mutations to rescue wild-type levels of CHIKV replication was investigated in both Huh7 and C6/36 cells (Fig 4.2A-C). Within the nsP1-encoding region, restoration of SL85, SL102, SL165 and SL194 significantly rescued CHIKV replication compared to the stem-loop disrupted mutants in Huh7 cells (Fig 4.2A). *SL85mut-Comp* and *SL102mut-Comp* displayed significantly increased CHIKV replication relative to their disrupted counterparts ($p \leq 0.0001$ and $p \leq 0.0005$, respectively). Similarly, *SL165mut-Comp* and *SL194mut-Comp* displayed significantly increased CHIKV replication relative to their disrupted mutants ($p \leq 0.0005$ and $p \leq 0.0001$, respectively). Likewise, compared to *SL246mut,* SL246 was restored to wild-type replication levels in C6/36 cells by compensatory mutations ($p \leq 0.05$) (Fig 4.2B). These results suggest that enhancement of replication by RNA stem-loops in *nsp1* requires base-pairing in the stem regions, rather than the primary nucleotide sequence.

Within the 5′ UTR, restoration of base-pairing in the duplex stem of SL47 restored wild-type levels of CHIKV replication compared to *SL47mut* ($p \leq 0.005$) in mosquito-derived C6/36 cells at the permissive temperature of 28°C (Fig 4.2C)i)). In contrast, compensatory mutations in SL47 did not rescue wild-type levels of replication in human-derived Huh7 cells at 37°C (Fig 4.2C)ii)).

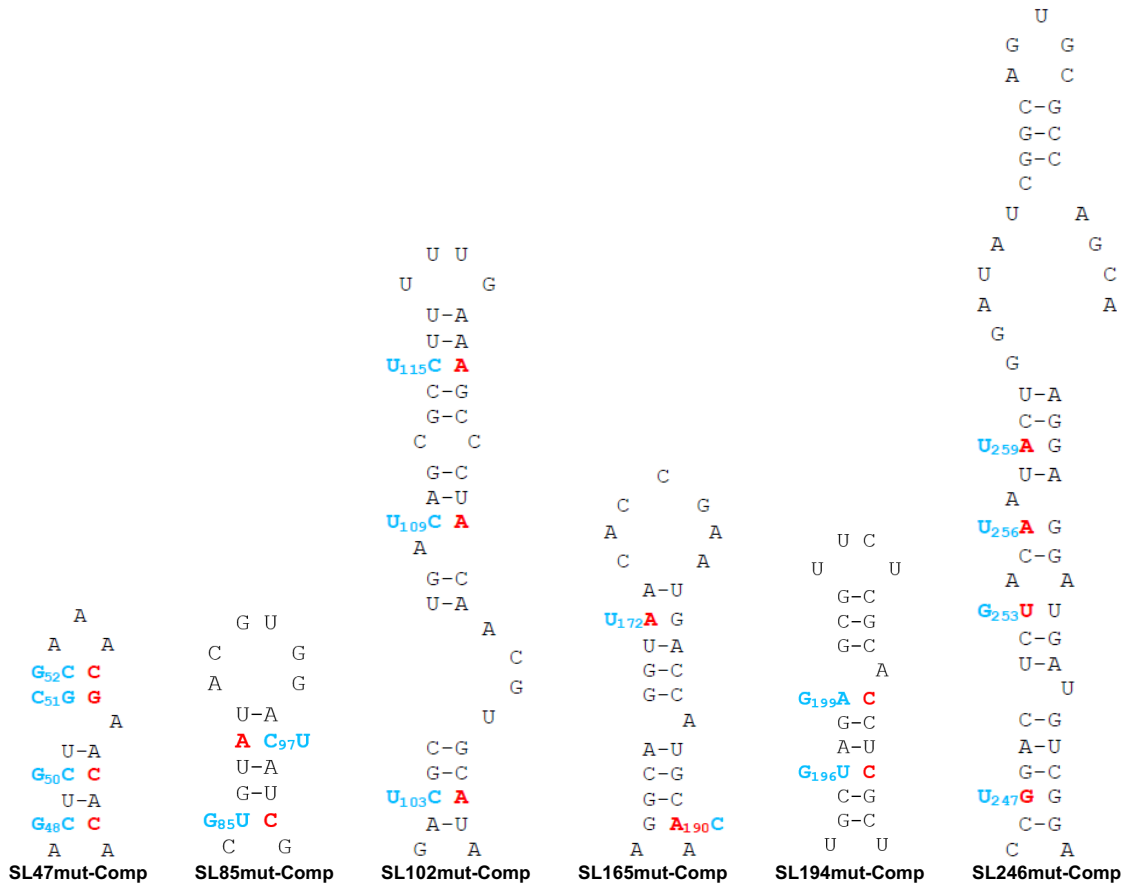


Figure 4.1: Compensatory mutants restoring RNA secondary structure. Mutations predicted to destabilise the heteroduplex stem of CHIKV genomic RNA structures (red) and compensatory mutations predicted to restore base-pairing (blue). All mutations in the nsP1-encoding region are synonymous. Compensatory mutations (blue) are labelled on the outside of the stem, while the sequence in the corresponding SL-*mut* (red) is displayed as part of the stem; for example, in SL165, $U_{172}A$ denotes an A>U mutation relative to *SL165mut* at nt 172 while $A_{190}C$ on the other side of the stem denotes an A>C mutation at nt 190 relative to *SL165mut*. In SL194, $G_{199}A$ denotes an A>G mutation relative to *SL194mut* which restores base-pairing in combination with the C_{211} mutation of *SL194mut* (Fig 3.4A). Mutations to SL47 were introduced by H. Khalid (University of Leeds).

Lower temperatures increase thermodynamic stability, thus making formation of a stable RNA structure more favourable. In order to determine whether the formation of a functional stem-loop in *SL47mut-Comp* was dependent on the lower temperatures at which C6/36 cells are maintained, Huh7 cells were infected at 28°C. Full rescue was observed when Huh7 cells were maintained at 28°C for the duration of the assay ($p \leq 0.005$) (Fig 4.2C)iii)). Temperature-dependent rescue of wild-type phenotype in *SL47mut-Comp* is consistent with *in silico* UNAfold-predicted free energies of folding at 37°C and 28°C. At 37°C, *SL47mutComp* is predicted to be less thermodynamically stable than the wild-type (wild-type $\Delta G = -2.9$ kcal/mol and *SL47mut-Comp* $\Delta G = -2.3$ kcal/mol). At 28°C, restoration of base-pairing in the stem of *SL47mut-Comp* is predicted, despite remaining relatively less stable compared to the wild-type sequence (wild-type $\Delta G = -4.35$ kcal/mol and *SL47mutComp* $\Delta G = -3.75$ kcal/mol). The ability of *SL47mut-Comp* to replicate to wild-type levels in Huh7 and C6/36 cells at 28°C suggests a structure-dependent mechanism for SL47 in both host species. Together, these results demonstrate that base-pairing in the stem of each stem-loop, rather than the primary sequence within the stem, is integral to enhancement of CHIKV replication.

4.2.2 Stem-loops in the 5′ region enhance viral genome replication in a structure-dependent manner

The ability of compensatory mutations to rescue wild-type levels of genome replication specifically, rather than virus infection in general, was confirmed using the sub-genomic replicon system in both Huh7 and C6/36 cells (Fig 4.3A-C). For each stem-loop within *nsp1*, restoration of base-pairing by compensatory synonymous-site substitutions restored wild-type levels of CHIKV genome replication in the relevant host cell line (Fig 4.3A & B). There were no significant differences between compensatory mutants and wild-type sub-genomic replicon replication. Similarly, the temperature-dependence of *SL47mut-Comp* was observed at the level of genome replication; compensation of wild-type phenotype was observed during assays at 28°C but not at 37°C (Fig 4.3C). These results confirm that structure-dependent enhancement of CHIKV replication by stem-loops in the 5′ region occurs specifically at the level of genome replication.

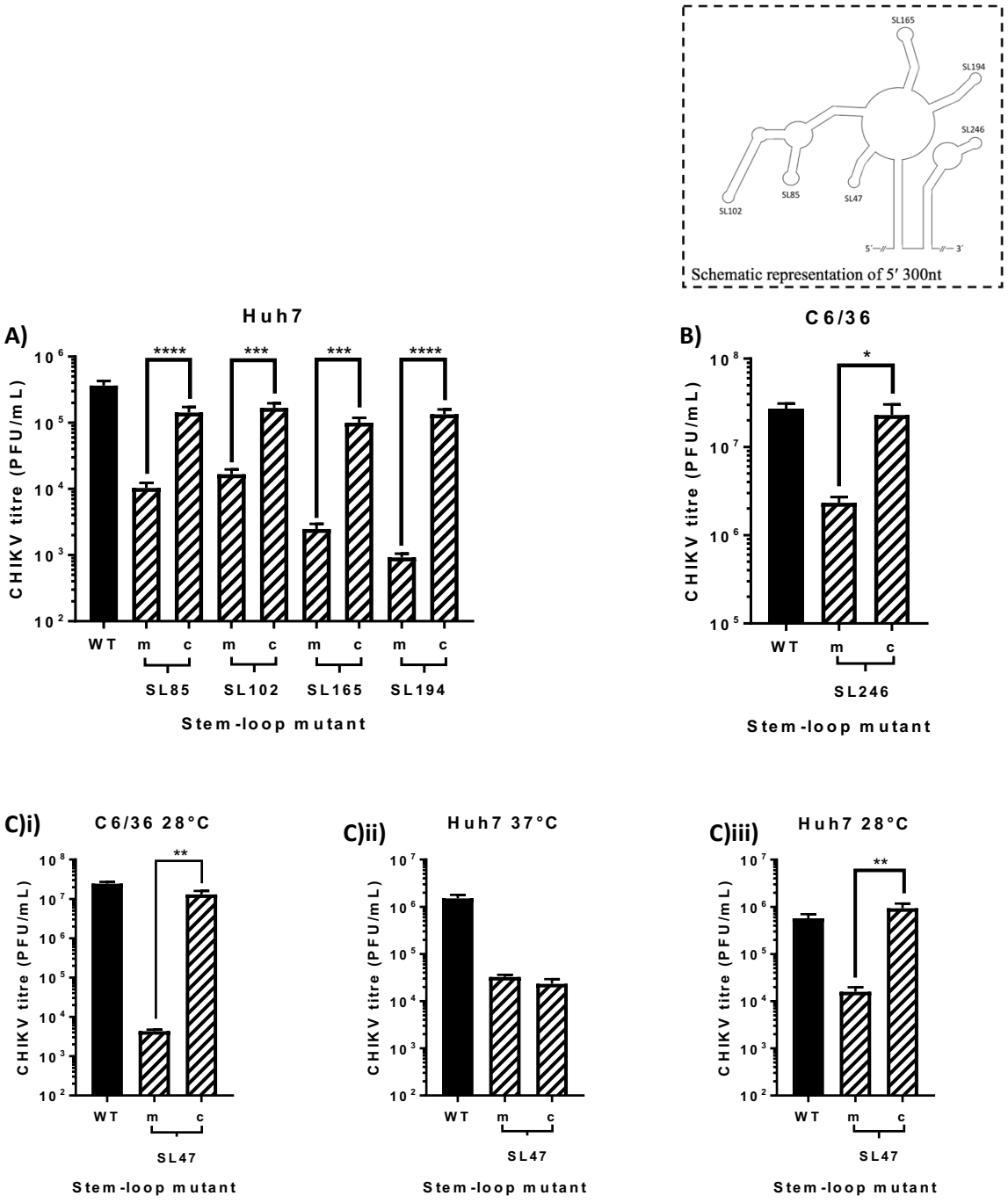


Figure 4.2: SL47-SL246 enhance viral replication in a structure-dependent manner. Replication phenotype of WT CHIKV (n=3) (black) compared to virus containing mutations (hatched bars) predicted to destabilise (m) or restore (c) the heteroduplex stem of genomic RNA structures following infection of Huh7 and C6/36 cells for 24 hours at MOI=1. Results are shown

only for cell lines in which destabilisation of stems exerts an effect on replication: (**A**) Huh7, (**B**) C6/36 cells and (**C**) both cell types. Replication phenotypes for *SL47mut* (m) and *SL47mut-Comp* (c) shown in both cell types, alongside data from Huh7 cells at 28°C. A schematic representation of the 5′ structure of the CHIKV genome is shown for reference (top right). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*), ≤ 0.01 (**), ≤ 0.001 (***). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

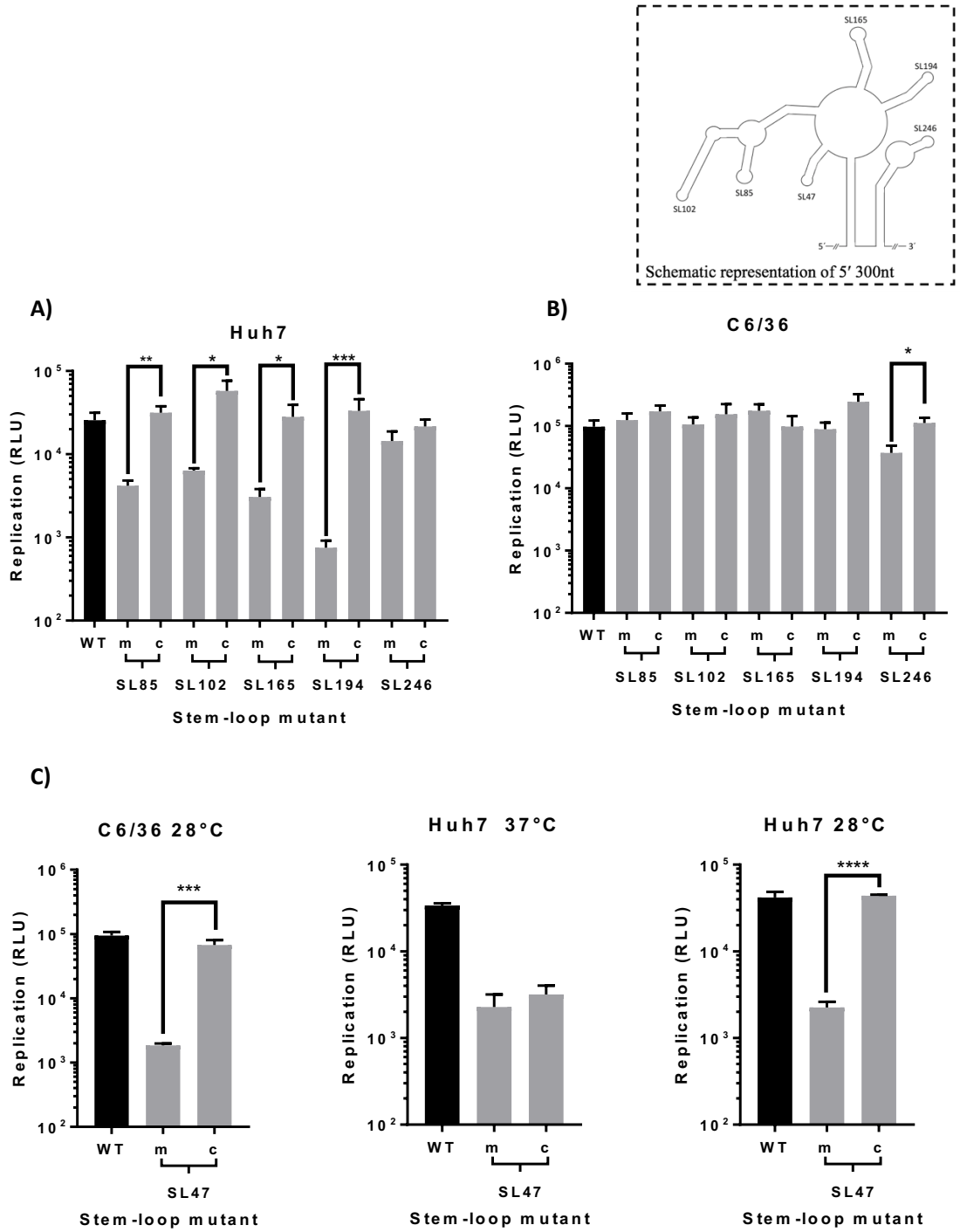


Figure 4.3: SL47-SL246 enhance genome replication in a structure-dependent manner. Replication phenotype of sub-genomic CHIKV replicons (n=3) in (**A**) Huh7 and (**B**) C6/36 cells, at 6 and 24 hours post-transfection respectively. Replicons with WT RNA structure (black) are compared to replicons bearing mutations (grey) predicted to destabilise (m) or restore (c) the heteroduplex stem of individual genomic RNA structures. (**C**) Replication

phenotypes for *SL47mut* (m) and *SL47mut-Comp* (c) shown in both cell types, alongside data from Huh7 cells at 28°C. A schematic representation of the 5′ structure of the CHIKV genome is shown for reference (top right). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.3 Enhancement of viral genome replication requires the terminal loop sequence of SL194

Several of the RNA replication elements mapped by SHAPE-constrained modelling were predicted to contain unpaired terminal-loop and bulge regions (Fig 3.4A). In order to investigate the role of these single-stranded regions within the *nsP1* stem–loops, substitutions were introduced (Fig 4.4). Replication phenotypes, compared to wild-type, were measured in the sub-genomic replicon system in both Huh7 and C6/36 cells (Fig 4.5). Levels of CHIKV genome replication for mutants with substitutions in the single-stranded regions of SL102, SL165 and SL246 were indistinguishable from wild-type in Huh7 and C6/36 cells - suggesting that the stem–loop structures themselves, rather than the primary sequence of unpaired motifs is important for efficient CHIKV genome replication. Increasing the number of unpaired nucleotides in the terminal-loop of SL165, by substitutions predicted to disrupt base-pairing in the apex of the duplex stem (*SL165mut-Loop-IV*), resulted in a small, consistent but non-significant ($p=0.09$) reduction in replication of the sub-genomic replicon (Fig 4.5). However, disruption of base-pairing in the duplex stem could not be excluded as a contributing factor. The only substitution in a single-stranded region of the replication elements which significantly inhibited genome replication was a synonymous C>U substitution within the terminal loop of SL194 (*SL194mut-Loop*). This substitution reproduced a UUUU sequence motif that is conserved at this locus in divergent alphaviruses (including SINV, SFV and VEEV). This substitution significantly inhibited replication in Huh7 cells by ~1 log compared to wild-type ($p\leq0.05$), indicating that the primary terminal loop sequence of SL194, in addition to the requirement for an intact duplex stem, is important for its function as an RNA replication element during CHIKV genome replication.

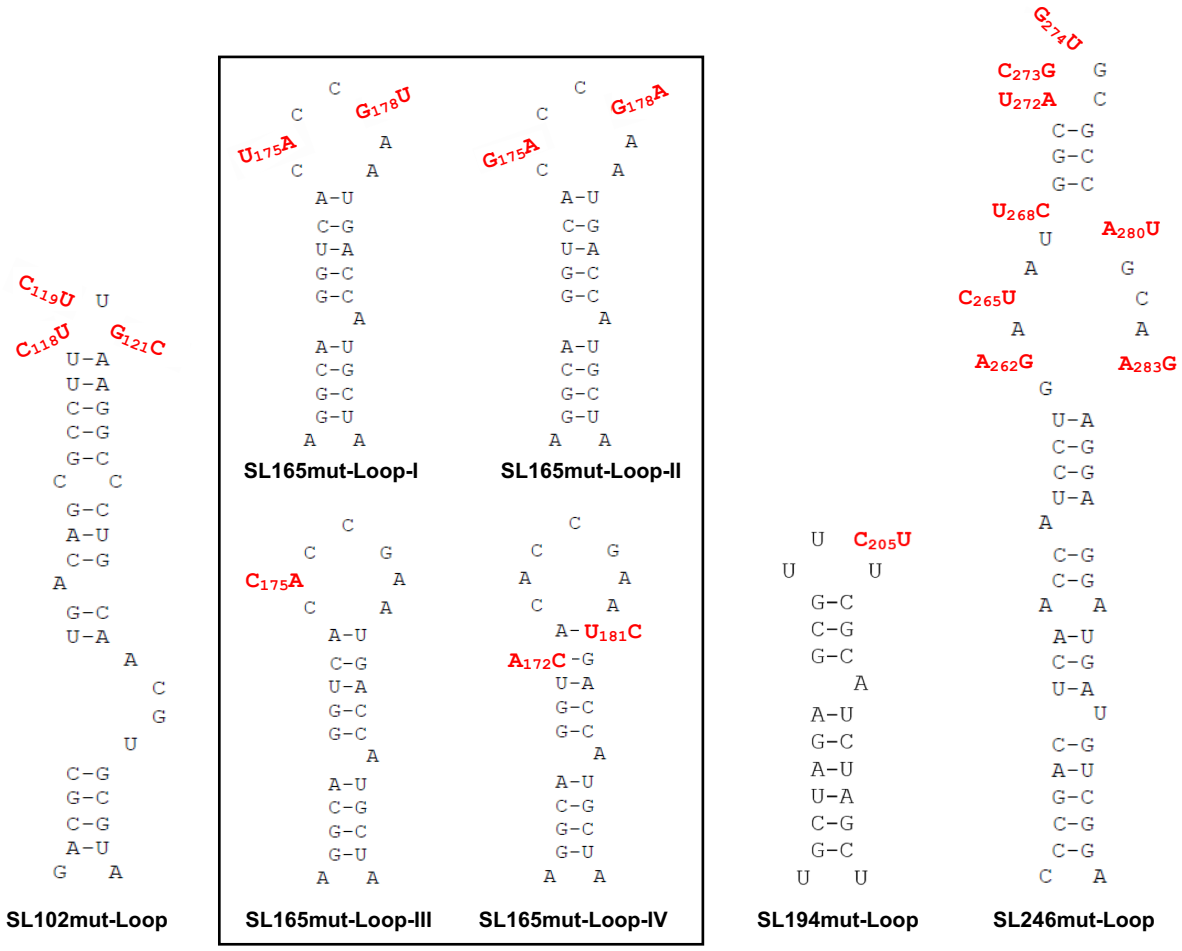


Figure 4.4: Mutations in single-stranded regions of SL102-SL246. Substitutions (red) in single-stranded regions of CHIKV genomic RNA structures. Black box indicates several distinct loop mutants in SL165. Mutations are synonymous in SL165 and SL194. Mutations are more extensive and non-synonymous in SL102 and SL246, where mutations were introduced by H. Khalid (University of Leeds). Mutations are labelled on the outside of the stem, while the wild-type sequence is displayed as part of the stem; for example, in SL165mut-Loop-I, $U_{175}A$ denotes an A>U mutation at nt 175 while $G_{178}U$ on the other side of the loop denotes a G>U mutation at nt 178.

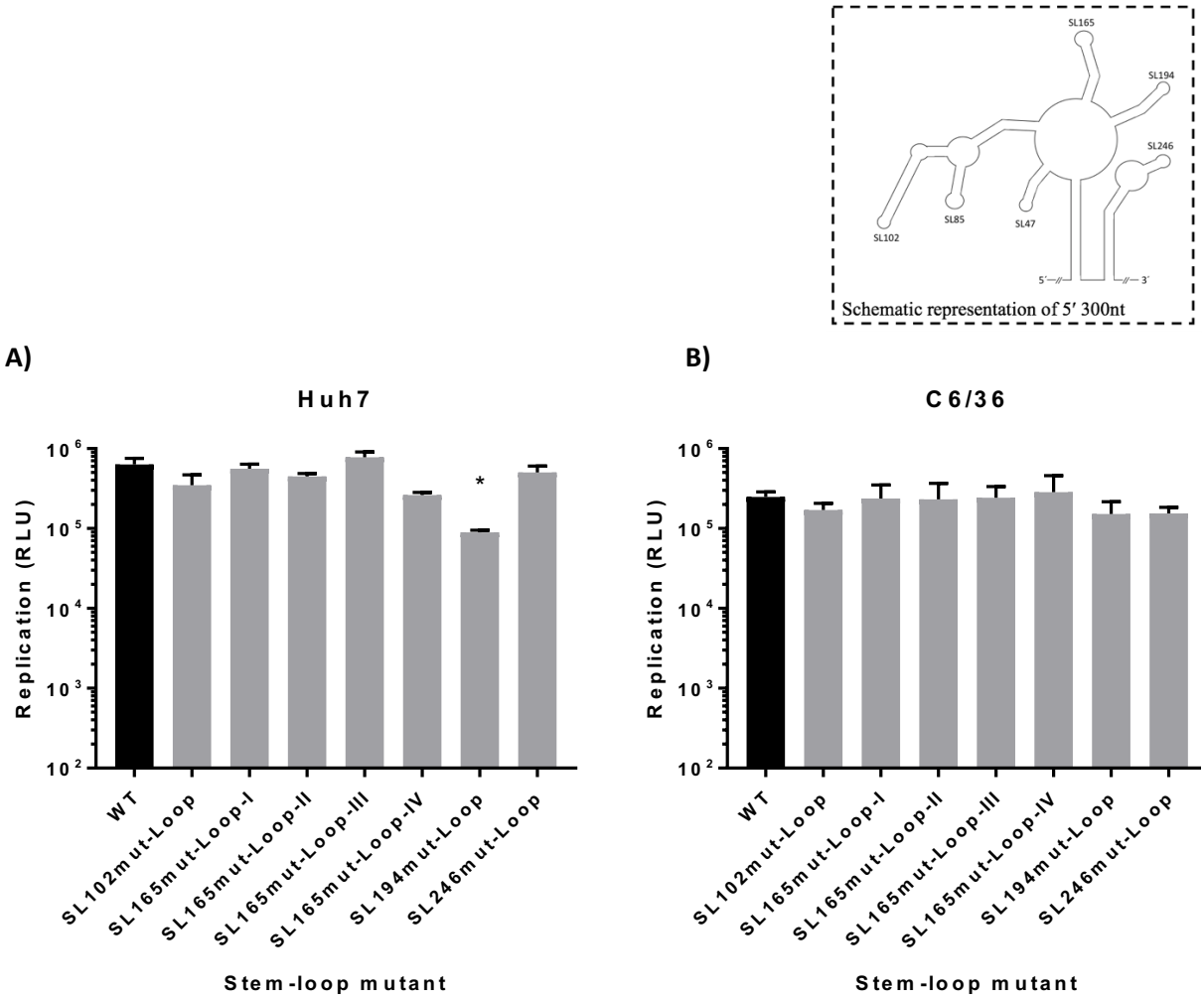


Figure 4.5: The terminal loop sequence of SL194 enhances viral genome replication in Huh7 cells. Replication phenotype of sub-genomic CHIKV replicons (n=3) with wild-type RNA structure (black) compared to replicons bearing mutations (grey) in terminal loop and single-stranded bulge regions of RNA structures in (**A**) Huh7 and (**B**) C6/36 cells, at 6 and 24 hours post-transfection respectively. A schematic representation of the 5′ structure of the CHIKV genome is shown for reference (top right). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.4 Enhancement of CHIKV replication requires the terminal loop sequence of SL194 and the terminal loop size of SL165

The mutants *SL194mut-Loop* and *SL165mut-Loop-IV*, which exhibited a consistent reduction in CHIKV genome replication compared to WT, were assayed for their effect on virus replication in Huh7 cells (Fig 4.6). Virus replication was significantly inhibited for *SL165mut-Loop-IV* compared to wild-type ($p \leq 0.0001$), although this was not observed at the level of genome replication (Fig 4.5). *SL165mut-Loop-IV* contains two substitutions which increase the size of the terminal loop, decreasing the number of base-pairs in the stem and thus reducing its stability (Fig 4.4). These factors complicate the conclusions which can be drawn about the terminal loop of SL165. Significant inhibition of CHIKV replication was observed for *SL194mut-Loop* compared to wild-type ($p \leq 0.005$), recapitulating the phenotype seen during assays of genome replication. Interestingly, disruption of base-pairing in the stem (*SL194mut*) reduced viral replication to a greater degree than alterations to the primary sequence of the terminal loop (*SL194mut-Loop*) ($p \leq 0.01$). This suggests that the intact duplex stem of SL194 is significantly more important to viral replication than the terminal loop sequence, UUCU.

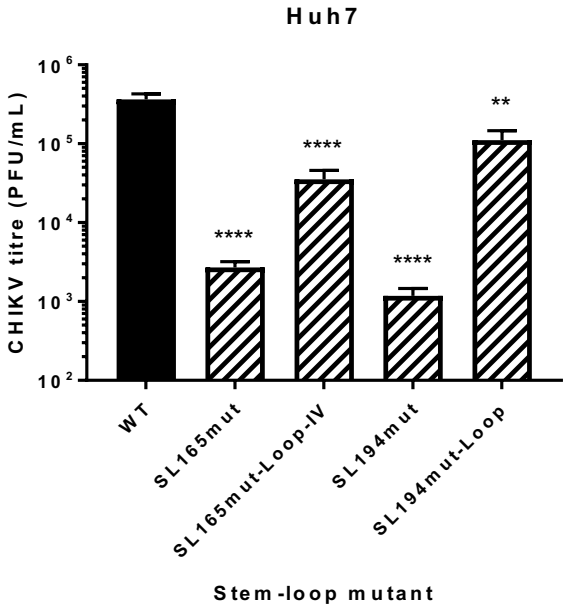


Figure 4.6: The terminal loop sequences of SL165 and SL194 enhance viral replication in Huh7 cells. Replication phenotype of WT CHIKV (black) compared to viruses containing mutations (hatched bars) in terminal loop and single-stranded bulge regions of genomic RNA structures following infection of Huh7 cells for 24 hours. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*), ≤ 0.01 (**), ≤ 0.001 (***), ≤ 0.0001 (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.5 Passaging stem-loop mutants in mammalian cells leads to phenotypic reversion within 10 passages

Passaging, by multiple rounds of harvest/re-infection, can reveal selection pressures acting on specific regions of the viral genome. Selection of escape mutants may result in restoration of wild-type phenotype through direct or indirect reversion of a mutated sequence, to restore wild-type sequence or stem-loop structure respectively. Alternatively, selection pressure can act to restore functional RNA/RNA or RNA/protein interactions, through pseudo-revertant mutations in a long-range RNA interacting sequence or virally expressed protein *trans*-activating partner. Escape mutants provide information about the mechanisms involved in viral replication, whether they act to conserve wild-type sequence, structure or interactions via direct, compensatory or pseudo-reversion respectively. In order to explore the potential routes of reversion, and thus gain mechanistic insight, mutant viruses were passaged repeatedly in cell lines representing appropriate host species, assayed for phenotypic reversion and sequenced by Sanger sequencing to determine the basis of reversion (Figs 4.7 & 4.8).

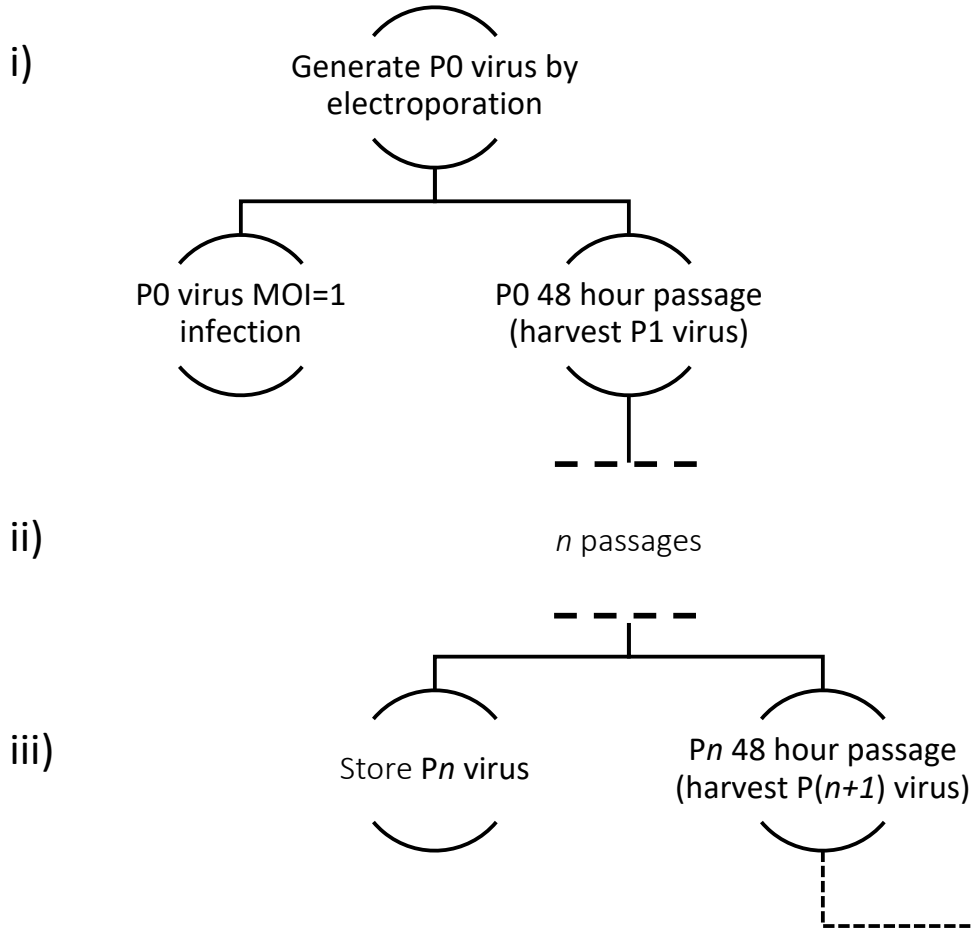


Figure 4.7: Workflow during repeated infection experiments to produce passaged virus. i) P0 virus was generated by electroporation of BHK-21 cells with *in vitro* transcribed CHIKV_IC RNA, titred by plaque assay and one-step growth assay carried out in the appropriate cell line to confirm replication phenotype at the outset of passaging. The first passage was carried out at MOI=1; ~2x10^6 cells were infected with P0 virus (WT and mutants) over 48 hours in 10ml. ii) 9ml of passage supernatant was stored and the remaining 1ml used to infect ~2x10^6 cells. This process was repeated to generate several passages, as denoted by dashed lines. iii) At intervals, passage supernatant was titred to probe for potential phenotypic reversion. Passaging was continued until no significant difference was observed between passaged WT and mutants, as denoted by a dotted line.

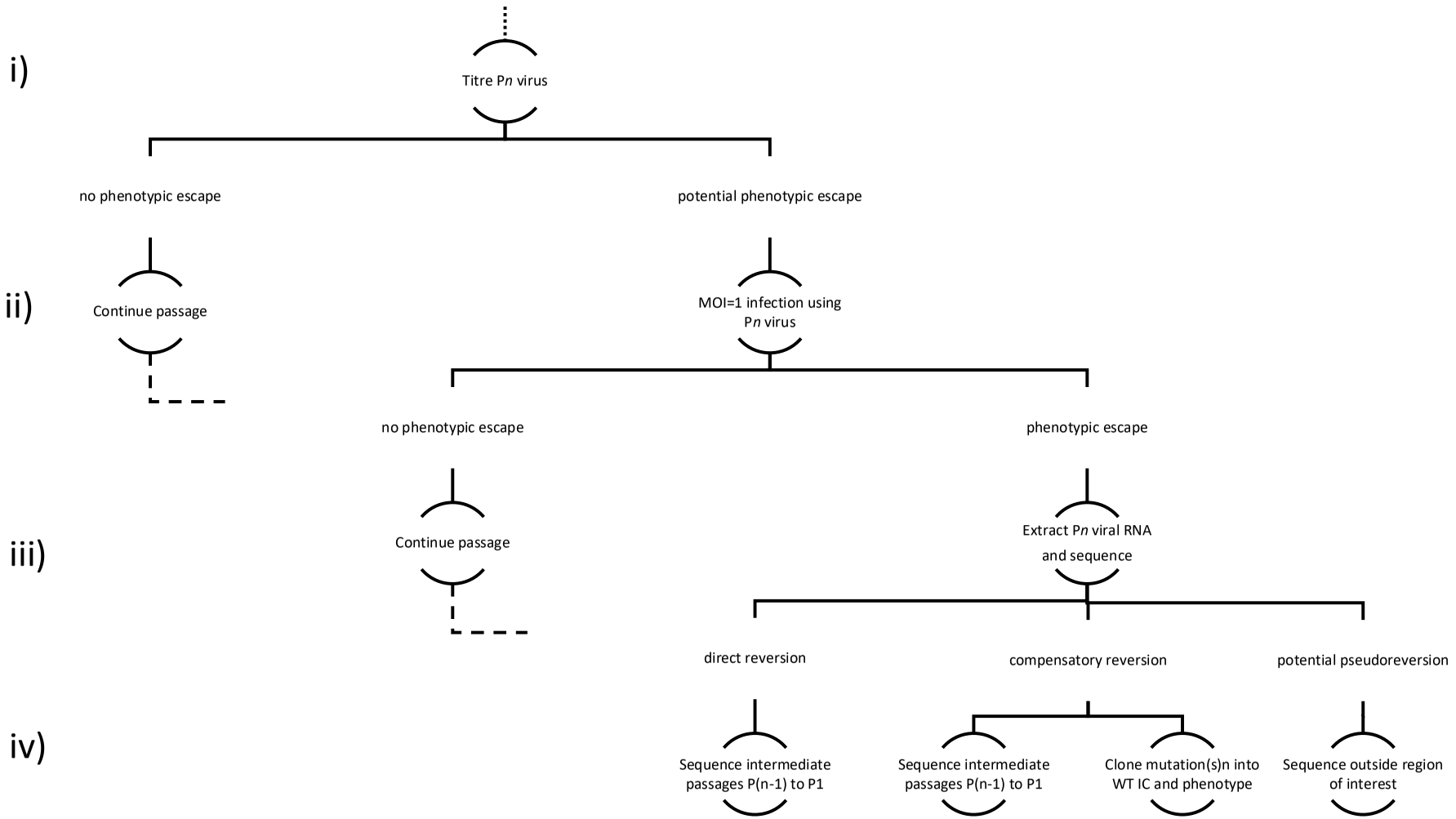

i)
Titre P*n* virus
no phenotypic escape
potential phenotypic escape
ii)
Continue passage
MOI=1 infection using
P*n* virus
no phenotypic escape
phenotypic escape
iii)
Continue passage
Extract P*n* viral RNA
and sequence
direct reversion
compensatory reversion
potential pseudoreversion
iv)
Sequence intermediate
passages P(n-1) to P1
Sequence intermediate
passages P(n-1) to P1
Clone mutation(s)n into
WT IC and phenotype
Sequence outside region
of interest

Figure 4.8: Workflow during passage experiments to analyse passaged virus phenotype and genotype. i) Passage supernatant (P*n* virus) was generated as described in Fig 4.7 (dotted line) and titred by plaque assay. Potential phenotypic escape was determined by comparison of viral titre in 200µl of passaged WT and mutant virus. Where a significant difference in viral titre remained, passaging was resumed (dotted line). Where no significant difference in viral titre was observed, phenotypic escape was confirmed by one-step growth assay. ii) If no significant replication phenotype was observed, the basis of phenotypic escape was determined, otherwise passaging was resumed. iii) ~2x10^6 cells of the relevant cell line were infected with P*n* virus over 48 hours. Infected cells were harvested in Trizol, intracellular viral RNA extracted and reverse transcribed. The region of interest was amplified and sequenced in duplicate. Sequencing the region of interest revealed direct revertants (wild-type sequence in the region of interest), compensatory revertants (substitutions or additional mutations in the region of interest) and potential pseudorevertants (original mutant sequence in region of interest). iv) For direct and compensatory revertants, the accumulation of mutations was examined by sequencing of previous passages. Compensatory revertants, mutations were cloned into CHIKV_IC to confirm that these mutations were the basis of phenotypic reversion. Where no sequence reversion was observed in the region of interest, the non-structural ORF-1 was sequenced to detect mutations outside the region of interest.

Stem-loop mutant viruses were generated by electroporation of *in vitro* transcribed RNA into BHK-21 cells, as described in 2.4.1, to produce virus (P0) for serial passage experiments. Stem-loop mutants were passaged and assayed alongside wild-type virus in order to discount host adaptation as the cause of genotypic changes. As expected, each of the mutants *SL47mut, SL85mut, SL102mut, SL165mut* and *SL194mut* exhibited significantly reduced replication at P0 compared to wild-type virus during a one-step growth assay in Huh7 cells ($p \leq 0.01$)(Fig 4.9A). Huh7 cells were then infected with P0 virus at MOI=1 for 48 hours, before $1/10^{th}$ of the supernatant was transferred to a new flask of uninfected cells. Passaging continued for a total of 10 rounds in Huh7 cells. Several passages were titred between P0 and P10 to sample for potential phenotypic reversion. However, as significant reversion to wild-type phenotype was not observed, passaging was continued blind until P10. Analysis at P10 indicated that mutant viruses *SL47mut, SL85mut, SL102mut*, *SL165mut* or *SL194mut* were replicating to wild-type levels, suggesting phenotypic reversion. In order to confirm the presence of escape mutants, a one-step growth assay was carried out with P10 virus (Fig 4.9B). No significant difference in viral titre was observed for any of the stem-loop mutants compared to wild-type, demonstrating that phenotypic reversion had occurred.

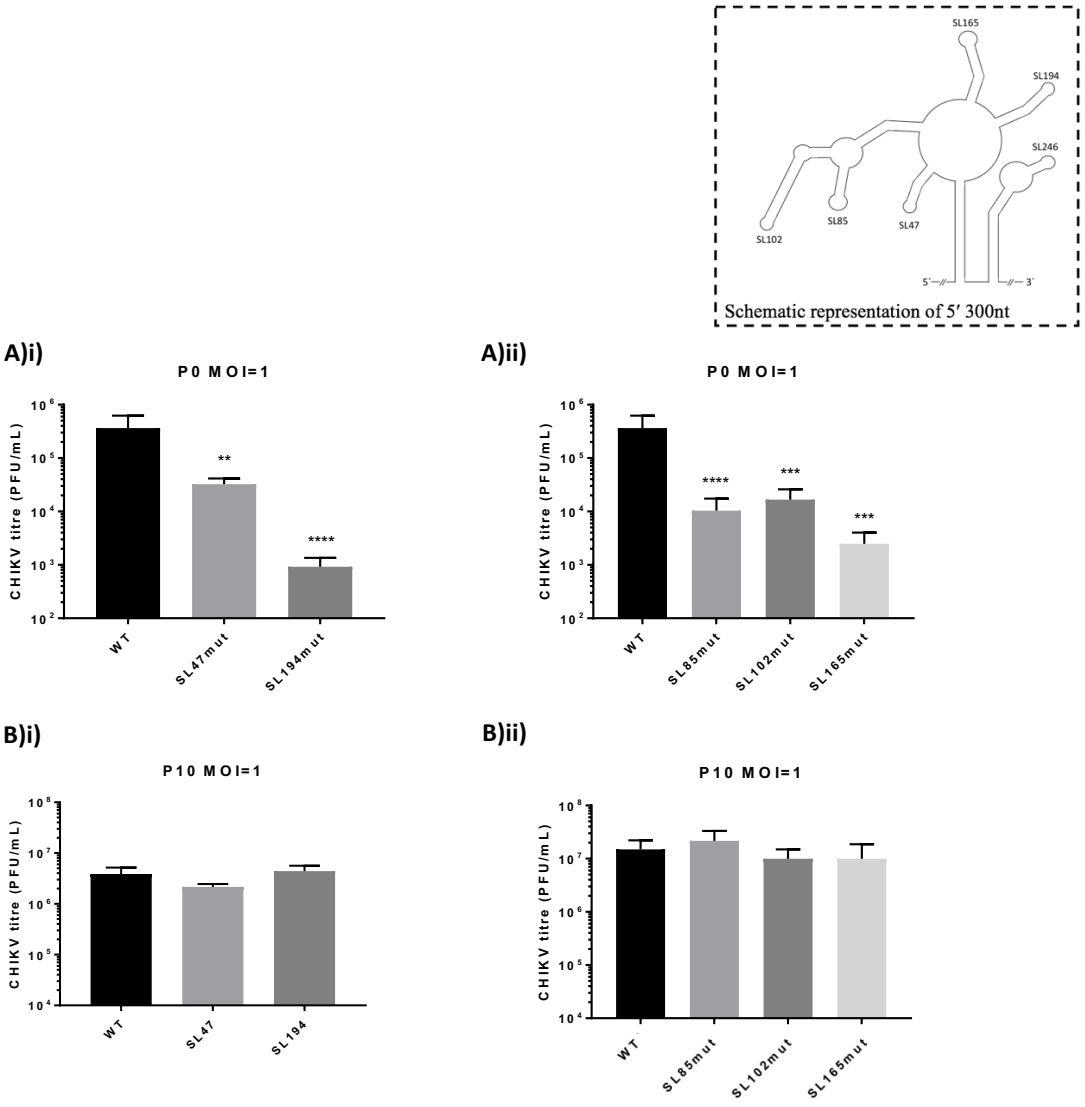


Figure 4.9: Mammalian cell passaging leads to phenotypic reversion within ten passages. Results of passaging stem-loop mutants i) *SL47mut* and *SL194mut* and ii) *SL85mut*, *SL102mut* and *SL165mut* in mammalian cells. Replication phenotype of WT CHIKV (black bar) compared to virus bearing mutations predicted to destabilise the heteroduplex stem RNA structures (grey bars) following MOI=1 infection of Huh7 cells over 24 hours with **(A)** P0 and **(B)** P10 virus. A schematic representation of the 5′ structure of the CHIKV genome is shown for reference (top right). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***), ≤0.0001 (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.6 Escape mutants exhibit direct and indirect reversion in the 5′ region in mammalian cells

In order to determine the genotypic basis of reversion, RNA was extracted from Huh7 cells infected with P10 *SL47mut, SL85mut, SL102mut, SL165mut* and *SL194mut* viruses. The first 1 kb of each P10 viral genome was reverse transcribed, amplified by PCR and sequenced by Sanger sequencing. Direct reversion to wild-type sequence was observed for *SL47mut-SL165mut* at P10 (Fig 4.10). *SL194mut* contained the two synonymous substitutions in the duplex stem at P10, alongside an additional compensatory mutation predicted to restore base-pairing.

4.2.6.1 SL47 mutant reversion

The sequential genetic basis of reversion for each mutant was then mapped by sequencing earlier passages and modelling corresponding thermodynamically-predicted structures using UNAfold software. *SL47mut* retained four substitutions until P5 (Fig 4.11A). Two revertant mutations occurred during P6 ($C_{56}G_{57}$>$G_{56}C_{57}$), although these did not affect the RNA structure of the region, which remained single-stranded (Fig 4.11B). This suggests that there is selection pressure to restore the sequence, independent of the structure. A further revertant mutation during P7 (C_{60}>G_{60}) resulted in the formation of a short stem-loop (Fig 4.11C), followed by a final reversion (C_{61}>G_{61}) during P9 which restored the wild-type sequence and therefore the structure in full by re-forming the base-pair at the base of the stem (Fig 4.11D).

Stem-loop	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
SL47	×	×	×	×	×	×(2/4)	×(3/4)	×(3/4)	✓	✓
SL85	×	×	×	×	×	✓	✓	✓	✓	✓
SL102	×	×	×	×	×	×(1/3)	✓	✓	✓	✓
SL165	×	×	×	×	× (1/2 comp)	✓	✓	✓	✓	✓
SL194	×(1/2)	×(1/2)	×(1/2 + comp)	✓(comp)	✓(comp)	✓(comp)	✓(comp)	✓(comp)	✓(comp)	✓(comp)

Figure 4.10: Development of reversion in Huh7 cells over 10 passages (P1-P10) for each stem-loop. Ticks represent direct or compensatory reversion to wild-type sequence or structure respectively. Crosses represent no or incomplete reversion. Number of mutations reverted to the wild-type sequence from P0 is expressed as a fraction (eg 2/4 mutations reverted at P6). 'Comp' represents a compensatory mutation which reforms a base-pair but does not restore wild-type sequence. Yellow dashed lines denote sampling for phenotypic reversion in passage supernatant, yellow filled boxes denote phenotypic reversion during one-step growth assay.

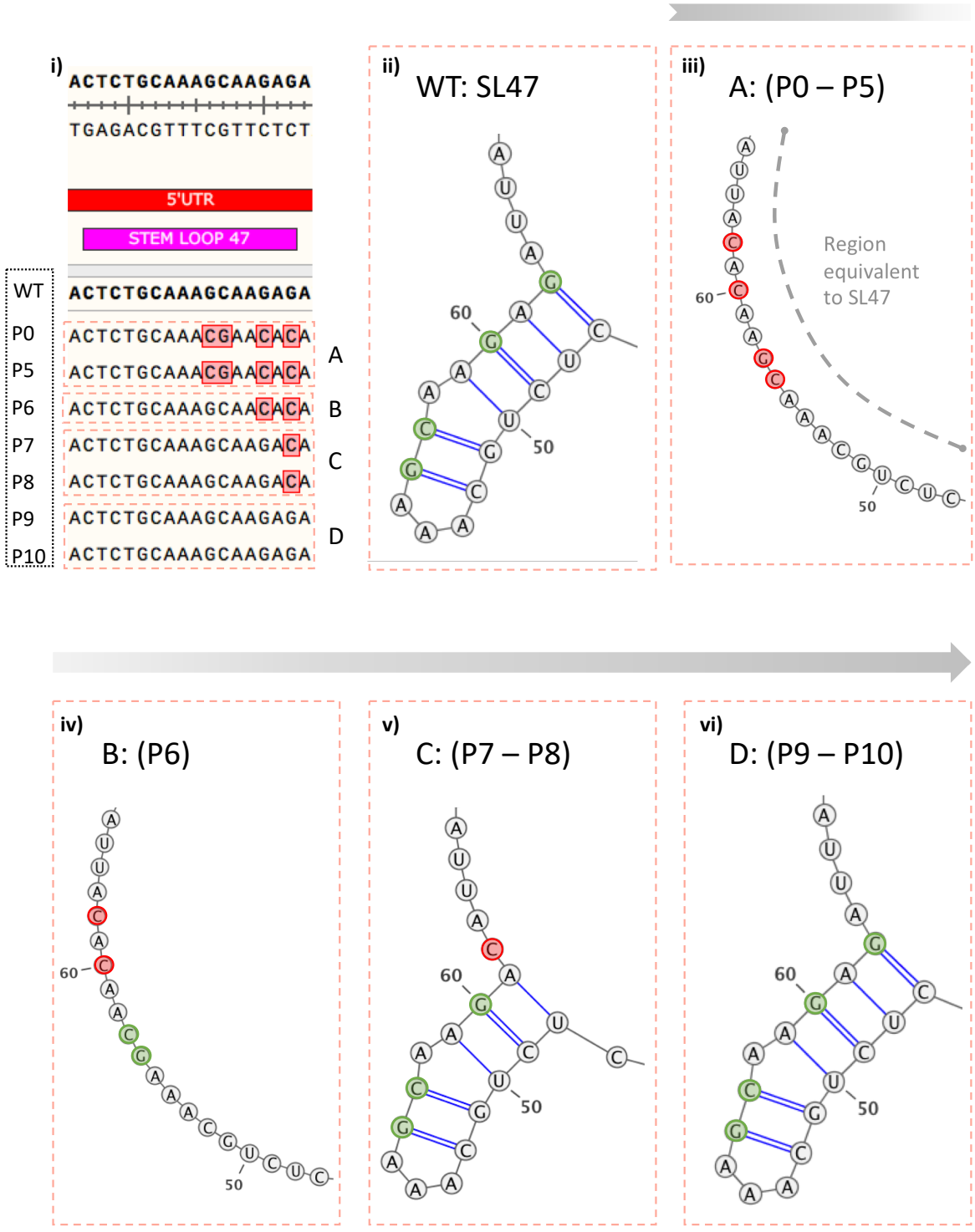


Figure 4.11: Development of reversion in SL47 from passage 0 to passage 10 (P0-P10) in Huh7 cells. i) The passage sequences of P0 and P5-P10 are aligned below the WT sequence. The mutations of *SL47mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups

containing the same sequence are labelled A:D). **ii)** The wild-type structure of SL47 is displayed, with the sites targeted for mutation in P0 highlighted (green circles). **iii-vi)** The region equivalent to SL47 is denoted by a grey dotted line in the P0 predicted structure. Mutated nucleotides are displayed for passage sequences (red circles). Each panel A:D displays the UNAfold-predicted structure in the region for the passage sequence shown in the alignment. The order of sequential passages is denoted by a grey arrow.

4.2.6.2 SL85 mutant reversion

SL85mut retained both substitutions until P5 (Fig 4.12). The presence of both mutations resulted in an elongated stem preceding SL102 (Fig 4.13A). During P6, direct reversion occurred to restore wild-type sequence and thus structure of SL85 (Fig 4.13B).

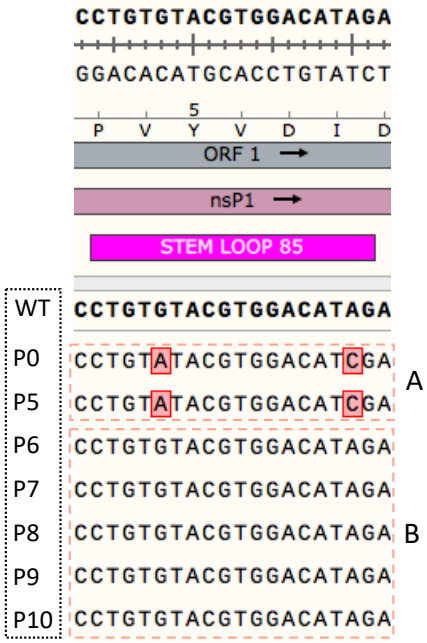


Figure 4.12: Development of reversion in SL85 from passage 0 to passage 10 (P0-P10) in Huh7 cells. The passage sequences of P0 and P5-P10 are aligned below the WT sequence. The mutations of *SL85mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups containing the same sequence are labelled A and B).

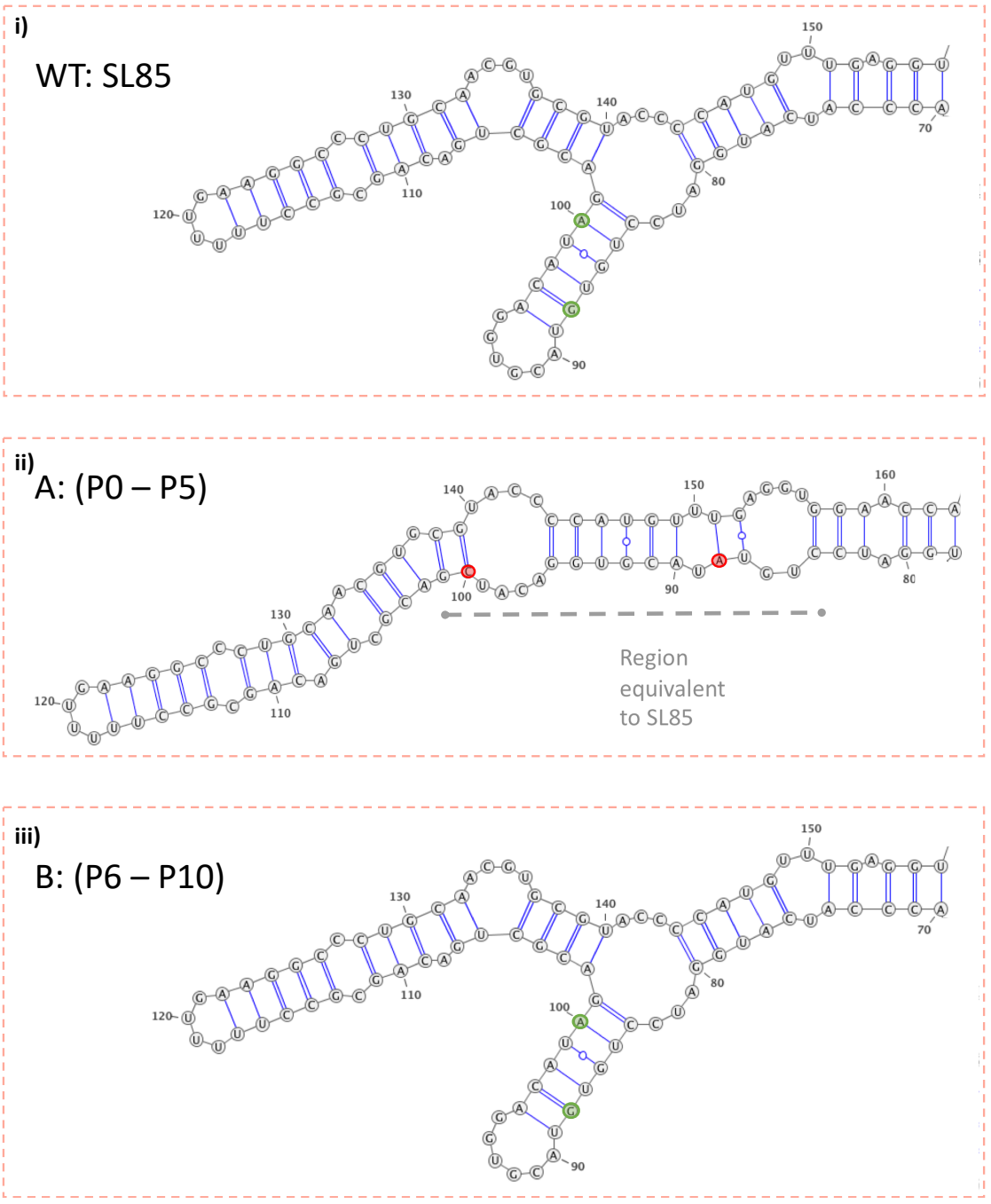


Figure 4.13: Incremental restoration of SL85 structure during passaging. i) The wild-type structure of SL85 is displayed, with the sites targeted for mutation in P0 highlighted (green circles). **ii-iii)** The region equivalent to SL85 is denoted by a grey dotted line in the P0 predicted structure. Mutated nucleotides are displayed for passage sequences (red circles). Each panel A:B displays the UNAfold-predicted structure in the region for the passage sequence shown in the alignment.

4.2.6.3 SL102 mutant reversion

SL102mut retained three G>A stem mutations until P5 (Fig 4.14). The combination of these mutations resulted in the disruption of SL85, the elongation of SL102, more prominent and frequent bulges in the stem and increase in the size of the terminal loop from 4 to 7nt (Fig 4.15A). During P6, the G_{130}>A_{130} substitution of *SL102mut* detectably reverted to the wild-type A (Fig 4.14). This reversion drastically altered the structure of SL102, producing an expanded bulge which contained the sequence of the original terminal loop in a single-stranded configuration (Fig 4.15B), whereas it had been base-paired in *SL102mut* (Fig 4.15A). During P7, the final two G>A mutations were reverted to the wild-type sequence (Fig 4.14). Direct reversion restored the wild-type structure for both SL102 and the preceding stem-loop SL85.

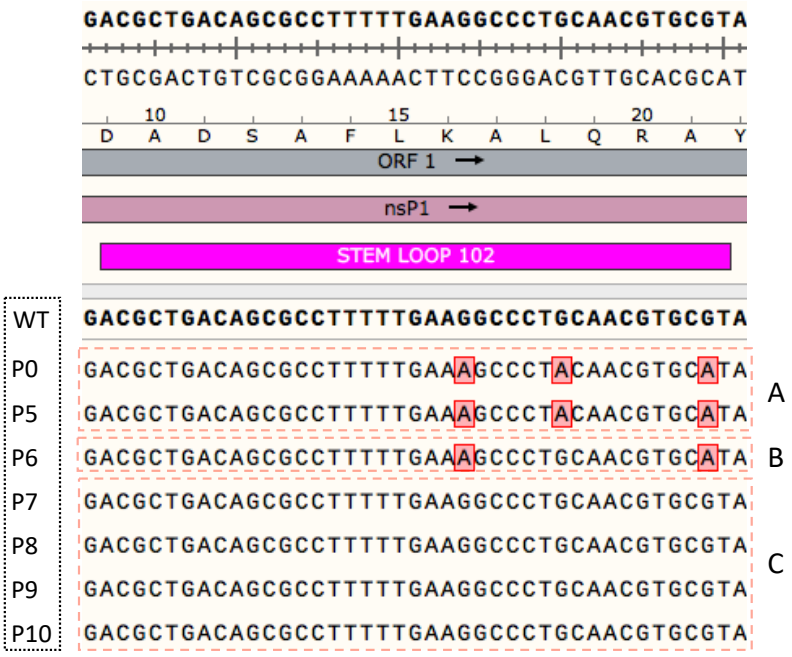


Figure 4.14: Development of reversion in SL102 from passage 0 to passage 10 (P0-P10) in Huh7 cells. The passage sequences of P0 and P5-P10 are aligned below the WT sequence. The mutations of *SL102mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups containing the same sequence are labelled A:C).

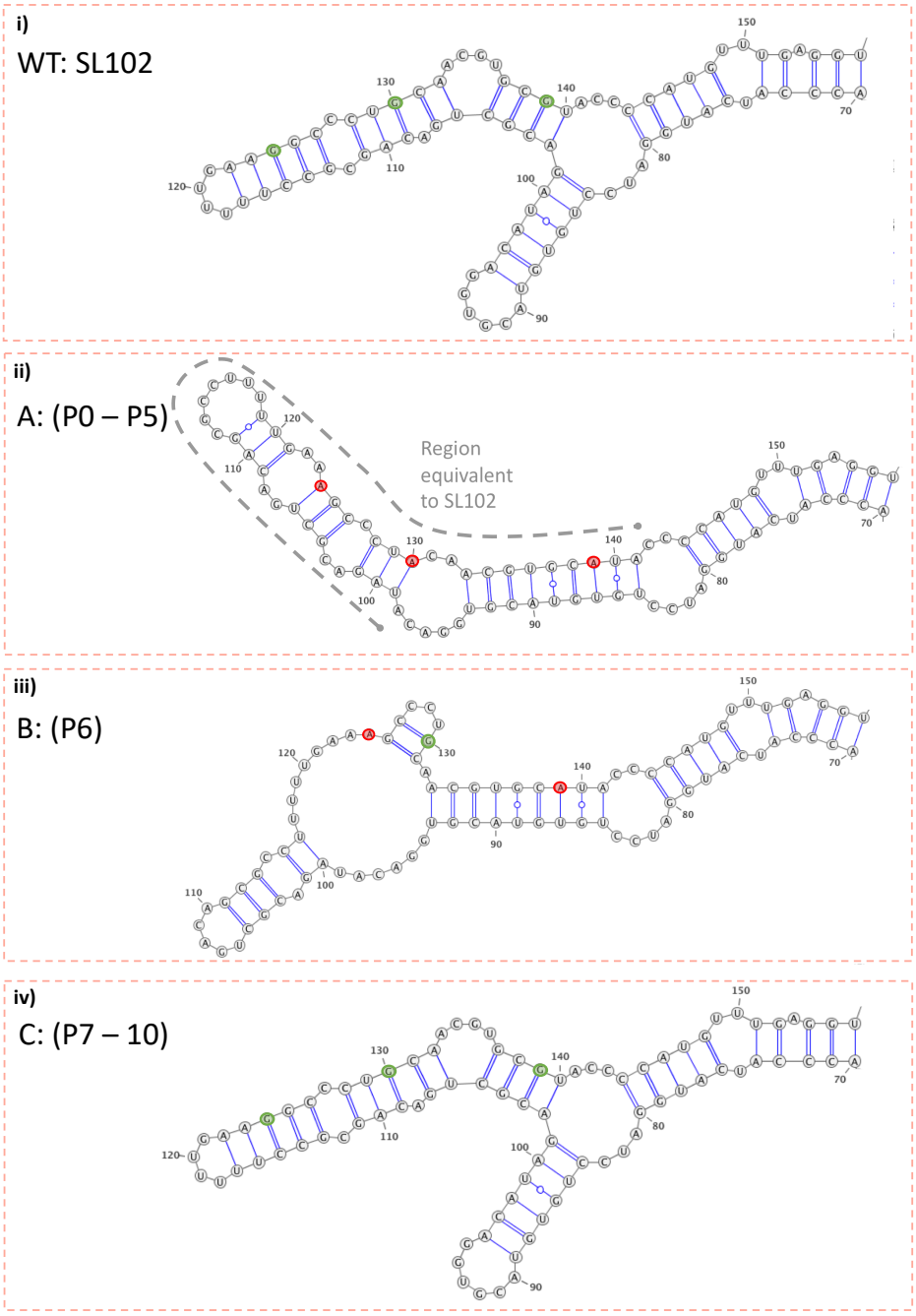


Figure 4.15: Incremental restoration of SL102 structure during passaging. i) The wild-type structure of SL102 is displayed, with the sites targeted for mutation in P0 highlighted (green circles). **ii-iv)** The region equivalent to SL102 is denoted by a grey dotted line in the P0 predicted structure. Mutated nucleotides are displayed for passage sequences (red circles). Each panel A:C displays the UNAfold-predicted structure in the region for the passage sequence shown in the alignment. The order of sequential passages is denoted by a grey arrow.

4.2.6.4 SL165 mutant reversion

SL165mut retained both original substitutions (C_{172}>A_{172} and U_{190}>A_{190}) until P4 (Fig 4.16). These mutations are sufficient to disrupt the structure of the stem-loop (Fig 4.17A). Indirect reversion occurred during P5, whereby an A_{190}>C_{190} mutation restored base-pairing in the base of the duplex stem via a G_{165}:C_{190} pair (Fig 4.17B). This synonymous compensatory mutation produced a more stable base-pair than the wild-type non-Watson Crick base-pair G:U at this site, allowing restoration of base-pairing within the stem, despite the continued presence of a C_{172}>A_{172} mutation at the apex. The presence of the C_{172}>A_{172} mutation increased the size of the terminal loop of SL165 by two nucleotides compared to wild-type SL165. Direct reversion occurred during P6 for both the C_{172}>A_{172} mutation at the apex of the stem and the compensatory A_{190}>C_{190} mutation at the base (Fig 4.16). Direct reversion at the base of the stem restored the weaker G:U base-pair seen in wild-type SL165 (Fig 4.17C).

Figure 4.16: Development of reversion in SL165 from passage 0 to passage 10 (P0-P10) in Huh7 cells. The passage sequences of P0 and P5-P10 are aligned below the WT sequence. The mutations of *SL165mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups containing the same sequence are labelled A:C).

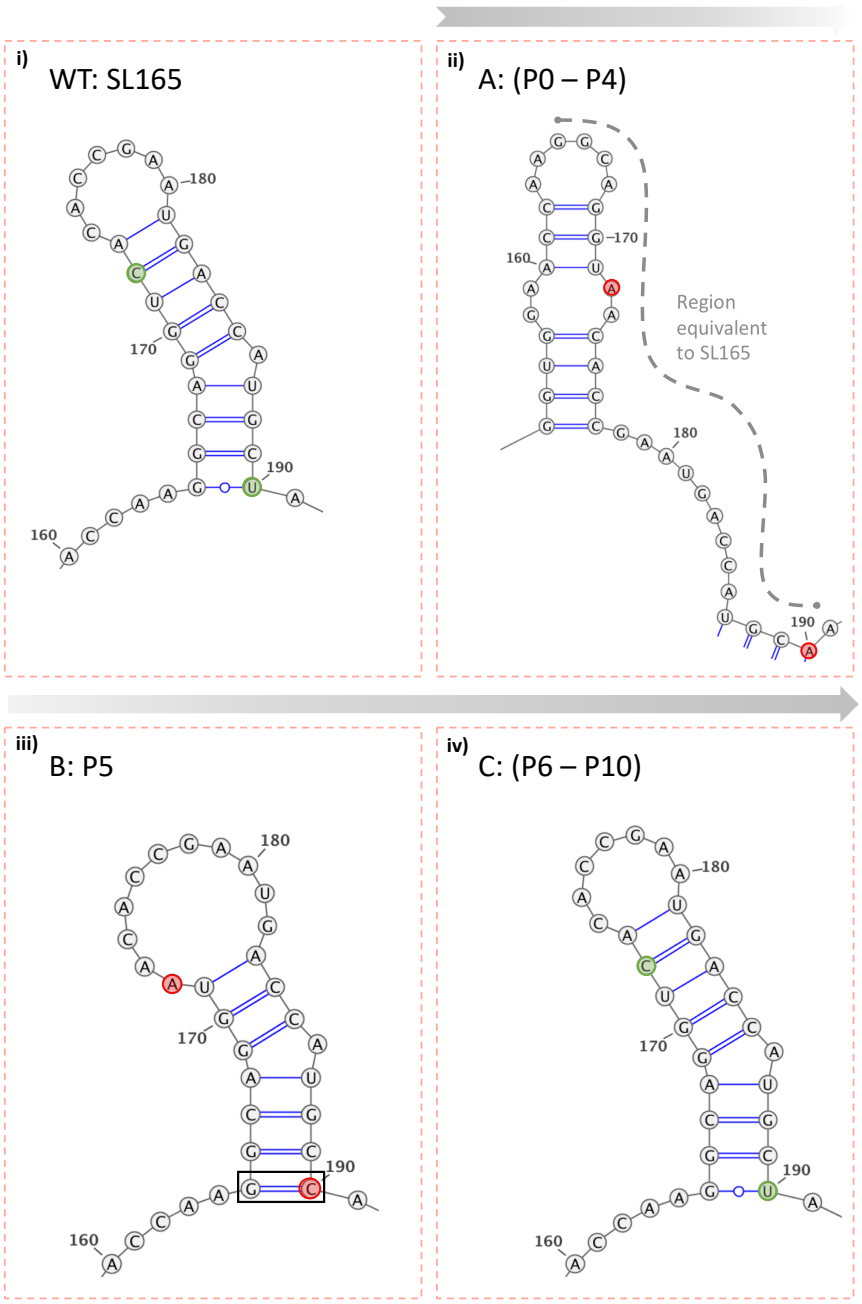


Figure 4.17: Incremental restoration of SL165 structure during passaging. i) The wild-type structure of SL165 is displayed, with the sites targeted for mutation in P0 highlighted (green circles). **ii-iv)** The region equivalent to SL165 is denoted by a grey dotted line in the P0 predicted structure. All mutated (non- wild-type) nucleotides are displayed for passage sequences (red circles). Compensatory mutations which do not restore wild-type sequence but reform base-pairing in the stem associated with wild-type structure are boxed in black across the base-pair. Each panel A:C displays the UNAfold-predicted structure in the region for the passage sequence shown in the alignment. The order of sequential passages is denoted by a grey arrow.

4.2.6.5 SL194 mutant reversion

SL194mut contained two substitutions, U_{211}>C_{211} and A_{214}>C_{214} (Fig 4.18). These mutations are sufficient to disrupt the structure of the stem-loop (Fig 4.19A). Direct reversion of the U_{211}>C_{211} mutation was detectable in P1 (Fig 4.18), which partially reformed the duplex stem, allowing presentation of the wild-type terminal loop and A bulge (Fig 4.19B). However, this mutant still contained the A_{214}>C_{214} mutation at the base of the stem, which resulted in a U_{196}/C_{214} bulge. During P3, a compensatory mutation altered the A_{199}:U_{211} base-pair to a G_{199}:U_{211} base-pair via synonymous A_{199}>G_{199} substitution (Fig 4.18). This substitution retained base-pairing in the stem (Fig 4.19C). Interestingly, during P4, the original U_{211}>C_{211} substitution present in *SL194mut* was restored (Fig 4.18). This U_{211}>C_{211} return to the *SL194mut* sequence produced a co-variant G:C bond, more stable than the original A:U bond present in wild-type SL194 (Fig 4.19D). The compensatory A_{199}>G_{199} mutation, in combination with the original substitutions present in *SL194mut*, was detectable as the predominant sequence in the population in all passages thereafter (Fig 4.18). In order to demonstrate that the compensatory A_{199}>G_{199} mutation alone was sufficient to cause phenotypic reversion, the compensatory mutation was incorporated into *SL194mut* to form *SL194mut(A199G)* (Fig 4.20A). A one-step growth assay was carried out with *SL194mut(A199G)* virus in Huh7 and C6/36 cells (Fig 4.20B & C). There was no significant difference between WT and *SL194mut(A199G)* CHIKV replication, indicating that the A_{199}>G_{199} mutation in P10 *SL194mut* was responsible for restoration of wild-type CHIKV replication. Surprisingly, the compensatory mutation detected at P4 did not result in phenotypic reversion during sampling of P5 passage supernatant (Fig 4.10). Phenotypic reversion was detected during sampling of P10 supernatant and subsequent one-step growth assay. In light of the phenotypes observed for *SL194mut(A199G),* it is likely that A_{199}>G_{199} did not become fixed in the population until after P5 and thus a small but significant replication phenotype remained.

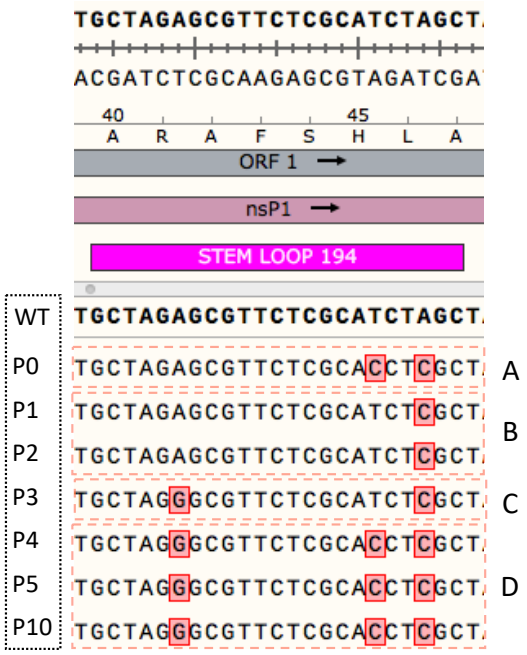


Figure 4.18: Development of reversion in SL194 from passage 0 to passage 10 (P0-P10) in Huh7 cells. The passage sequences of P0-P5 and P10 are aligned below the WT sequence. The mutations of *SL194mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups containing the same sequence are labelled A:D).

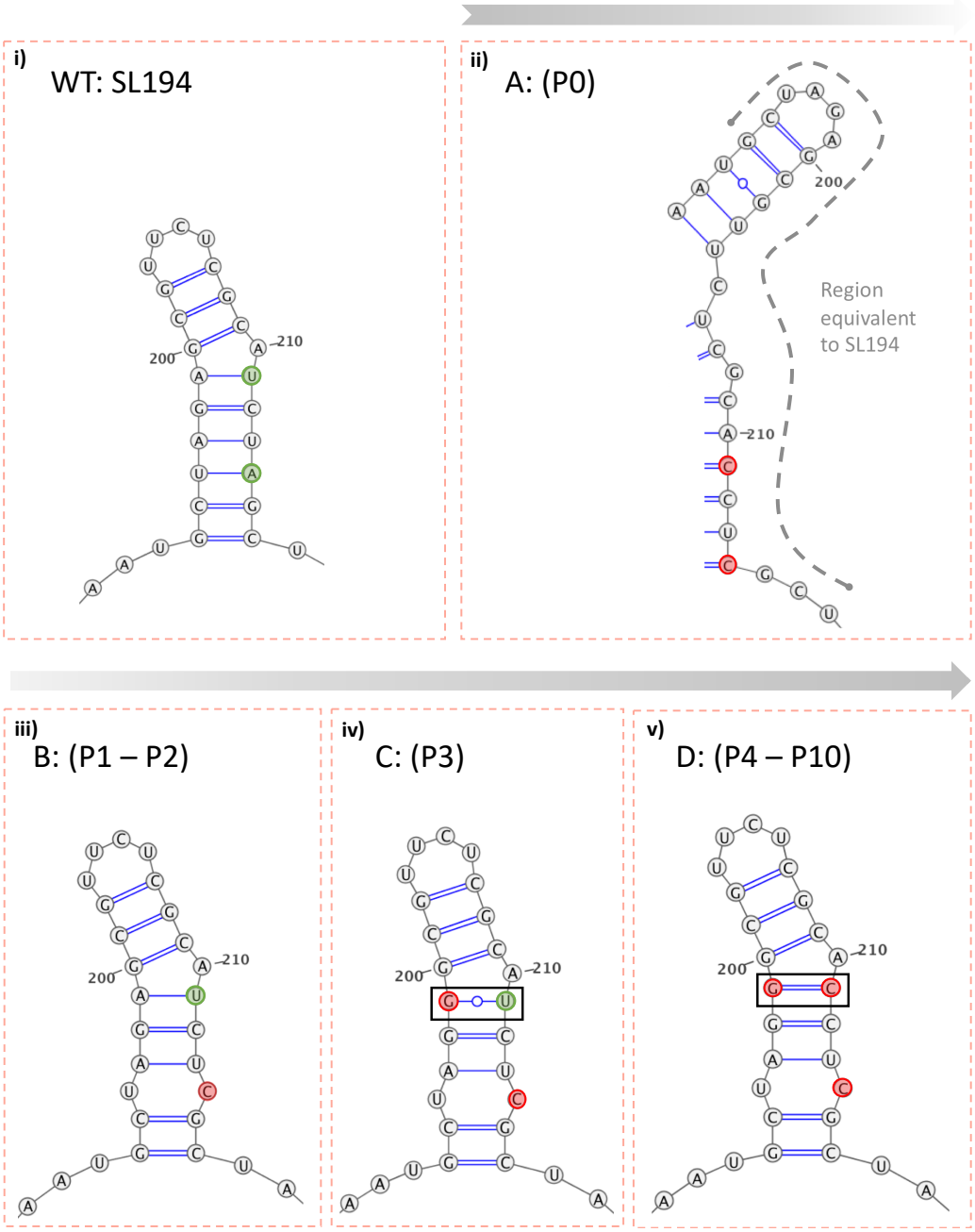


Figure 4.19: Incremental restoration of SL194 structure during passaging. i) The wild-type structure of SL194 is displayed, with the sites targeted for mutation in P0 highlighted (green circles). **ii-v)** The region equivalent to SL194 is denoted by a grey dotted line in the P0 predicted structure. All mutated (non- wild-type) nucleotides are displayed for passage sequences (red circles). Compensatory mutations which do not restore wild-type sequence but reform base-pairing in the stem associated with wild-type structure are boxed in black across the base-pair. Each panel A:D displays the UNAfold-predicted structure in the region for the passage sequence shown in the alignment. The order of sequential passages is denoted by a grey arrow.

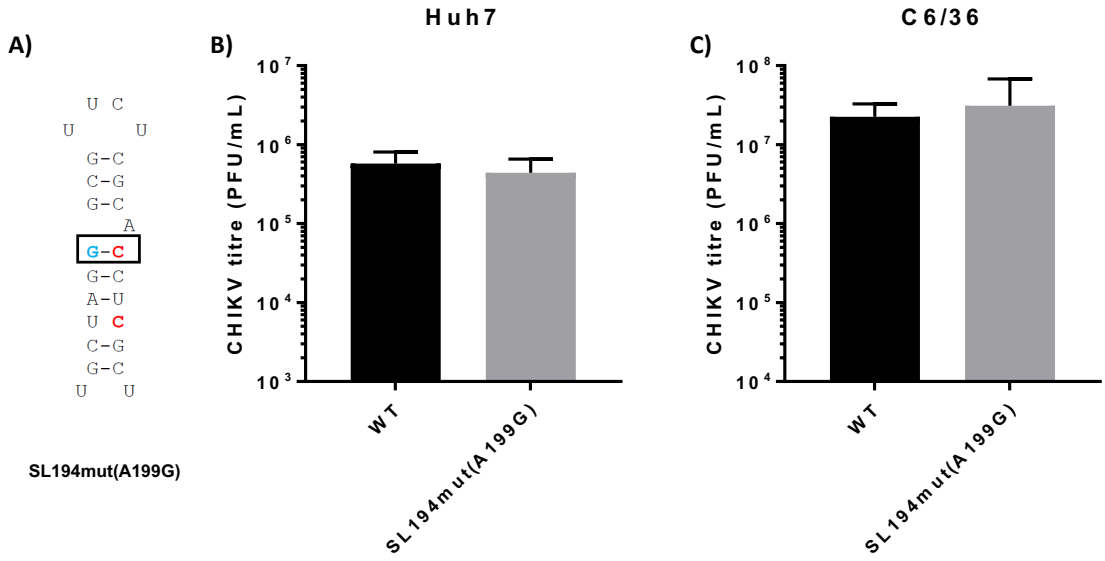


Figure 4.20: Phenotypic analysis of *SL194mut(A_{199}>G_{199}).* (A) Schematic of A_{199}>G_{199} compensatory mutation (blue) incorporated into *SL194mut* (mutations in red) to produce *SL194mut(A_{199}>G_{199})*. Replication phenotype of WT virus (black bar) compared to *SL194mut(A_{199}>G_{199})* (grey bar) following one-step growth assay in **(B)** Huh7 and **C)** C6/36 cells over 24 hours. Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.7 Passaging stem-loop mutants in *Ae. albopictus* cells leads to phenotypic reversion within 3 passages

In order to explore the potential routes of reversion in mosquito cells, mutant viruses were passaged repeatedly in C6/36 cells and assayed for phenotypic reversion, as previously carried out in mammalian cells (Figs 4.7 & 4.8). As expected, both *SL47mut* and *SL246mut* exhibited significantly reduced replication compared to wild-type virus following a one-step growth assay in C6/36 cells with P0 virus ($p \leq 0.01$) (Fig 4.21A). C6/36 cells were then infected with P0 virus at MOI=1 for 48 hours, before $1/10^{th}$ of the supernatant was transferred to a flask of uninfected cells. Passaging continued for a total of 3 passages in C6/36 cells. Each passage was titred between P0 and P3 to sample for potential phenotypic reversion. Significantly inhibited replication phenotypes for both stem-loop mutants were observed at P1 and P2. At P3, there was no significant reduction in viral titre in the passage supernatant relative to wild-type for *SL47mut* or *SL246mut,* suggesting potential phenotypic reversion. In order to confirm this, one-step growth assays were carried out with P3 virus (Fig 4.21B). No significant difference in viral fitness was observed for either of the stem-loop mutants compared to wild-type, demonstrating that phenotypic reversion had occurred.

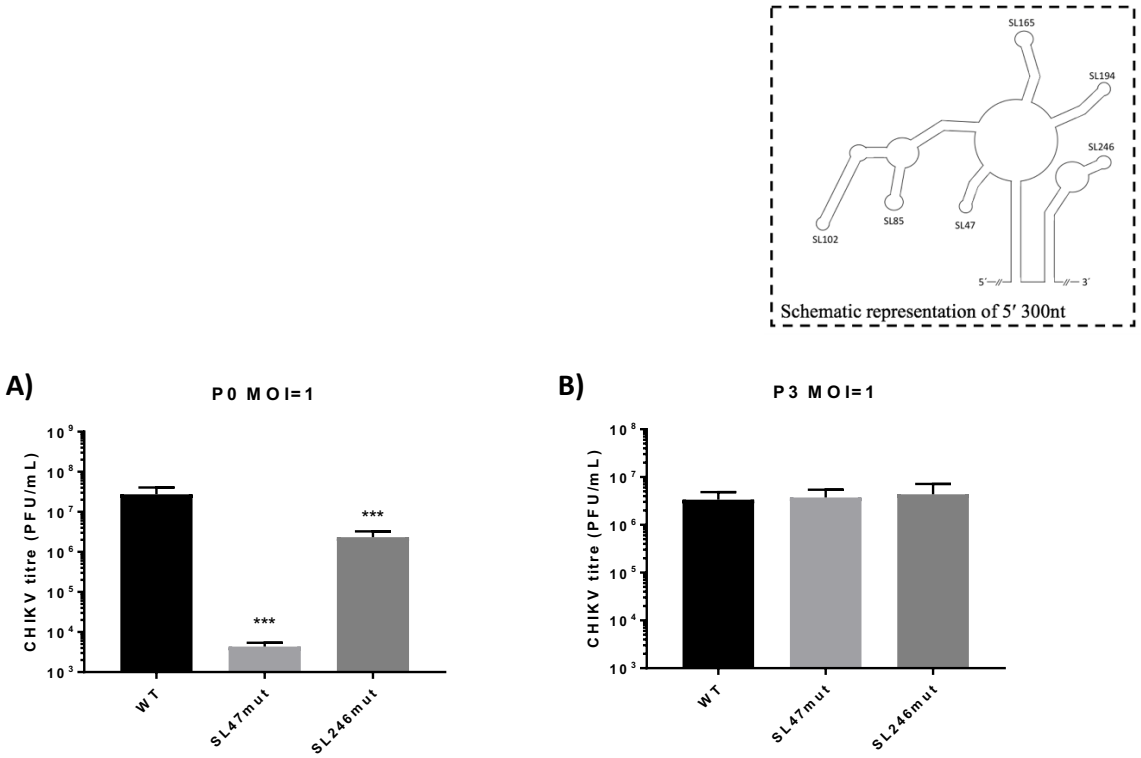
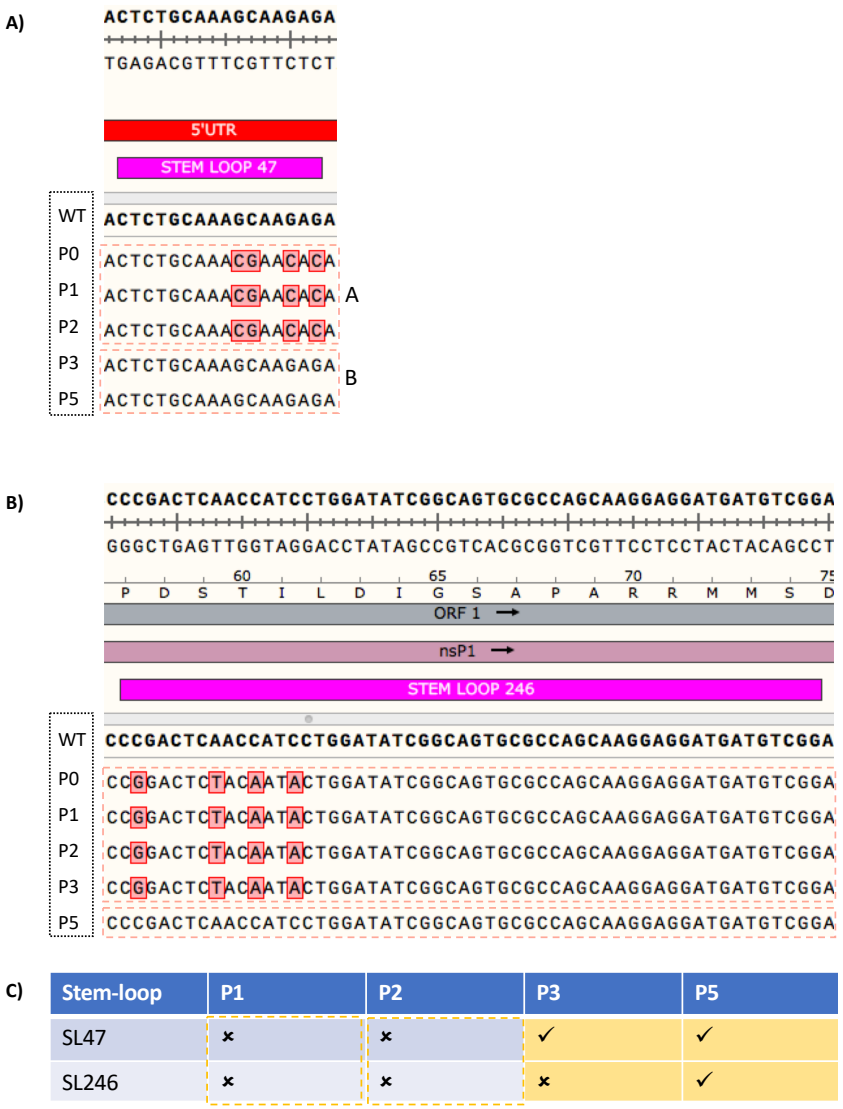


Figure 4.21: Mosquito cell passaging leads to phenotypic reversion within three passages. Replication phenotype of WT CHIKV (black bar) compared to virus bearing mutations predicted to destabilise the heteroduplex stem RNA structures (grey bars) following one-step growth assay of C6/36 cells over 24 hours using **(A)** P0 and **(B)** P3 virus. A schematic representation of the 5′ structure of the CHIKV genome is shown for reference (top right). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***), Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

4.2.8 Escape mutants accumulate necessary revertant mutations over 5 passages in *Ae. albopictus* cells

In order to determine the genotypic basis of reversion in P3 *SL47mut* and *SL246mut* viruses, RNA was extracted from infected C6/36 cells. The first 1 kb of each P3 viral genome was reverse transcribed, amplified by PCR and sequenced by Sanger sequencing in duplicate. The *SL47mut* escape mutant exhibits direct reversion at P3 (Fig 4.22A). Interestingly, unlike in mammalian cells, *SL47mut* escape mutants in C6/36 cells occurred over a single 48-hour passage with no detectable genetic intermediate.

Although *SL246mut* had phenotypically reverted to wild-type levels of CHIKV replication by P3 (Fig 4.21E), each of the 4 synonymous substitutions in the duplex stem were still present and no compensatory mutations were observed in the stem-loop (Fig 4.22B). Further experiments were carried out using the P3 *SL246mut* escape mutant to explore the potential for pseudo-reversion within the non-structural protein-encoding ORF-1, which spans the first 6 kb of the CHIKV genome. In order to provide enough viral RNA for analysis of this length of the genome, the infection was extended from 48 to 96 hours, effectively 2 further passages. Unfortunately, while this successfully provided high quantities of viral RNA, it allowed direct reversion to occur during the growth period. The sequence of ORF-1 in viral RNA extracted from this infection, termed P5 due to the length of the growth period, was identical to the wild-type sequence throughout. This included direct reversion of SL246 and presumably any potential pseudo-reversions in ORF-1. These results are summarised according to the passage at which direct genetic reversion was detectable for *SL47mut* and *SL246mut* (Fig 4.22C).



C)

Stem-loop	P1	P2	P3	P5
SL47	×	×	✓	✓
SL246	×	×	×	✓

Figure 4.22: Development of reversion in SL47 and SL246 from passage 0 to passage 5 (P0-P5) in C6/36 cells. The passage sequences P0-P5 of **A)** *SL47mut* and **B)** *SL246mut* are aligned below the WT sequence. The original mutations of *SL47mut* and *SL246mut* (red boxes) are shown in P0. Passage sequences are grouped (dotted red lines) according to the sequence changes occurring over the course of the experiment (groups containing the same sequence are labelled A and B). **C)** Development of reversion in C6/36 cells over 5 passages (P0-P5). Ticks represent direct reversion to wild-type sequence. Crosses represent no reversion. Yellow dashed lines denote sampling for phenotypic reversion in passage supernatant, yellow filled boxes denote phenotypic reversion during one-step growth assay.

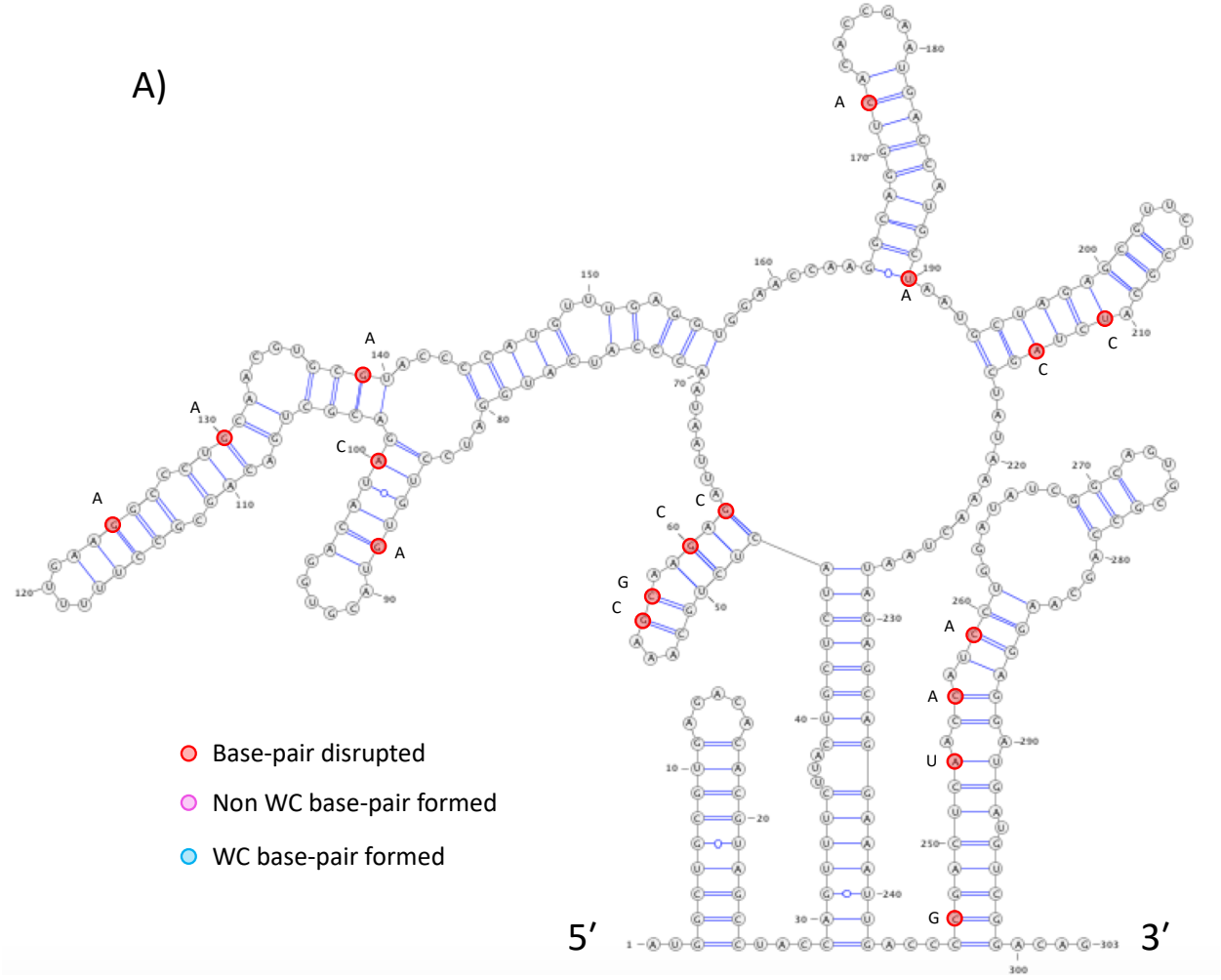
4.3 Discussion

The previous chapter demonstrated that stem-loop 47 (SL47) within the 5′ UTR of the CHIKV genome enhances replication of the viral genome in human- and *Ae. albopictus*-derived cells. In contrast, stem-loops within the adjacent nsP1-encoding region (SL85, SL102, SL165, SL194 and SL246) enhance CHIKV genome replication in a host-dependent manner (Figs 3.5 & 3.6). In this study, the relative importance of primary RNA sequence and secondary structure of each stem-loop was investigated in the appropriate host(s), by reverse genetic analysis and serial passage, to investigate evolutionary routes of escape.

4.3.1 Structure-led reverse genetic analysis

The secondary structure of each duplex stem was found to be integral to enhancement of CHIKV genome replication (Figs 4.2 & 4.3). Compensatory mutations, designed to restore base-pairing in the stem without returning to the wild-type RNA sequence, were introduced into infectious virus and replicon constructs (Fig 4.1), allowing the effects of RNA sequence and structure in the stem to be distinguished. Rescue of wild-type levels of CHIKV genome replication by compensatory mutants demonstrated that the stem-loops enhance CHIKV genome replication in a structure-dependent manner (Figs 4.2 & 4.3). However, mutations designed to disrupt RNA structure do so by altering primary RNA sequence. Primary nucleotide sequence has been demonstrated to affect the stacking energy of RNA and thus the stability of secondary structures with equivalent base-pairing. Thermal denaturation studies of a stable RNA hairpin from the bacteriophage T4 demonstrated that the orientation (C:G or G:C) of the closing base-pair determined the stability of the structure (403). Compensatory mutations in this study do not identically reproduce wild-type structure in the stem-loops, rather maintain the number of base-pairs. This disparity may account for the preference for direct reversion to wild-type sequence observed in escape mutants as opposed to compensatory mutation (Fig 4.10), as well as the relative reduction in CHIKV titre for compensatory mutants compared to WT (Fig 4.2). In addition, compensatory mutations in SL165 and SL246 replaced Watson-Crick base-pairs with G:U base-pairs (Fig 4.23A&B). Non-Watson-Crick G:U base-pairs result in less stable RNA secondary structures compared to G:C/A:U pairs. Therefore, *SL165mut-Comp* and *SL246mut-Comp* are not ideal comparisons for the wild-type

structures. However, maintaining synonymy within nsP1 placed limitations of the design of compensatory mutations. Despite the relative instability of G:U base pairs, the degree of compensation observed for *SL165mut-Comp* and *SL246mut-Comp* was not significantly different to that of the other compensatory mutants (Figure 4.2). Thus, although the strength of G:U pairing is lower, no significant effect of decreased stability was observed for compensatory mutants which utilised G:U.


A)
Base-pair disrupted
Non WC base-pair formed
WC base-pair formed
5′
3′

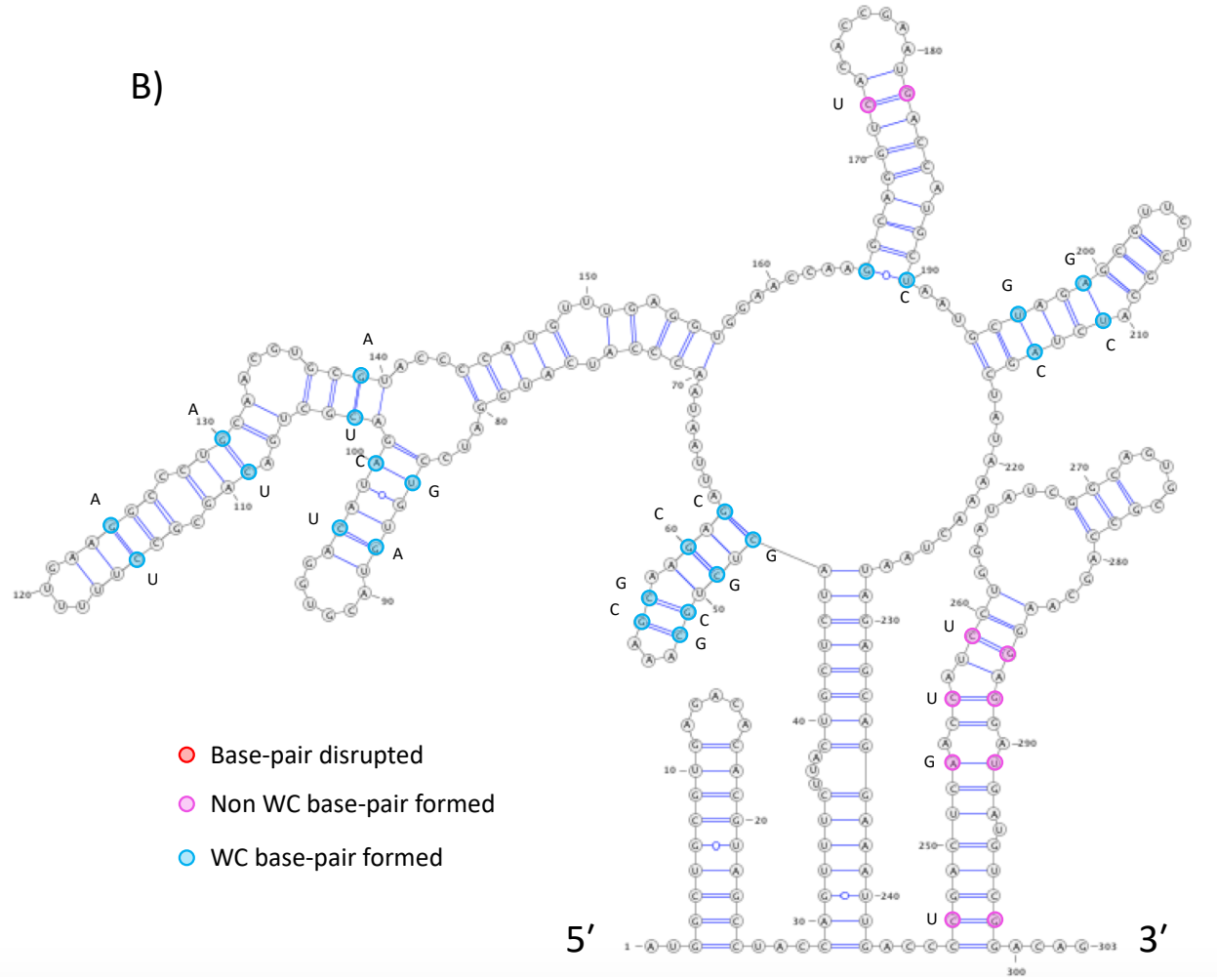


Figure 4.23: Types of base-pair formed by mutation in the positive strand. UNAfold prediction of wild-type positive-strand RNA structure overlaid with (A) disruptive and (B) compensatory mutations denoted by coloured circles which depict the type of alteration made in the positive strand structure: red circles indicate that base-pairing is absent or disrupted, pink circles indicate a non-Watson-Crick base-pair formed by an introduced mutation, turquoise circles indicate a Watson-Crick base-pair formed by an introduced mutation. The structure is labelled 5′ to 3′. Base-pairing in the predicted wild-type structure is denoted by blue lines: one line for A:U pairs, two lines for G:C pairs and a line with a white circle for G:U pairs.

As discussed in Chapter 3, synonymous mutations alter codon usage and dinucleotide frequencies, thus the replication phenotypes observed may be the result of reduced translational efficiency or enhanced immune recognition of the genome - rather than disruption of RNA structure. Compensatory mutation typically increased the number of CpG and UpA dinucleotides (Fig 4.25). For example, *SL85mut-Comp* contained a greater number of CpG and UpA dinucleotides (+2 CpG and +3 UpA relative to WT) than *SL85mut* (+1 CpG and +1 UpA relative to WT). In spite of increased CpG and UpA dinucleotide frequencies, *SL85mut-Comp* rescued wild-type levels of CHIKV genome replication. This suggests that CpG and UpA frequencies are not responsible for the replication phenotype observed for *SL85mut*. Overall, changes to dinucleotide frequency did not correlate with replication phenotype. *SL165mut* and *SL165mut*-Comp contain the same frequency of CpG and UpA as wild-type SL165, but *SL165mut* exhibited a replication defect while *SL165mut-Comp* did not. *SL102mut-Comp* contains 5 more UpA dinucleotides than WT but this had no significant effect on replication.

In addition to dinucleotide content, the role of codon efficiency must be considered. Codon efficiency is known to affect translation and stability of host and viral mRNA (418,419). Synonymous codons are not used at equal frequency, as demonstrated in *Escherichia coli, Saccharomyces cerevisiae* and vertebrate-derived muscle and liver tissue (420–422). Codon usage bias refers to the preference displayed by different organisms for particular synonymous codons. Substitution of codons with vastly different frequencies in the host genome can impact translational efficiency (401). Codon deoptimisation has been shown to impact CHIKV replication, attenuating the virus (402). Codon efficiency and dinucleotide usage in the compensatory mutants is similar to the disruptive mutants, thus the restoration of wild-type replication is likely due to reformation of the stem-loop structures.

As discussed in Chapter 3, UNAfold modelling predicts that the corresponding structure of the region of interest in the negative strand differs greatly from that of the positive strand (Fig 3.15). A structure equivalent to SL194 is present, as well as less homologous versions of SL3, SL102 and SL246. In contrast, homologues of SL47, SL85 and SL165 are not predicted to form in the negative strand. As compensatory mutations were designed based on a structural model of the positive sense genomic copy of the CHIKV genome, rescue of wild-type replication by this

approach suggests the importance of stem-loops in the positive strand, rather than unforeseen disruption of potential functional elements in the negative strand intermediate of the viral genome. For example, compensatory G:U pairing in *SL246mut-Comp* would represent C:A at the same position in the negative strand, which would not form a base-pair (Fig 4.1). Given that *SL246mut-Comp* rescues wild-type levels of CHIKV genome replication, it is likely that SL246 in the positive strand is responsible for the enhancement of genome replication in *Ae. albopictus,* since the corresponding structure in the negative strand could not have reformed via these compensatory mutations. However, *SL102mut-Comp* and *SL194mut-Comp* reform Watson-Crick base-pairing in the negative strand (Fig 4.24). Therefore, despite the major differences between the positive- and negative-strand secondary structure in this region, compensatory effects due to restoration of negative strand structures for SL102 and SL194 cannot be ruled out based on the evidence presented here. Further SHAPE-constrained thermodynamic modelling could be carried out to determine the secondary structure of the negative strand 3′ terminus, which corresponds to the 5′ region of the positive strand.

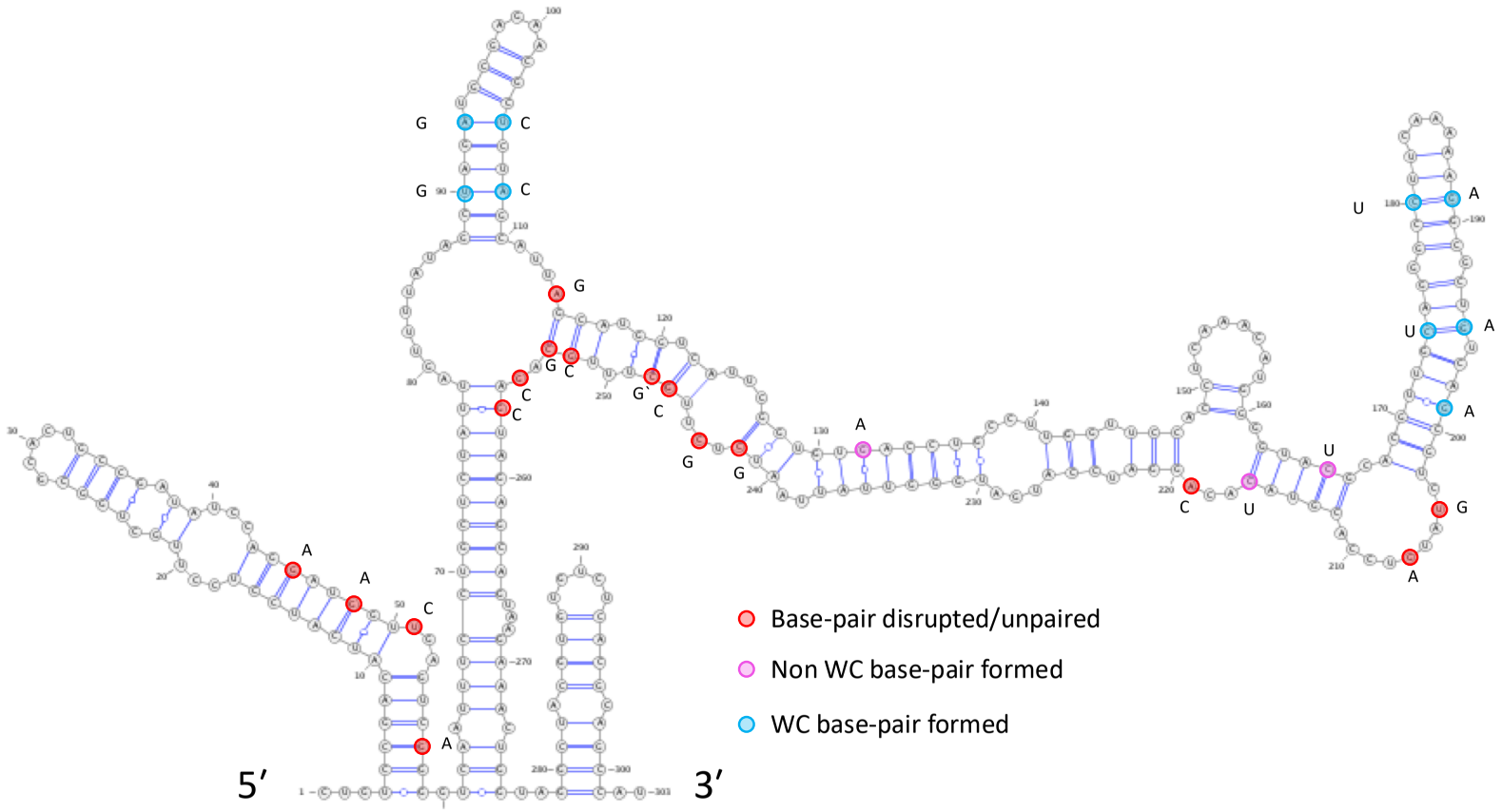


Figure 4.24: Types of base-pair formed by compensatory mutation in the negative strand. UNAfold prediction of secondary structure in the 3′ 300nt of the wild-type negative strand intermediate. Compensatory mutations introduced are displayed and denoted by coloured circles which depict the type of alteration made in the negative strand structure: red circles indicate base-pairing remains absent or disrupted, pink circles indicate a non-Watson-Crick base-pair formed by an introduced mutation, turquoise circles indicate a Watson-Crick base-pair formed by an introduced mutation. The structure is displayed 5′ to 3′, whereby the 5′-most structure corresponds to SL246 (the 3′-most structure in the positive strand). Base-pairing in the predicted wild-type structure is denoted by blue lines: one line for A:U pairs, two lines for G:C pairs and a line with a white circle for G:U pairs.

Stem-loop mutant	CpG	UpA
SL47mut	0	0
SL47mut-Comp	0	0
SL85mut	1	1
SL85mut-Comp	2	3
SL102mut	-2	2
SL102mut-Comp	-3	5
SL165mut	0	0
SL165mut-Comp	0	0
SL194mut	1	-2
SL194mut-Comp	2	-3
SL246mut	0	2
SL246mut-Comp	0	1

Figure 4.25: Changes in CpG and UpA dinucleotide frequency in compensatory structure mutants. Net change in CpG and UpA dinucleotide frequencies in each of the stem-loop mutants alongside the corresponding compensatory mutant, expressed as the net number of dinucleotide pairs introduced (positive) or removed (negative) by the mutations displayed in previous chapters (Figs 3.4A & 4.1).

	Percentage change in codon efficiency	
	Human	*D. melanogaster*
SL47mut	-	-
SL47mut-Comp	-	-
SL85mut	-2	5
SL85mut-Comp	-7	17
SL102mut	-13	-2
SL102mut-Comp	-7	5
SL165mut	-8	1
SL165mut-Comp	4	8
SL194mut	16	-4
SL194mut-Comp	4	-2
SL246mut	-15	-3
SL246mut-Comp	-11	-7

Figure 4.26: Changes in codon efficiency in compensatory structure mutants. Net change in codon efficiency for each stem-loop mutant and its corresponding compensatory mutant in human- and *D. melanogaster*-derived cells. Changes to codon efficiency are displayed as a relative percentage increase (positive) or decrease (negative) in codon efficiency in each organism upon mutation. Each mutant contains a minimum of two substitutions relative to WT and thus each percentage is an average of the change induced by all substitutions per mutant. Standard deviation for each average was equal to or less than 10%. Codon efficiency is measured as codon frequency in the genome of each organism, as drawn from the codon frequency tool from Genscript (405). SL47 is in the untranslated region, thus codon usage is null.

4.3.2 SL47

SL47mut exhibits a larger phenotype than the stem-loop mutants in the nsP1-coding region. It is possible that SL47 represents a uniquely important stem-loop with a larger biological effect than the adjacent stem-loops. However, it must also be considered that *SL47mut* contains the greatest number of mutations. 25% of the nucleotides in SL47 were mutated in *SL47mut*, the highest percentage of any of the stem-loops. In part, this was due to the position of SL47 within the 5′ UTR which allowed for extensive disruption without altering the protein-coding sequence. Nonetheless, the larger effect size may be partially due to greater disruption. One argument against this is that *SL47mut-Comp*, which contains only 50% of the original SL47 sequence, was able to replicate to wild-type titres. Regardless, the higher number of mutations in SL47 likely played a large role in the temperature sensitivity of *SL47mut-Comp*. Only *SL47mut-Comp* demonstrated temperature-sensitivity during CHIKV replication, suggesting strict thermodynamic constraints on folding structure. *SL47mut-Comp,* although it contains four sequence substitutions relative to *SL47mut* and eight relative to wild-type SL47, maintains four C:G base-pairs in its stem. Given that the number of hydrogen bonds in the stem remains equivalent to wild-type, it is interesting that *SL47mut-Comp* exhibits such dramatic temperature sensitivity. De-stabilisation of the compensatory mutant at higher temperatures may be explained by changes in stacking energy and other constraints on RNA structure formation (403), as demonstrated by the predicted free energy of folding at 37°C relative to the wild-type structure (wild-type $\Delta G = -2.9$ kcal/mol and *SL47mutComp* $\Delta G = -2.3$ kcal/mol). The temperature-sensitivity of *SL47mut-Comp* demonstrates that the formation of SL47 is under strict sequence constraints, despite sequence-independent functionality (Figs 4.2C & 4.3C). The wild-type sequence of SL47 is therefore an evolutionarily stable element under strong purifying selection in mammalian cells; sequence variation causes a steep fitness decrease due to temperature-sensitivity of the structure. This sequence constraint may be an incidental by-product of thermodynamic stability constraints or may be functional as the preferred binding context of a *trans*-activating factor.

Interestingly, during passage of *SL47mut* in Huh7 cells, partial sequence reversion was fixed in the population prior to any structure restoration, at P6 and P7 respectively (Fig 4.11). There is selection for $G_{56}C_{57}$ prior to structure reformation, which suggests a preference for this sequence

independent of stable stem-loop formation. However, it is possible that the P6 mutant containing $G_{56}C_{57}$ is capable of transiently forming the apical region of SL47. While most of the *in silico* predicted structures of P6 *SL47mut* in the context of the entire structured region do not favour the formation of the apical region without a stable base, analysis of the P6 *SL47mut* sequence in isolation predicts a structure containing three base pairs $U_{50}G_{51}C_{52}$:$G_{56}C_{57}A_{58}$ with a terminal loop $A_{53\text{-}55}$ (Fig 4.27B). The UNAfold-predicted free energy of folding for this structure is very small (ΔG= -0.20kcal/mol), indicating that the structure is unstable at 37°C. Nevertheless, the potential ability of the P6 *SL47mut* mutant to transiently form the structure may represent a fitness increase relative to P0-P5 *SL47mut*, which is not predicted to form any stable secondary structure ($\Delta G \geq$ 0kcal/mol). The potential formation of the apical region of SL47 in P6 *SL47mut* could be investigated via SHAPE mapping. Furthermore, the replication phenotypes of passage mutants P6 and P7 during one-step growth assays of Huh7 cells could shed light on the level of rescue achieved during the process of reversion. In particular, the replication phenotype of P7 *SL47mut* merits further exploration, given the temperature-sensitivity of *SL47mut-Comp*. P7 *SL47mut* forms ¾ of the hydrogen bonds present in *SL47mut-Comp* and maintains the wild-type orientation of these bonds. Comparison of the replication phenotypes of P7 *SL47mut* and *SL47mut-Comp* would determine the relative importance of number and orientation of bonds in the duplex stem.

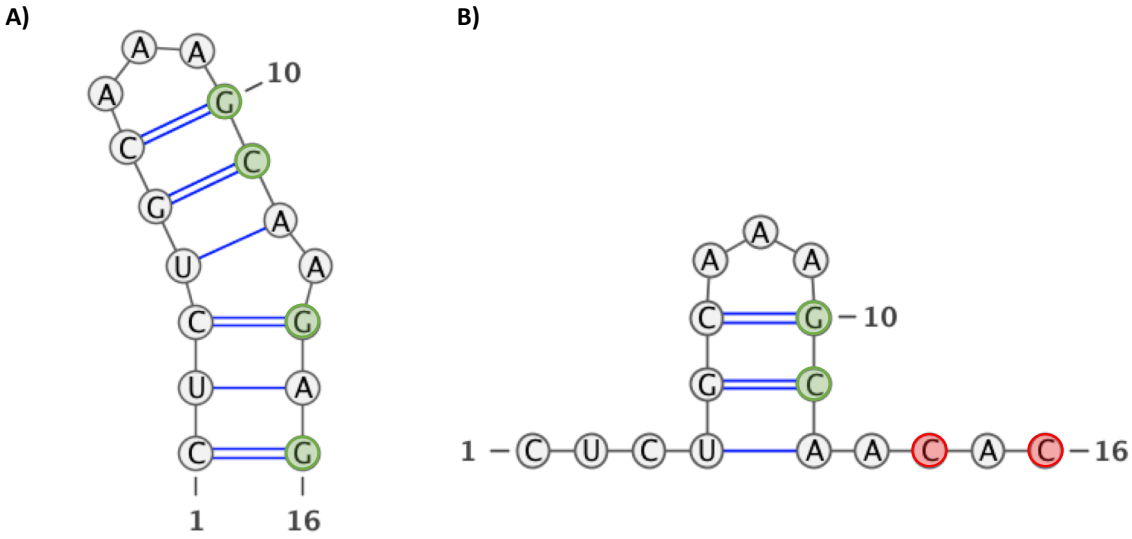


Figure 4.27: The apex of SL47 reforms at P6. UNAfold-predicted structure of **(A)** *SL47mut* P6 and **(B)** wild-type SL47. Nucleotide 1 in the prediction corresponds to nt 47 in the CHIKV genomic RNA. Sites of mutation in *SL47mut* (P0) are circled and coloured according to sequence: wild-type (green) and mutant (red).

4.3.3 SL85 and SL102

In Chapter 3, SHAPE-constrained thermodynamic analysis demonstrated that the RNA structure of CHIKV immediately downstream of the AUG start codon differs from that previously observed for the New World alphaviruses SINV and VEEV. SHAPE mapping of SINV and VEEV has demonstrated a single, long RNA element (13)– deletion of which does not inhibit SINV genome replication (350). In contrast, the homologous region of the CHIKV genome was demonstrated to contain two distinct RNA replication elements (SL85 and SL102 (Fig 3.2A), both of which are involved in efficient genome replication in human-derived cells but not those from *Ae. albopictus* (Fig 3.4). In the case of SL85, the *in silico* thermodynamically predicted structure, SHAPE mapping and reverse genetic analysis were not in close agreement (Fig 4.2). Raw SHAPE reactivity data for SL85 was not consistent with the structure predicted by combined thermodynamic and reverse genetic analysis, suggesting that this region of the genome may be structurally dynamic and able to form alternative interactions. Structure-led reverse genetic analysis, destabilising SL85 and then restoring base-pairing with compensatory substitutions, demonstrated that SL85 is essential for efficient CHIKV genome replication in human-derived cells and functions through a structure-dependent mechanism (Figs 4.2A & 4.3A). The ability of *SL85mut-Comp* to rescue wild-type CHIKV replication demonstrates that the structure forms during the CHIKV lifecyle *in vivo*. Further investigation of the spatio-temporal dynamics of SL85 formation may reveal at which stage of the viral lifecycle SL85 exerts its effects and the location of its formation within the cell i.e. in membrane-bound replicative complexes. The potential for alternative structure formation in this region is investigated in Chapter 5.

The substitutions made in *SL85mut* and *SL102mut* each result in a single, longer RNA element similar to that observed in SINV and VEEV (Figs 4.11 & 4.13; (13)). As demonstrated in Chapter 3, *SL85mut* and *SL102mut* exhibit replication defects relative to the wild-type structure, indicating that this region functions differently in CHIKV than in VEEV and SINV. During passage of *SL85mut* and *SL102mut* in Huh7 cells, direct reversion to wild-type sequence was observed over 6 and 7 passages respectively (Figs 4.12-15). No intermediate changes were observed for *SL85mut*. *SL102mut* exhibited an intermediate structure at P6, which is conformationally distinct to that of wild-type SL102 (Fig 4.15). The P6 intermediate structure restores the single-stranded conformation of nt 118-121 ($U_{118}U_{119}U_{120}G_{121}$) - which represents the terminal loop of wild-type

SL102. Interestingly, this region was heavily mutated during reverse genetic analysis of the terminal loop sequence of SL102 (Fig 4.4) and found to be dispensable to CHIKV genome replication (Figs 4.5 & 4.6). Thus, it is unlikely that the A_{130}>G_{130} reversion detected in the population during P6 is due to selection for the unpaired terminal loop sequence UUUG, rather than the primary nucleotide sequence. One element of the P6 structure, which bears similarity to the wild-type structure is the apical region. While the lower portion of the P6 structure diverges greatly from both the wild-type and P0 structures, the apical region of P6 appears far closer to the wild-type structure (Fig 4.28); the terminal loop is smaller (7 nt, 5 nt and 4 nt in P0, P6 and wild-type respectively) and the first bulge becomes smaller (4 nt, 2 nt and 2 nt in P0, P6 and wild-type respectively). Overall, P6 represents a truncated form of the apex of SL102.

In light of the structure-dependent replication phenotypes observed for SL102, partial restoration of the apical structure may represent a sufficient increase in fitness to allow predomination of the P6 mutant. A comparison of the binding partners of P6 *SL102mut* and WT SL102 may allow the recognition of interacting protein partners, which recognise the apical structure. The replication phenotype of P6 *SL102mut* would also provide insight into the relative importance of an elongated SL102 as compared to a truncated stem of similar apical dimensions.

4.3.4 SL165 and SL194

SL165mut and *SL194mut* each selected compensatory mutations during passage in Huh7 cells. The compensatory mutations detected during passaging recapitulated substitutions made during the design of compensatory mutants, A_{190}>C_{190} in SL165 and A_{199}>G_{199} for in SL194 (Fig 4.1). *SL165mut* acquired a compensatory A_{190}>C_{190} mutation during P5, which stabilised the base of the structure by formation of a G_{165}:C_{190} pair (Fig 4.17B) in place of a wild-type G_{165}:U_{190}. During P6, direct reversion was observed, producing G_{165}:U_{190} (Fig 4.17C). There appears to be selection for U_{190}, either as a sequence component or due to preference for weaker bonding at the base of the structure. The preference for weaker bonding at this site could be investigated by phenotypic analysis of the single substitution U_{190}>C_{190}. In lieu of this, compensatory mutant analysis can provide information about the effect of the A_{190}>C_{190} mutation, as this mutation is one of two present in the compensatory mutant, *SL165mut-Comp* (Fig 4.1). *SL165mut-Comp* demonstrated increased CHIKV genome replication relative to *SL165mut* ($p \leq 0.05$) with no significant difference

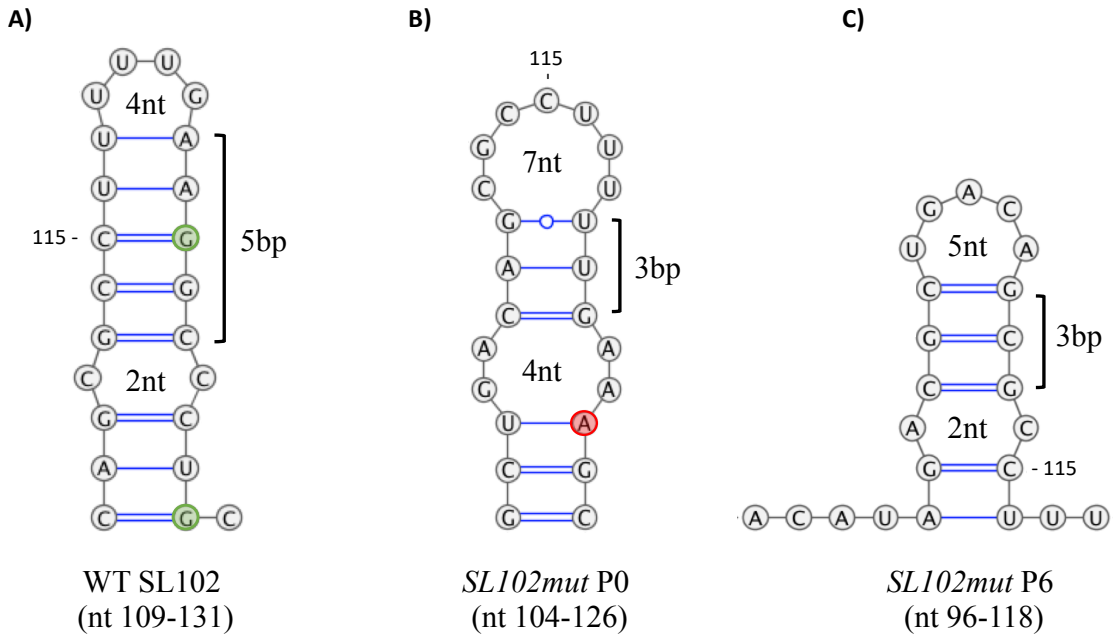


Figure 4.28: The apical structure of SL102, but not sequence, reforms at P6. UNAfold-predicted structure of the apex of SL102 in **(A)** WT SL102. **(B)** *SL102mut* P0 and **(C)** *SL102mut* P6. Nucleotide 115 is labelled in each structure for comparison. Sites of mutation in each structure are circled and coloured according to sequence: wild-type (green) and mutant (red).

to wild-type SL165 (Fig 4.3). Thus, it can be concluded that the sequence U_{190} is not required to enhance CHIKV genome replication, since *SL165mut-Comp* contains C_{190} and can replicate to wild-type levels. The destabilised apex of P5 *SL165mut,* which contains the disruptive mutation C_{172}>$A_{172,}$ may select for the stronger G_{165}:C_{190} pair in P5 *SL165mut*. Reversion of the C_{172}>A_{172} mutation in P6 *SL165mut* would remove the selection pressure for increased hydrogen bonding in the base of the stem and allow selection for G_{165}:U_{190} to prevail. In this way, it appears that the compensatory mutation exhibited by P5 *SL165mut* acts as an intermediate stage allowing direct reversion via incremental fitness increase.

Uniquely, passage of *SL194mut* in Huh7 cells resulted in a fixed compensatory mutation, which predominated in the population by P4. During P1, one of the two substitutions in *SL194mut* directly reverted to the wild-type sequence (C_{211}>U_{211}), reforming the A_{199}:U_{211} base-pair of SL194 (Fig 4.19B). This single substitution was sufficient to reform the terminal loop, apical region of the stem and A_{210} bulge. The other mutation in *SL194mut*, A_{214}>C_{214}, remained throughout all passages, representing the only disruptive mutation detectable by P10 in any of the mutants passaged in mammalian cells. During P3, a compensatory mutation altered the A_{199}:U_{211} base-pair to G_{199}:U_{211} via synonymous substitution (Fig 4.19C). G_{199} represents a substitution made during the design of *SL194mut-Comp.* Fixation of G_{199} during P3 is interesting, as it replaces the wild-type A:U base-pair with a G:U base-pair thus forming a structure with a less negative free energy of folding (P2 ΔG= -8.00kcal/mol and P3 ΔG= -7.10kcal/mol). Non-Watson-Crick G:U base-pairing has been shown to alter helical geometry in certain RNA structures (423). It is possible that the U_{196}/C_{214} bulge alters the steric context of the stem-loop in such a manner that a G:U base-pair is favourable but the free energy data does not support this. Codon usage frequencies are also unable to explain selection of G_{199}, since the usage of AGG and AGA codons is equivalent in humans, at 20% total arginine tRNAs per codon. Another possibility is that the U_{196}/C_{214} bulge and the A_{199}:U_{211} base-pair in combination form a sequence- or structure- recognition signal that is unfavourable to CHIKV replication. This could explain selection for G:U, which would alter the sequence context of the region as well as the structure around the base-pair. A final substitution occurred during P4, which was present in all passages thereafter (Fig 4.19D). This substitution reversed that made during P1, reverting to the *SL194mut* sequence at nt 211 (U_{211}>C_{211}). The substitution made during P4 produces a G_{199}:C_{211} pair. In a similar manner to the sequential

changes in *SL165mut*, it can be hypothesised that a greater number of hydrogen bonds in one region of the stem offsets the ongoing destabilisation in another region. Although a single substitution would be sufficient to produce P4 *SL194mut* from its P0 counterpart, the route of reversion observed was less direct.

Passage sequences presented in this study represent the predominant sequence in the region of interest at each passage, as determined in duplicate following RT-PCR amplification and Sanger sequencing. Due to the limitations of Sanger sequencing for analysis of quasispecies, variation present within each passage is not presented or quantified here. Sampling of passage supernatant did not suggest full phenotypic reversion for P5 *SL194mut* relative to WT, but the compensatory G_{199} mutation was detected during P4. Each passage sequence P4-P10 contained the compensatory mutation G_{199} and the original mutations of *SL194mut*, indicating that these mutations were present in the population. When this set of mutations was incorporated into CHIKV_IC via site-directed mutagenesis and used to infect cells of human and *Ae. albopictus* origin, wild-type titres of CHIKV were recovered (Fig 4.20). This demonstrates that the compensatory substitution G_{199} is sufficient to rescue *SL194mut* replication in human-derived cells, and that the structure of the apical region is responsible for SL194-mediated enhancement as opposed to the base of the stem.

4.3.5 Passaging in *Ae. albopictus*

Passaging in mammalian cells led to direct genetic reversion in the majority of stem-loop mutants through the accumulation of revertant mutations, incrementally reforming the wild-type structure. In one case, a compensatory revertant predominated. In contrast, passaging of *SL47mut* and *SL246mut* in mosquito cells led to direct genetic reversion without detectable intermediate mutations (Fig 4.22A & B). During passaging of *SL47mut,* direct genetic reversion of four mutations was detected during a single 48-hour passage, P3 (Fig 4.22A), which corresponded to phenotypic reversion during one-step growth assay in C6/36 cells with P3 virus (Fig 4.22B). Interestingly, the phenotypic reversion observed at P3 for *SL246mut* (Fig 4.22B) did not correspond to direct or compensatory reversion in SL246 when intracellular viral RNA was extracted from infected cells (Fig 4.22B).

Previous passage studies with RNA structural mutants of SINV and VEEV have demonstrated the development of compensatory 'pseudorevertant' mutations in non-structural proteins nsP1-4, which restore interactions with RNA (350,358). In order to investigate the genetic basis of phenotypic reversion in P3 *SL246mut* virus, the non-structural protein-encoding ORF-1 was sequenced in 1 kb overlapping fragments. A contiguous sequence of ORF-1 from P3 *SL246mut* did not reveal pseudorevertant mutations in ORF-1. Phenotypic reversion in P3 may be due to pseudoreversion of a structural protein encoded by ORF-2, restoring an interaction between a structural protein and RNA stem-loops, or by an unrelated beneficial mutation whereby increased infectivity or egress compensate for decreased genome replication. During the 2005-6 epidemic in La Réunion, mutation in structural proteins resulted in increased CHIKV replication in *Ae. albopictus*; virus extracted from infected mosquitoes contained the A226V mutation in E1 and replicated to higher titres in *Ae. albopictus* than previously documented (45). However, the ORF-1 sequence exhibited direct genetic reversion within SL246 (Fig 4.22C). Larger quantities of P3 viral RNA for analysis were generated by extension of the infection period from 48 to 96 hours, prior to extraction from infected cells, which may have allowed the development of direct reversion. The ORF-1 sequence was termed P5 due to the elongated period of infection. The presence of direct genetic reversion within SL246 of P5 *SL246mut* limits the conclusions that can be drawn about pseudoreversion in P3, as further mutation clearly occurred during P5. Pseudorevertant mutations in ORF-1 may have directly reverted over the course of P5 and thus not have been detected during sequencing. The speed of reversion in *Ae. albopictus*-derived cells presents a technical challenge to the accurate genotyping of phenotypic revertants in mosquito host-cells.

In contrast to the high genetic stability seen in mammalian cells, phenotypic reversion in mosquito cells for *SL47mut* and *SL246mut* occurred at P3 (Figs 4.21 & 4.22C) and genetic reversion at P3 and P5, respectively. Higher rates of CHIKV reversion have been observed in mosquito cell lines compared to mammalian-derived cells during passaging studies of nsP3 (197). During a study by Gao and colleagues, revertant mutations in nsP3 were detectable in C6/36 and U4.4 *Ae. albopictus* cells within 24 hours post-transfection with a CHIKV replicon, becoming fixed in the population by 72 hours post-transfection. Over the same time course, no reversion was detected in murine C2C12 cells. It is clear that rapid reversion is possible in *Ae. albopictus*-derived cells, highlighting

the need for dual-host studies of vaccine candidates for alphaviruses. Rapid reversion in mosquito cells may be the result of increased variation within the genetic pool or a greater number of replication cycles per passage. Mosquito-specific interactions, either protein:RNA or protein:protein interactions, could decrease the fidelity of the replicase during replication. An increase in error rate would increase variation within the quasispecies, widening the pool of genetic variants under selection. In this way, revertants would arise at a faster rate. Similarly, differing sites of replication within mosquito cells may favour rapid reversion by increasing genetic variability. In mammalian cells, replication complexes form at the plasma membrane and are internalised shortly after transcription of the negative strand. This may represent a checkpoint for successful replication which ensures cytopathic vacuoles contain only active replication complexes. In mosquito cells, spherules form directly on the membranes of cytopathic vacuoles. Thus, adjacent replication complexes may interact differently in mosquito cells, potentially increasing the likelihood of recombination between viral genomes or functional interactions between neighbouring spherules.

Alternatively, rapid reversion may be the result of generally increased replication rates and reduced cell death in mosquito cells. Mammalian cells typically undergo lysis or apoptosis as a result of infection within 48 hours. In contrast, mosquito cells do not undergo apoptosis as a result of CHIKV infection, remaining productively infected over long periods. Therefore, mosquito cells continuously replicate the virus over each passage with little cell death, enabling many more rounds of replication in an average population of cells over a given period. In addition, the establishment of productive CHIKV infection takes longer in mosquito cells but eventually produces a much greater quantity of infectious virus, possibly by the establishment of a greater number of replication complexes. The reasons for these differences are not well understood but could contribute to more rapid reversion.

4.3.6 Passaging approach

Previous studies of passaging in arboviruses have investigated single-host passaging compared to alternating passaging, as a method of analysing the evolutionary dynamics imposed by natural cycling between human and mosquito hosts. During these studies, single-host passaging often resulted in the accumulation of host-specific mutations and reduced replication in the bypassed

host (392,394). Multi-host passaging between primate and mosquito cell lines has been shown to limit rates of CHIKV evolution relative to single-host passaging; beneficial mutations were selected during alternating passaging of a 3′ UTR deletion mutant, which provided fitness increases in both host species (417). In this study, single-host passaging of stem-loop mutants demonstrated phenotypic reversion within a maximum of ten passages. The potential for, and relative speed of, reversion during alternating passaging could be investigated in future work.

The predominant viral genome sequence in each passage was determined by Sanger sequencing of amplified cDNA. RNA viruses exist as a 'quasispecies' during infection; a population of non-identical viral genome sequences exists in dynamic equilibrium where viable mutants arise at a high rate and are strongly selected against by competition with an optimal wild-type sequence (424). In light of such variation within the pool of viral RNAs following infection and particularly during passage of mutants, analysis of a single predominant sequence by Sanger sequencing provides limited information. This approach could be expanded by sequencing of multiple full-length cDNA clones. Alternatively, next-generation sequencing can provide a greater depth of sequence information within heterogeneous populations, by sequencing a vast number of viral genomes utilising a sequencing-by-synthesis approach such as pyrophosphate release (425). Whilst there is more genetic variation within the pool of viral RNAs than can be observed by sequencing the predominant genome, the chronological development of escape mutants was clear.

4.3.7 Single-stranded regions

In contrast to the other structures in the nsP1-encoding region, the function of SL194 as an RNA replication element during CHIKV genome replication is dependent on both the structure of the stem–loop and the primary sequence of the single-stranded terminal loop. A single synonymous C>U substitution within this unpaired region significantly inhibited both sub-genomic replicon and infectious virus replication (Figs 4.5A & 4.6). As this single substitution reproduced a UUUU terminal loop sequence conserved within the CSE elements of other divergent alphaviruses (including VEEV, SINV and SFV), it was predicted that it would not affect CHIKV replication - the fact that replication was inhibited suggests that the terminal loop of SL194 represents a CHIKV-specific signal motif. The single A_{210} bulge in SL194 and the A_{186} bulge in SL165 cannot be synonymously mutated. Each of these unpaired adenosines is positioned at the centre of a His codon (CAU), shares the same local steric context (G:C pair above, A:U pair below) and both exhibit very high reactivity to NMIA during

SHAPE mapping, indicating that they are consistently unpaired. Further study could determine the importance of the unpaired bulges, for example by targeting with locking nucleic acids (LNAs). A current study by O. Prosser (University of Leeds) has demonstrated the efficacy of an LNA targeting SL165 in reducing CHIKV replication. The LNA is designed to disrupt base-pairing of the stem but also blocks access to the A_{186} bulge. Alternatively, non-synonymous mutations could be designed, as non-synonymous mutation may exhibit no phenotypic consequence for CHIKV replication, as was the case for *SL102mut-Loop* and *SL246mut-Loop*.

While the sequence or size of the terminal loop of SL165 did not significantly affect genome replication directly, virus replication as a whole was inhibited by increasing the size of the unpaired terminal-loop by four nucleotides (Figs 4.4-6). The primary nucleotide sequence and structure of the nsP1 CSE, has previously been shown to be highly conserved between divergent alphaviruses (13). Reverse genetic evidence presented here suggests that the size of the terminal loop, rather than its primary sequence, may contribute to functionality of this RNA replication element during the viral lifecycle. Interestingly, although extensive non-synonymous mutations in single-stranded regions of SL102 and SL246 had no significant impact on CHIKV genome or virus replication, these regions were not subject to genetic drift during CHIKV passaging. Likewise, SL165 did not accumulate mutations in the terminal loop of *SL165mut* over 10 passages. Therefore, the single-stranded regions appear to be conserved during experimental passage, which parallels the conservation of this region observed during epidemics (395). Preference for the wild-type sequence in single-stranded regions may be a result of codon usage bias or selection for thermodynamic stability of the structure, rather than restoring functionality of the sequence as a *cis*-acting element.

4.3.8 Conclusion

Enhancement of CHIKV replication by the 5′ region of the genome is structure-dependent, as well as dependent on the terminal loop sequence of SL194. Passaging of stem-loop mutant viruses selected revertant mutations, which restored wild-type RNA sequence and, in one case, a compensatory revertant, which restored base-pairing of the stem - but not primary sequence. Further study will be necessary to determine potential interactions between stem-loops in the structured region (Chapter 5) and between stem-loops and *trans*-acting host or viral factors.

Chapter 5. Dynamic interactions between structures within the 5′ region of the CHIKV genome

5.1 Introduction

In previous chapters, the impact of disrupting individual stem-loops structures within the 5′ 300 nt of the CHIKV genome was investigated (Chapters 3 & 4). Stem-loop 47 (SL47) within the 5′ UTR of the CHIKV genome enhanced viral genome replication in human and *Ae. albopictus* derived cells, while stem-loops within the adjacent nsP1-encoding region (SL85, SL102, SL165, SL194 and SL246) enhanced replication in a host-specific manner (Figs 3.5 & 3.6). In addition to individual functionality, RNA stem-loop replication elements may act synergistically. Synergy can result from direct RNA:RNA interactions, for example between stem-loops F and G in the poliovirus genome. In addition to their individual functions, these two elements are required in close proximity to bind the p36 factor, which does not bind to either stem-loop alone (426). Alternatively, synergy can result from simultaneous protein:RNA interactions. For example, the *cis*-acting constitutive and RNA transport elements, CTE and RTE, in the HIV-1 genome synergistically increase expression of otherwise poorly expressed genes via increased binding affinity of the Nuclear RNA Export Factor 1 (NXF1) protein in the presence of both structures (427). Besides additive or synergistic interactions, RNA structures may exhibit functional redundancy, whereby the disruption of one structure is tolerated in the presence of an intact proximal structure. Pairwise disruption of RNA elements has been shown to elicit a different impact on virus replication compared to individual disruption in VEEV and DENV in a host-specific manner (376). Individual removal of stem-loops SL-I and SL-II in the 3′ UTR of DENV was tolerated in human cell lines, but removal of both structures significantly reduced genome replication (393).

Despite the fact that SL85, SL102, SL165 and SL194 were dispensable to CHIKV replication in mosquito cell lines, these stem-loops were conserved during single-host passaging in C6/36 cells. Similarly, SL246 was conserved during passaging in Huh7 cells. Genetic drift was not detected in any of the stem-loops in the 5′ 300 nt during single-host passaging, suggesting a selective advantage. Both functional redundancy and synergistic interactions may be revealed by pairwise or cumulative disruption of RNA structures.

RNA secondary structure dictates the tertiary structure of an RNA molecule. Regions of complex secondary structure can direct long-range RNA:RNA interactions (428), act as a recognition motif for *trans*-activating factors (429) and determine global folding of RNA (430). Several types of RNA tertiary structure have been described, the simplest being the hairpin-type pseudoknot (295). A hairpin-type pseudoknot is formed when the single-stranded terminal loop of a hairpin or stem-loop base-pairs with a complementary single-stranded region elsewhere in the RNA (Fig 5.1). Tertiary RNA structures, including pseudoknots and transient kissing interactions, are essential during translation and genome replication of several viruses. For example, HCV employs a 5′ internal ribosome entry site (IRES) composed of a four-way junction necessary to direct cap-independent translation (431). Meanwhile, SL2 in the 3′ UTR of HCV and a cruciform structure within the coding sequence of non-structural protein 5B (NS5B) form a kissing-loop interaction essential for genome replication (341).

Tertiary structures are highly stable compared to their secondary structure counterparts (432). However, the stability of a tertiary structure may be altered by protein binding, providing an elegant mechanism of differential gene expression (433). Dynamic shifts in tertiary RNA structure have been demonstrated to facilitate the switch from translation to replication of viral RNA in picornaviruses at the appropriate stage of the viral lifecycle (434). Similar phenomena have been observed in coronaviruses, for example, thermodynamic studies and *in silico* simulations have demonstrated the formation of two mutually exclusive RNA structures in the 3′ UTR of mouse hepatitis virus (MHV) (435). A conformational switch between the double-hairpin structure and a marginally stable *cis*-acting pseudoknot occurs only when the hairpins cannot form, initiating a switch from production of the negative-sense replication intermediate to positive-sense sub-genomic RNA. Differential formation of RNA structure occurs in different local conditions. For example, the tRNA-like structure (TLS) of turnip yellow mosaic virus (TYMV) is aminoacylated upon entry of the viral genome into cells, promoting translation via eukaryotic translation elongation factor 1 α (eEF1A) binding in conjunction with an upstream pseudoknot (436,437). As infection progresses, pseudoknot-dependent binding of the viral RdRp to the TLS initiates negative-strand synthesis (438). Increasing levels of RdRp outcompete eEF1A for binding of the TLS, signalling progression of the viral lifecycle, and genomic RNA is sequestered into spherules, further reducing the local concentration of eEF1A (439).

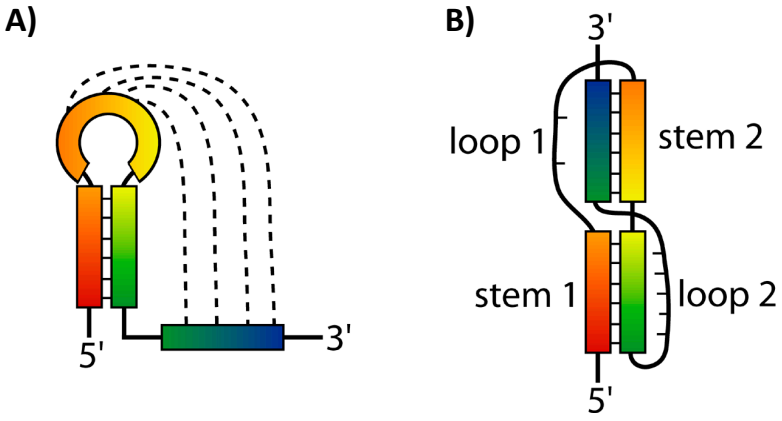


Figure 5.1: The structure of a hairpin-type pseudoknot (adapted from (440)**). (A)** The terminal loop of a hairpin base-pairs (dotted line) with a proximal single-stranded region, resulting in **(B)** a H-type pseudoknot with base-pairing within two stems separated by two intervening loops. Colour gradient (red to blue) represents the 5′ to 3′ polarity of the RNA molecule.

As previously noted, while reverse genetic studies support the predicted structure of SL85 and the fact that it functions in a structure-dependent mechanism (Figs 4.2 & 4.3), *in vitro* SHAPE reactivity data suggested that this region is also able to undergo alternative interactions (Fig 3.2). These observations may be explained by the dynamic formation of a pseudoknot competing with SL85. Pseudoknots and similar RNA tertiary motifs have been demonstrated to direct various forms of alternative translation including frameshifting (441), readthrough (442) and ribosomal shunting (443). The presence of RNA secondary or tertiary structures around an AUG start codon has been demonstrated to impact the rate of initiation from alternative (AUG and non-AUG) start sites, particularly by ribosomal shunting across the base of secondary structures (411). For example, plant pararetroviruses including cauliflower mosaic virus (CaMV) exhibit ribosomal shunting across an extended hairpin, resulting in translation from an alternative start codon which is essential for infectivity (444). The region of secondary structure containing SL85 and SL102 lies between two in-frame AUG start codons, both of which have been demonstrated to function in human cells (unpublished data, discussed in 5.2.4). Thus, we hypothesis that formation of dynamic higher order RNA-RNA interactions in this region could influence initiation of translation from alternative start codons, or represent a mechanism by which CHIKV switches from ORF-1 translation to genome replication.

The aim of the work described in this chapter was to identify potential tertiary interactions of the 5′ region of the CHIKV genome, including the possibility of stem-loop:stem-loop interactions and alternative pseudoknot interactions involving SL85.

5.2 Results

5.2.1 Stem-loops in the 5′ region are essential for CHIKV genome replication in mammalian and mosquito hosts

In order to investigate the potential for direct or indirect synergistic interactions between stem-loops in the 5′ region of the CHIKV genome, each stem-loop was disrupted cumulatively to produce a fully destabilised 5′ mutant, termed *CombmutA* (Fig 5.2). Infectious virus could not be recovered following electroporation of Huh7 and C6/36 cells with *in vitro* transcribed *CombmutA* CHIKV_IC RNA (Fig 5.3A). Similarly, infectious virus could not be recovered when this RNA was transfected by liposome delivery into Huh7 or C6/36 cells, demonstrating that the abolition of viral replication is independent of the RNA transfection method (Fig 5.3B).

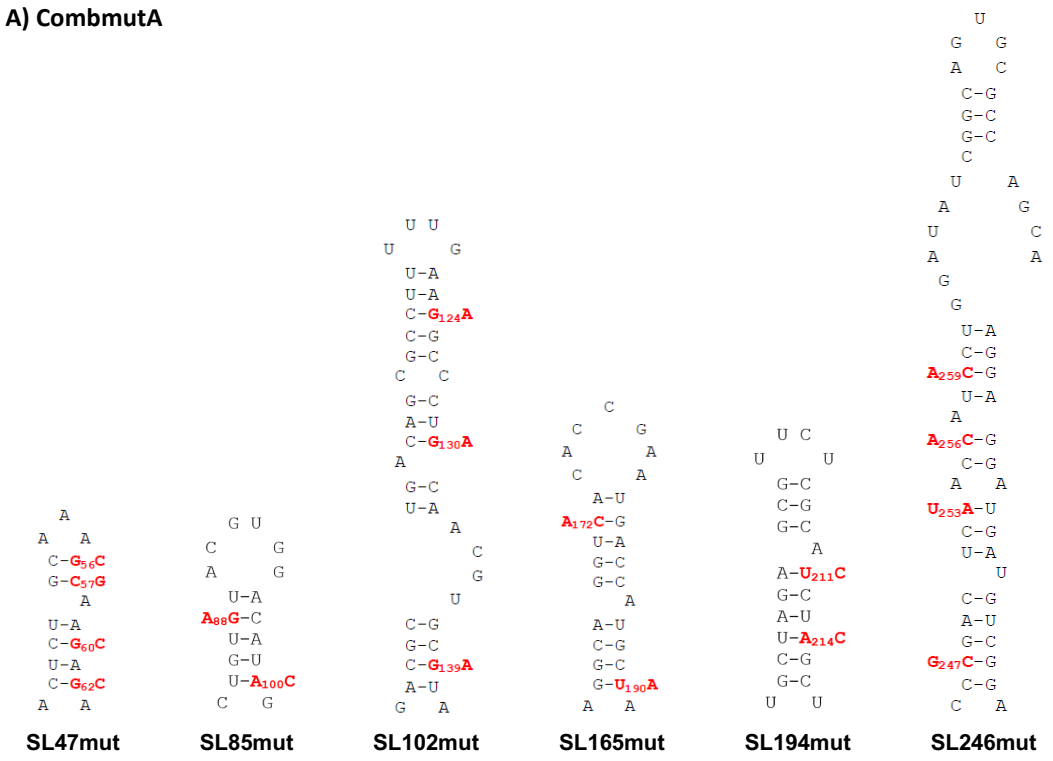


Figure 5.2: Combined mutant A. Each stem-loop mutant (labelled in bold) was incorporated into CHIKV_IC in combination to produce the *CombmutA*. Synonymous mutations (red) were designed to destabilise base-pairing of the heteroduplex stem.

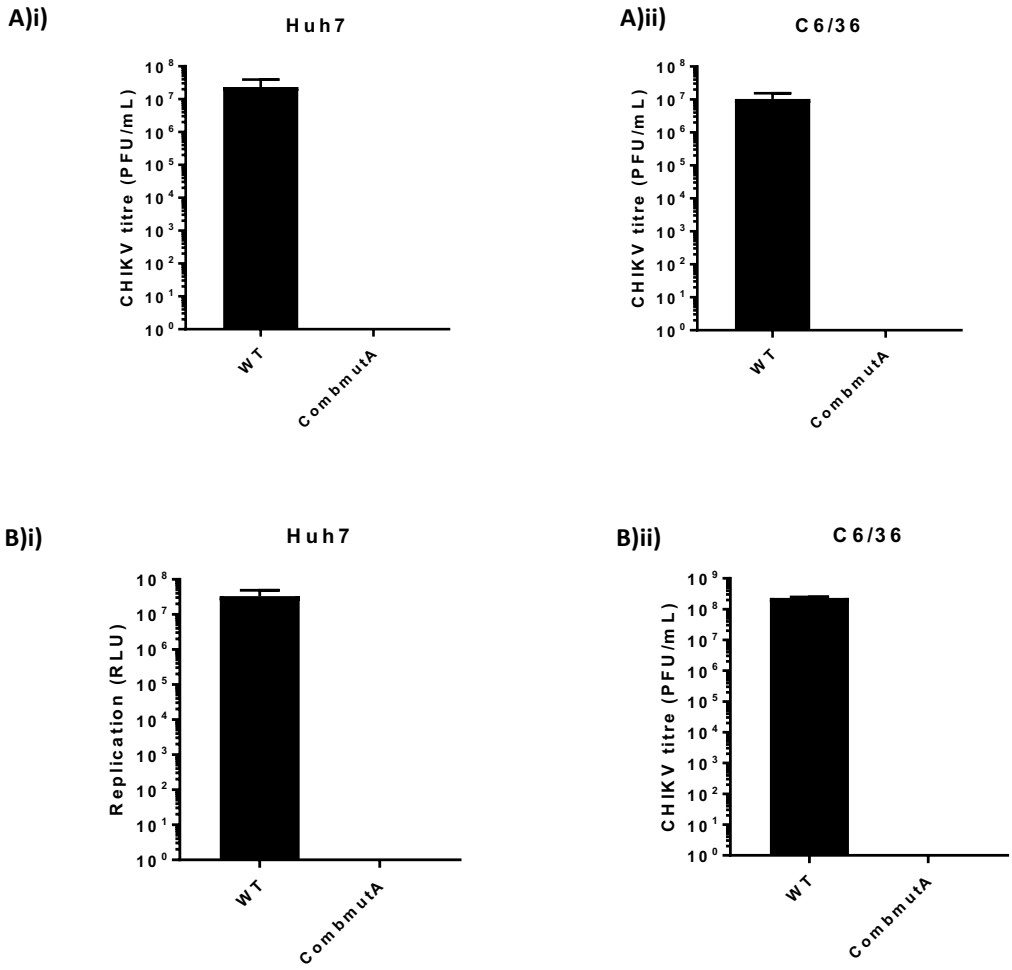


Figure 5.3: Combined disruption of SL47-SL246 abolished CHIKV replication. CHIKV titres following **(A)** electroporation and **(B)** transfection of **(i)** Huh7 and **(ii)** C6/36 cells with WT and *CombmutA* CHIKV_IC *in vitro* transcribed RNA over 24 hours (n=3). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

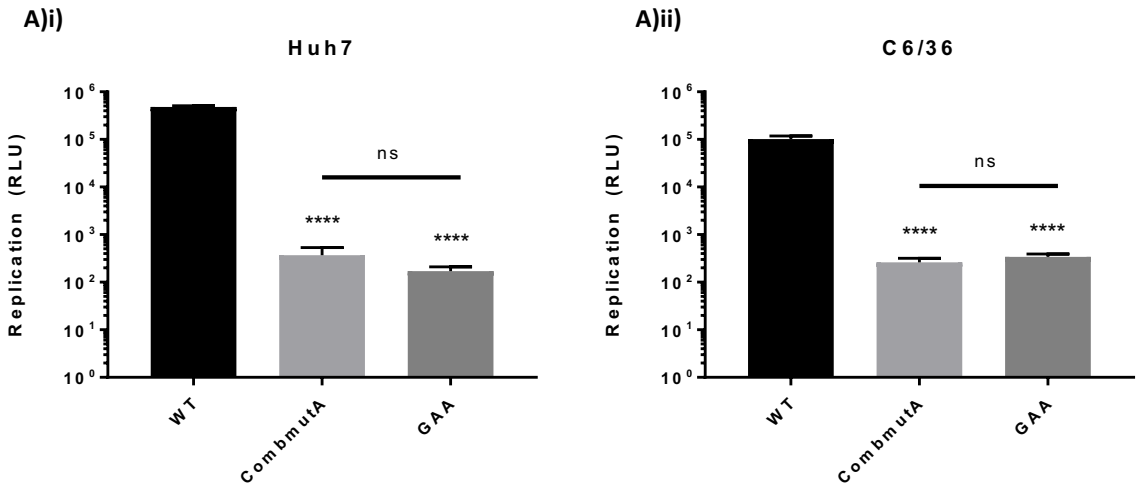


Figure 5.4: Combined disruption of SL47-SL246 abolished viral genome replication. (A) Replication phenotype of sub-genomic replicons (n=3) in **(i)** Huh7 and **(ii)** C6/36 cells over 24 hours post-transfection. CHIKV_Rep replicon with wild-type RNA structure (black) compared to *CombmutA* CHIKV_Rep replicon (light grey), alongside WT CHIKV_Rep(GDD>GAA) (dark grey) as a non-replicating control. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.0001$ (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

5.2.2 Combinations of stem-loops are essential for CHIKV replication in mammalian and mosquito hosts

Combined destabilisation of all stem-loops within the 5′ 300 nt of the CHIKV genome prevented virus and genome replication, demonstrating that base-pairing in this region was essential for CHIKV replication in both mammalian and mosquito hosts. In contrast, complete abolition of CHIKV replication was not seen following disruption of any one single stem-loop in either host (Figs 3.5 & 3.6). These data suggest that a combinatorial effect had occurred for *CombmutA* via direct or indirect interactions between stem-loops. Alternatively, the stem-loops may have distinct functions and an additive effect may have occurred.

In order to explore the interactions between stem-loops and determine the minimum set of stem-loops which prevented replication in each host, two further mutants were designed (Fig 5.5). Each mutant contained only those mutations which had been shown to impact CHIKV replication in either mammalian-, *CombmutB,* or mosquito-, *CombmutC,* derived cells. Infectious virus could not be recovered following electroporation of BHK-21 cells with *in vitro* transcribed *CombmutB* CHIKV_IC RNA (Fig 5.6A). In order to formally demonstrate that *CombmutB* was incapable of virus replication due to a lack of genome replication, as *CombmutA* was, the *CombmutB* mutations were cloned into the sub-genomic replicon CHIKV_Rep and transfected into Huh7 and C6/36 cells. Genome replication for *CombmutB* was indistinguishable from the non-replicating GAA control, indicating that *CombmutB* was non-functional at the level of genome replication, as had been demonstrated for *CombmutA* (Fig 5.6B, $p>0.05$). The inability of *CombmutB* to replicate in C6/36 cells was of particular interest, since individually disrupting stem-loops SL85, SL102, SL165 or SL194 had previously been demonstrated to have no impact on CHIKV replication in mosquito cells (Figs 3.5 & 3.6). Thus, it was predicted that *CombmutB* would be comparable in its replication to *SL47mut,* since an intact SL246 was present. Unexpectedly, *CombmutB* was completely incapable of producing infectious virus in C6/36 cells. Mutation of each of the stem-loops SL85, SL102, SL165 and SL194 only contributed to inhibition of CHIKV replication in mosquito cells in the combination with other mutated structures, indicating functional redundancy between two or more of the stem-loops.

In contrast to *CombmutB*, infectious virus was recovered following electroporation of BHK-21 cells with *in vitro* transcribed *CombmutC* CHIKV_IC RNA, although virus replication was significantly reduced compared to WT ($p \leq 0.05$) (Fig 5.6A). The ability of *CombmutC* to replicate in Huh7 and C6/36 cells was investigated in one-step growth assays (Fig 5.6C). Infections were carried out alongside the mutants *SL47mut* and *SL246mut,* (i.e. the two stem-loops destabilised in the combined mutant *CombmutC)*. Both *SL47mut* and *SL246mut* recapitulated the phenotypes previously observed in both Huh7 and C6/36 cells (Figs 3.5 & 3.6). CHIKV replication was significantly reduced in the combined mutant *CombmutC* compared to WT in Huh7 and C6/36 cells ($p \leq 0.001$). Interestingly, the reduction in CHIKV replication for *CombmutC* exceeded the expected additive phenotype of *SL47mut* and *SL246mut* in both cell lines, suggesting synergistic interaction between the stem-loops. In particular, since *SL246mut* was previously demonstrated not to impact CHIKV replication in mammalian cells, *SL47mut* should be directly comparable to *CombmutC* (Fig 5.6B)i)). This is not the case – CHIKV replication was significantly reduced in *CombmutC* compared to *SL47mut* in Huh7 cells ($p \leq 0.0001$). These data suggested an interaction between SL47 and SL246 in mammalian cells. In mosquito cells, the synergistic effect was much more modest, exceeding the expected additive phenotype by less than 0.5 logs.

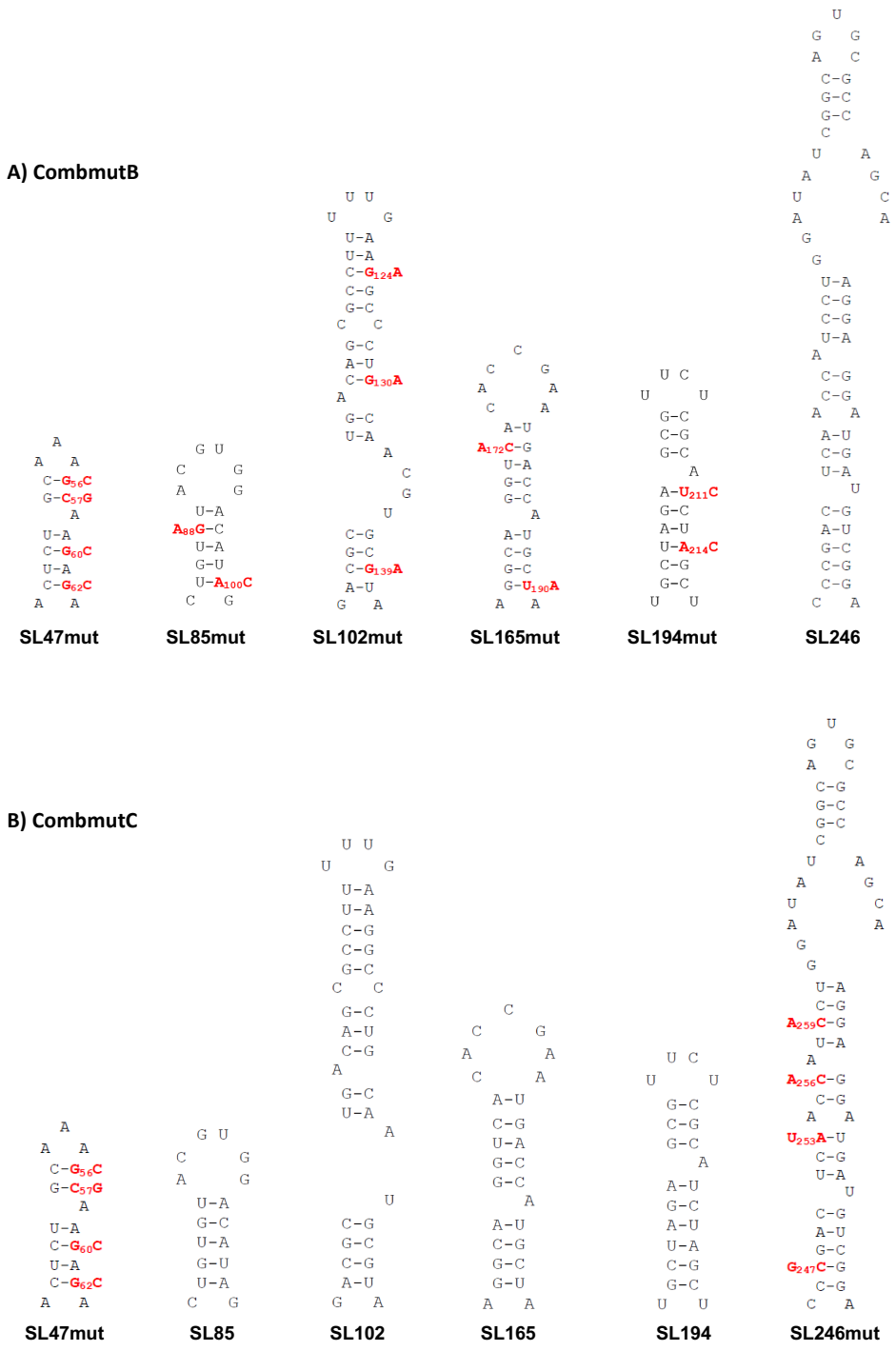


Figure 5.5: Combined mutants B and C. (A) *CombmutB* contains synonymous substitutions in all stem-loops shown to affect CHIKV replication in mammalian cell lines. **(B)** *CombmutC* contains synonymous substitutions in both stem-loops shown to affect CHIKV replication in mosquito cell lines. Mutated stem-loops labelled in red. Each stem-loop mutant (labelled in bold) was incorporated into CHIKV_IC in combination.

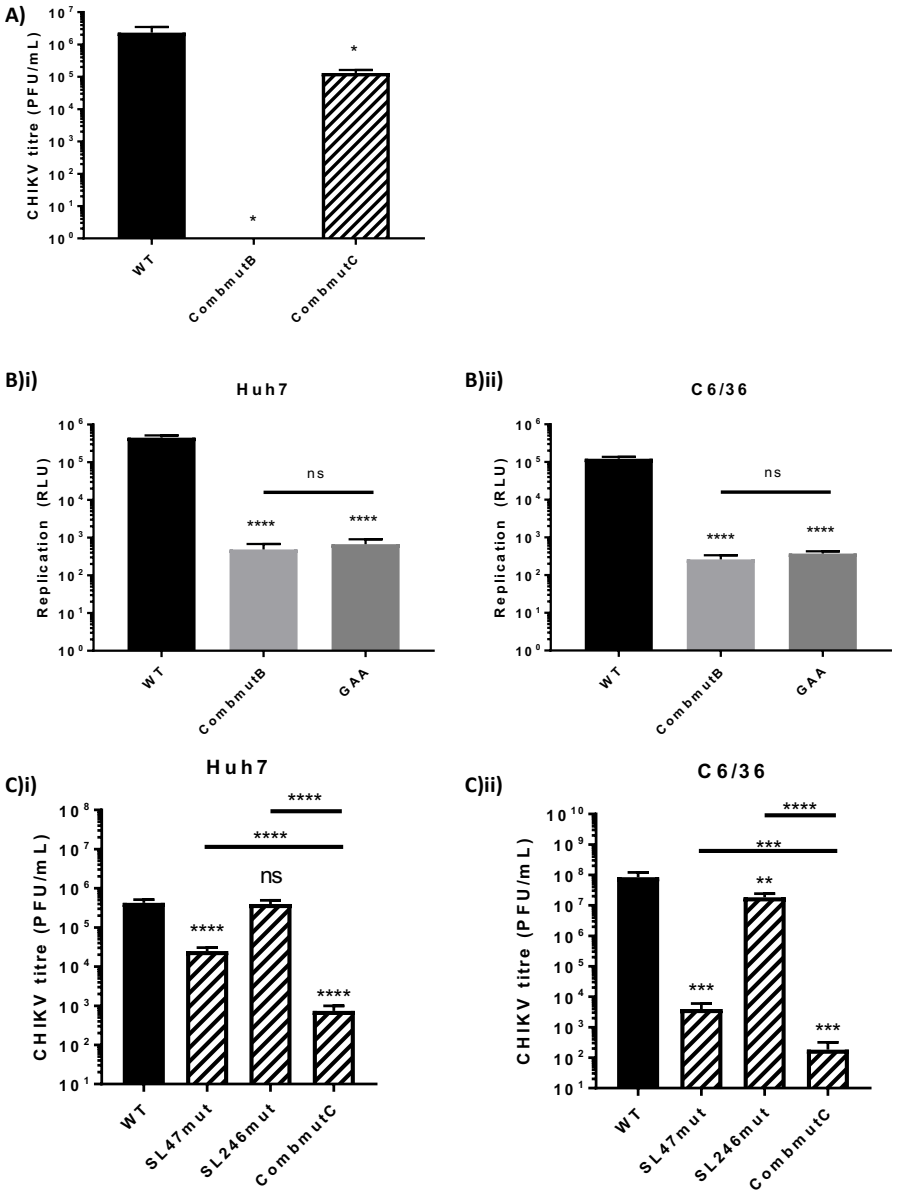


Figure 5.6: Combined mutations abolished or severely reduced CHIKV replication similarly in both host species. (A) CHIKV titres following electroporation of BHK-21 cells with WT, *CombmutB* and *CombmutC in vitro* transcribed CHIKV_IC RNA over 24 hours. **(B)** Replication phenotype of sub-genomic replicons in **(i)** Huh7 and **(ii)** C6/36 cells over 24 hours post-transfection. CHIKV_Rep replicon with wild-type RNA structure (black) compared to *CombmutB* (light grey), alongside WT CHIKV_Rep(GDD>GAA) (dark grey) as a non-replicating control. **(C)** CHIKV titres following infection of **(i)** Huh7 cells with WT, *SL47mut* and *CombmutC* virus and **(ii)** C6/36 cells with WT, *SL47mut, SL246mut* and *CombmutC* virus over 24 hours. n=3 for all experiments. . * represents statistical significance for each mutant compared to wild-type under a

two-tailed Student's *t*-test: $p \leq 0.05$ (*), ≤ 0.01 (**), ≤ 0.001 (***), ≤ 0.0001 (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

5.2.3 Potential pseudoknot formation

In Chapter 3, SHAPE-constrained thermodynamic analysis demonstrated that the RNA structure of the CHIKV genome immediately downstream of the AUG start codon contained two distinct RNA replication elements (SL85 and SL102 (Fig 3.2A)), both of which are involved in efficient genome replication in human-derived cells but not those from *Ae. albopictus* (Fig 3.4). However, SHAPE reactivity data for SL85 was not consistent with the structure predicted by SHAPE-constrained thermodynamic modeling, suggesting that this region of the genome may be structurally dynamic and able to form alternative interactions (Figs 5.7B & 5.8A). Structure-led reverse genetic analysis, de-stabilising SL85 and then restoring base-pairing with compensatory substitutions, demonstrated that SL85 is essential for efficient CHIKV genome replication in human-derived cells and functions through a structure-dependent mechanism (Figs 4.2A & 4.3A) (Fig 5.7A). The ability of *SL85mut-Comp* to rescue wild-type CHIKV replication demonstrates that the structure forms during the CHIKV lifecycle *in vivo*. Given the lack of agreement between the combined structural model/reverse genetic data and SHAPE reactivity data for SL85, the potential for alternative structure formation was explored. It was predicted that a pseudoknot may form in competition with the SL85 stem-loop and this was investigated by structural prediction and reverse genetic analysis.

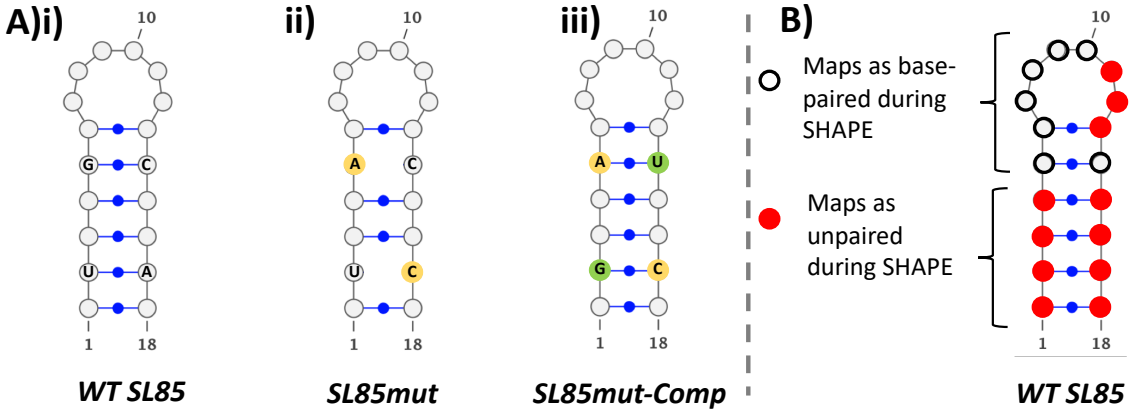


Figure 5.7: Disagreement between phenotypic studies and SHAPE reactivity for SL85. **(A)** Combined model of SL85 secondary structure, with base-pairing labelled in blue, including **(i)** the wild-type structure, **(ii)** *SL85mut* with disruption of two base-pairs in the stem (yellow) and **(iii)** *SL85mut-Comp* which restores of those base-pairs by compensatory mutations on the opposite side of the stem (green). The substitution of the wild-type G:C pair with a compensatory A:U pair distinguishes between structure- and sequence-based effects. WT and *SL85mut-Comp* have comparable phenotypes. Therefore, phenotypic data suggests that the stem-loop shown in the figure is likely to form. **(B)** Binary representation of SHAPE mapping data from A. Tuplin (University of Leeds) with unpaired nucleotides predicted in the stem (red circles) and paired nucleotides (circles) in the terminal loop. SHAPE mapping suggests that the stem-loop shown in the figure is highly dynamic or does not form.

Regions of the terminal loop predicted by thermodynamic modelling and reverse genetic analysis to be single-stranded but suggested by SHAPE mapping to be base-paired were analysed for sequence complementarity with proximal single-stranded regions. Two potential alternative RNA/RNA interactions were predicted, termed Pseudoknot-1 (Pk1) and Pseudoknot-2 (Pk2). Base-pairing of the terminal loop sequence of SL85 was hypothesised to involve upstream (Pk1) or downstream (Pk2) regions (Fig 5.8B). The potential pseudoknots Pk1 and Pk2 were investigated by reverse genetic analysis in a similar manner to that utilised during investigation of individual stem-loops. Unfortunately, the upstream single-stranded region, predicted to base-pair in Pk1, contained the AUG start codon of ORF-1, complicating reverse genetic analysis due to the presence of the start codon and surrounding Kozak context.

Synonymous mutations were designed to disrupt Pk1 from the upstream sequence (*Pk1$_S$mut*) and terminal loop (*Pk1$_L$mut*) regions and incorporated into CHIKV_IC (Fig 5.9A). A compensatory set of mutations was also designed to restore base-pairing within Pk1 (*Pk1-Comp*). One-step growth assays were carried out for WT CHIKV and Pk1 mutant virus in Huh7 and C6/36 cells over 24 hours. In order to assess the potential formation of Pk1 during CHIKV replication, the number of infectious virions released during infection was determined by plaque assay (Fig 5.9B-C). CHIKV replication was significantly reduced for *Pk1$_S$mut* compared to WT ($p \leq 0.01$) in Huh7 cells, whereas no significant difference in viral titre was observed in C6/36 cells (Fig 5.9B)i) & C)i)). *Pk1$_L$mut* virus replicated to wild-type titres in both Huh7 and C6/36 cells (Fig 5.9B)ii) & C)ii)). Subsequently, Huh7 cells were infected with *Pk1-Comp* virus to determine whether the replication phenotype of *Pk1$_S$mut* in these cells could be rescued by restoring base-pairing across the predicted pseudoknot. Replication of *Pk1-Comp* was significantly reduced relative to WT ($p \leq 0.0001$), representing an even greater replication phenotype than *Pk1$_S$mut* alone. Taken together, these data suggested that Pk1 did not form in the CHIKV genome and that phenotypes observed for the upstream region mutants may have been due to disturbance of the Kozak sequence.

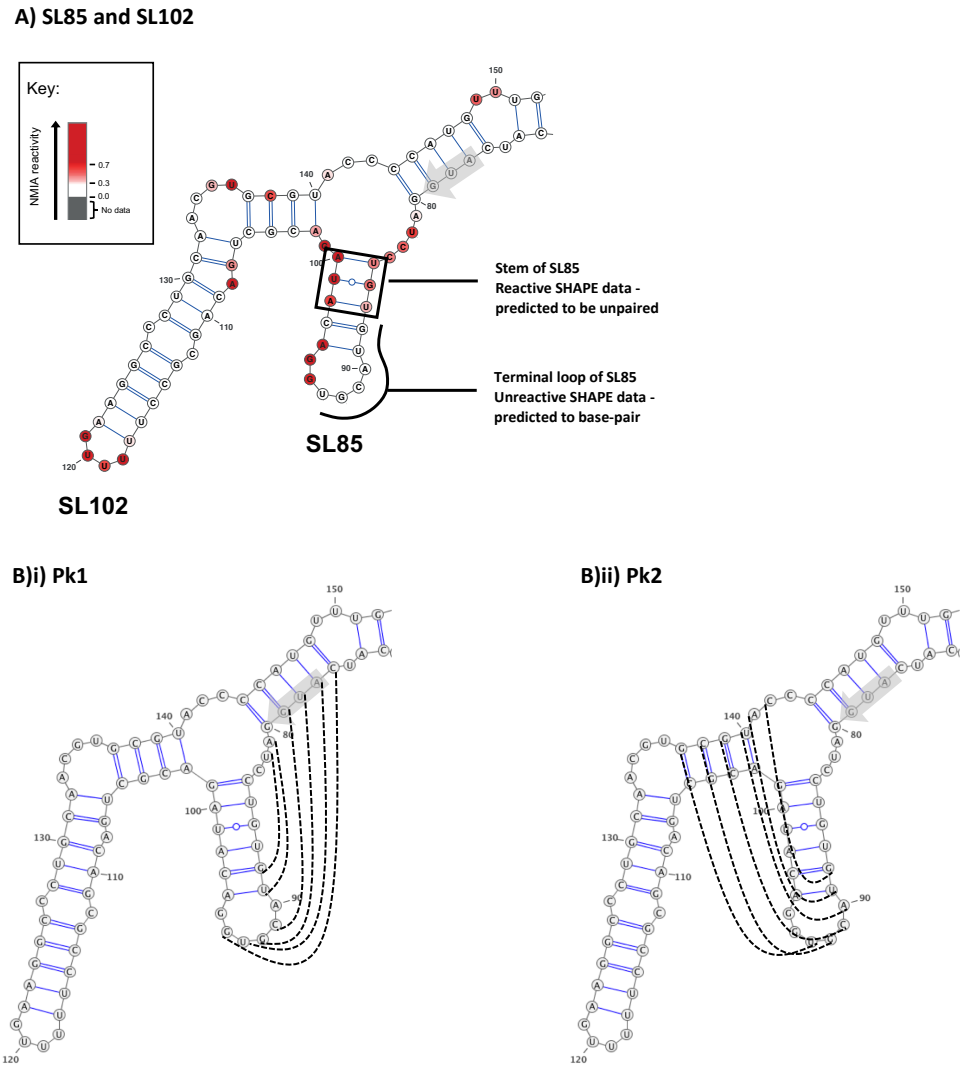


Figure 5.8: Potential pseudoknot formation. (A) SHAPE-constrained UNAfold-predicted structure of SL85 and SL102. 37°C SHAPE reactivities for individual nucleotides are overlaid onto a 37°C thermodynamically derived model of RNA folding, generated using SHAPE-directed constraints. The AUG start codon of nsP1 is denoted by a grey arrow. SHAPE reactivities are shown as a heat map: white SHAPE reactivities between 0–0.3 and increasing intensities from light pink to dark red indicate increasing SHAPE reactivities, as denoted by the key. High reactivity (red) denotes unpaired nucleotides whereas low reactivity (white) denotes base-paired nucleotides. **(B)** Schematic representation of two potential pseudoknots involving SL85: **(i)** pseudoknot-1 (Pk1) forming over the downstream AUG start codon of nsP1 (grey arrow) and **(ii)** pseudoknot-2 (Pk2) forming with an upstream sequence within SL102. Dotted lines represent potential base-pairing in each putative pseudoknot.

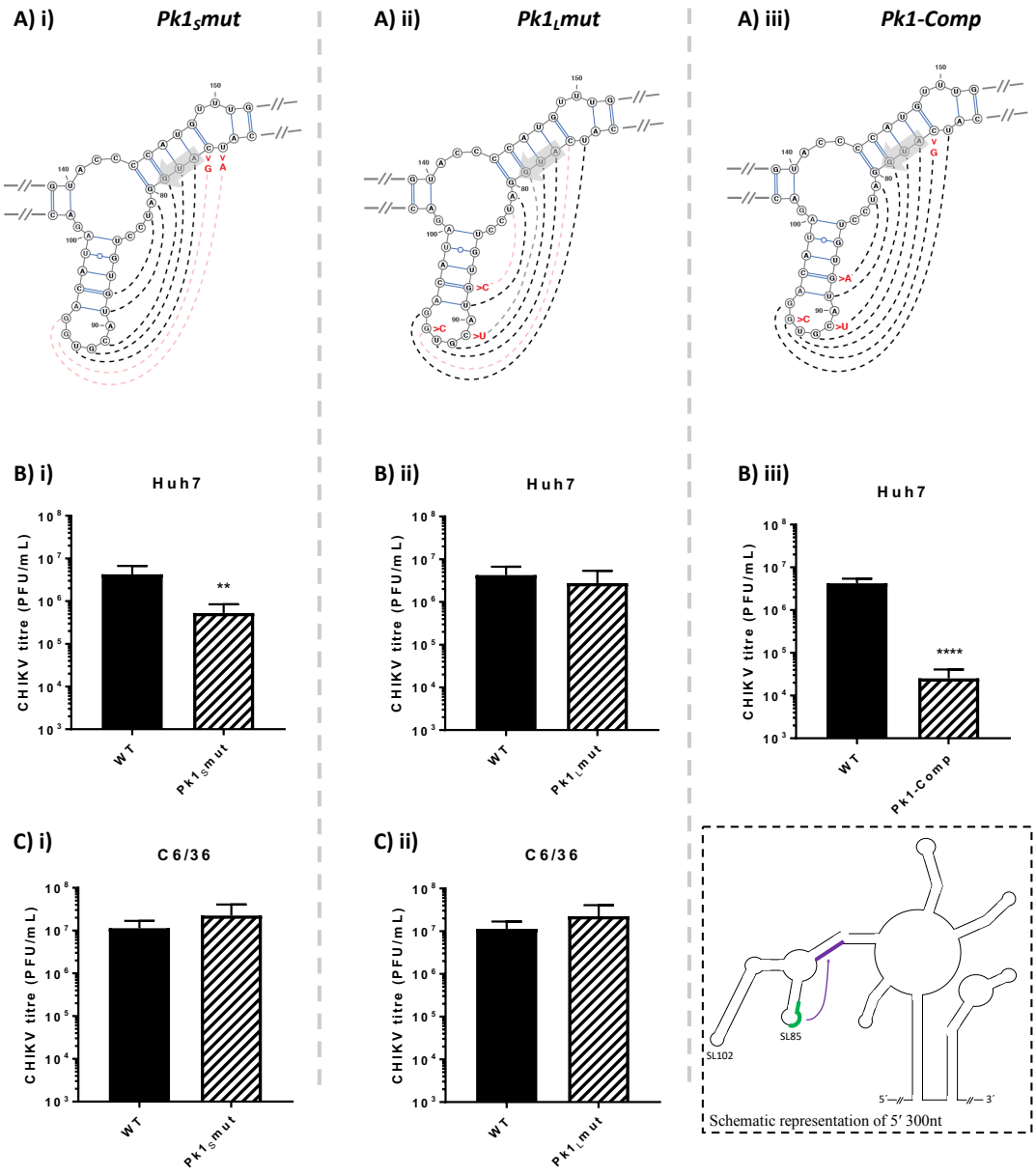


Figure 5.9: Phenotypic analysis of the putative pseudoknot, Pk1. (A) Mutations (red) predicted to destabilise the putative pseudoknot Pk1 from the **i)** upstream (*Pk1$_S$mut*) or **ii)** loop (*Pk1$_L$mut*) regions. **iii)** A compensatory set of mutations predicted to restore base-pairing of the pseudoknot: *Pk1-Comp*. The AUG start codon of nsP1 is denoted by a grey arrow; all mutations are synonymous. Dotted lines represent potential base-pairing in Pk1; black lines represent base-pairing, red lines represent disrupted pairing, grey lines represent a non-Watson-Crick base-pair produced by mutation. **(B-C)** CHIKV titres following MOI=1 infection with virus of wild-type RNA structure (black bar) and mutants (hatched bars) over 24 hours (n=3). **(B)** Infections were

carried out in Huh7 cells with **i)** *Pk1$_S$mut,* **ii)** *Pk1$_L$mut* and **iii)** *Pk1-Comp* virus and **(C)** C6/36 cells with **i)** *Pk1$_S$mut* and **ii)** *Pk1$_L$mut* virus. A schematic of the structured region (dashed box) is displayed for reference, with the loop region (green) and upstream sequence (purple) denoted. * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*), ≤ 0.01 (**), ≤ 0.001 (***), ≤ 0.0001 (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

Synonymous mutations were then designed to disrupt Pk2 from the downstream sequence (*Pk2$_s$mut*) and terminal loop regions. Unfortunately, the range of synonymous mutations which could be incorporated into the terminal loop of SL85 were limited. Only two nucleotides predicted to base-pair in Pk2 could be synonymously mutated. *Pk1$_L$mut,* although originally designed to disrupt Pk1, was predicted to disrupt base-pairing in Pk2 (Fig 5.10A)ii)) and had displayed no replication phenotype (Fig 5.9B)i) & C)i)). Huh7 and C6/36 cells were infected with either WTCHIKV or *Pk2$_s$mut* virus at MOI=1 for 24 hours. CHIKV replication was significantly reduced for *Pk2$_s$mut* compared to WT ($p \leq 0.01$) in Huh7 cells, whereas no significant difference in viral titre was observed in C6/36 cells (Fig 5.10B & C). These data suggest a function for the upstream single-stranded region during the viral lifecycle but do not demonstrate the formation of Pk2.

5.2.4 Alternative AUG start codon

Several studies undertaken by members of the group (O. Antoniak, E. Levitt and T. Nguyen) have demonstrated that the Met^{24} AUG codon in nsP1 functions as an alternative start site, producing a smaller, N-terminally truncated form of nsP1 (Fig 5.11). Western blot analysis and mutagenesis studies have determined the presence and importance of this truncated protein product respectively. Interestingly, Met^{1}>Ala mutation of the authentic 5′ AUG start codon abolishes virus replication completely, indicating that the truncated form of nsP1 is not sufficient for CHIKV replication. However, Met^{24}>Ala mutation significantly reduces CHIKV replication suggesting that the truncated nsP1 plays an important but non-essential role in the viral lifecycle.

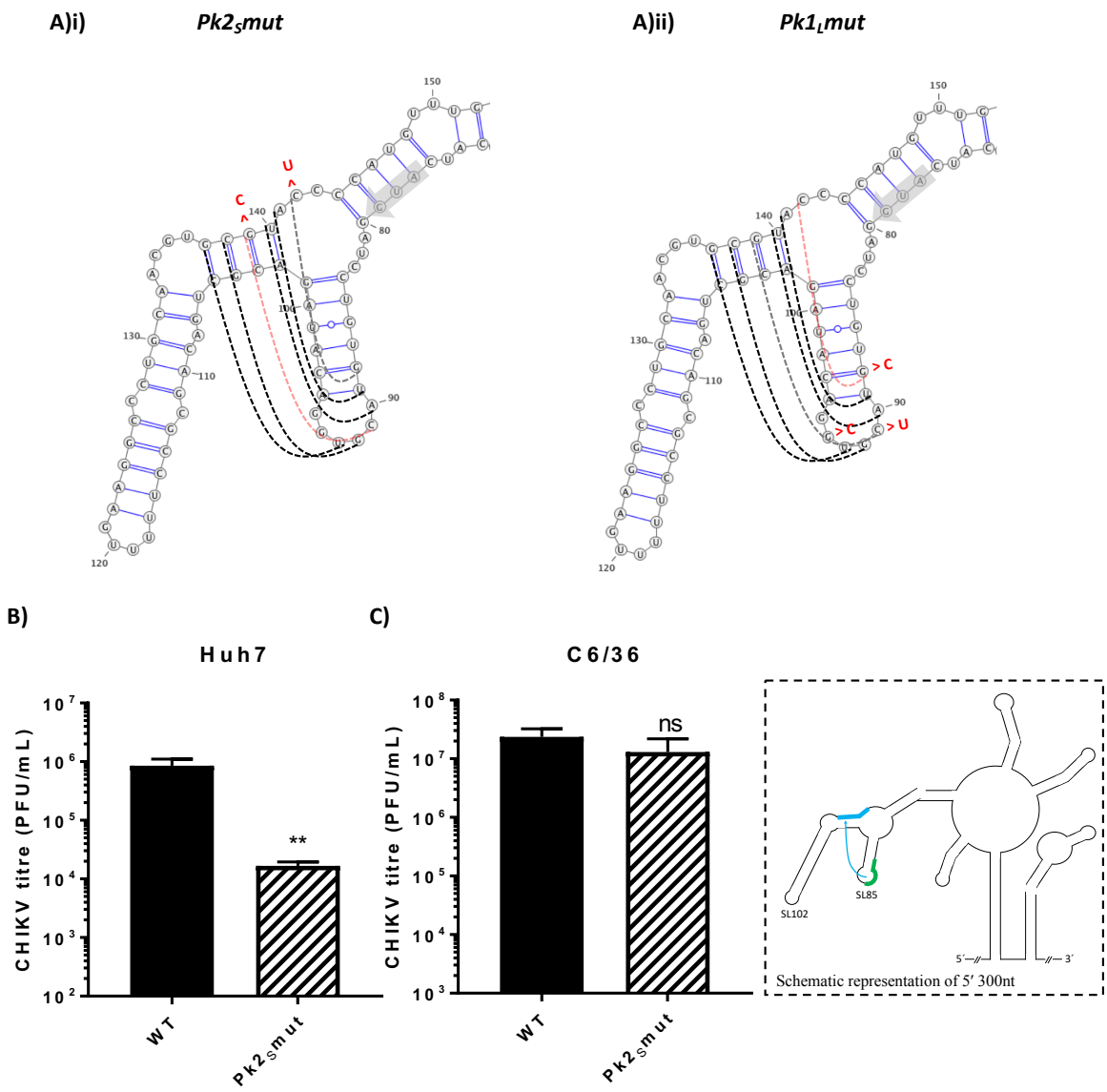


Figure 5.10: Phenotypic analysis of the putative pseudoknot, Pk2

(A) Mutations (red) predicted to destabilise the putative pseudoknot Pk2 from **(i)** the downstream (*Pk2Smut*) region and **(ii)** the terminal loop region (*Pk1Lmut)*. The AUG start codon of nsP1 is denoted by a grey arrow; all mutations are synonymous. Dotted lines represent potential base-pairing in Pk2; black lines represent base-pairing, red lines represent disrupted pairing, grey lines represent a non-Watson-Crick base-pair produced by mutation. **(B-C)** CHIKV titres following MOI=1 infection with virus of wild-type RNA structure (black bar) and *Pk2Smut* (hatched bars) over 24 hours in **(B)** Huh7 and **(C)** C6/36 cells (n=3). A schematic of the structured region (dashed box) is displayed for reference, with the loop region (green) and downstream sequence (blue) denoted. . * represents statistical significance for each mutant compared to wild-type under a two-

tailed Student's *t*-test: ns (non-significant), $p \leq 0.01$ (**). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

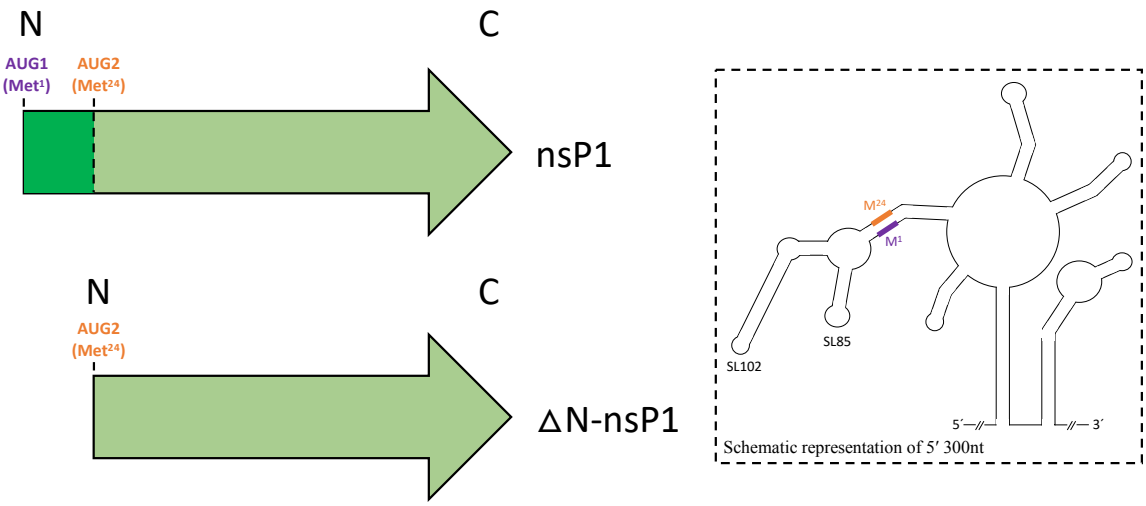


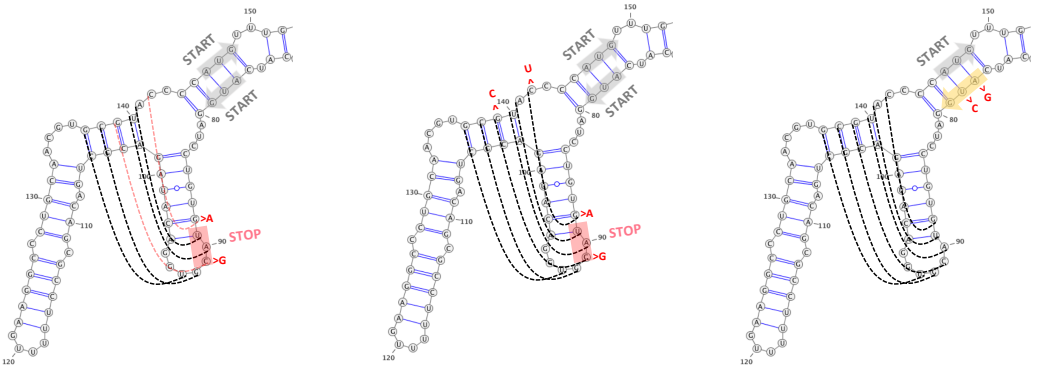
Figure 5.11: Alternative initiation of nsP1 translation

Schematic representation of CHIKV non-structural protein 1, showing wild-type nsP1 relative to the N-terminal truncated mutant (ΔN-nsP1). The N-terminus (N) and C-terminus (C) are labelled, along with each start codon AUG1 (Methionine^{1}) and AUG2 (Methionine^{24}). The N-terminal sequence is denoted in darker green in wild-type nsP1. A schematic of the structured region (dashed box) is displayed for reference, with AUG1 (M^{1}) codon (purple) and downstream AUG2 (M^{24}) codon (orange) denoted.

SL85 and SL102 lie between the two in-frame AUG codons, looping out to allow base-pairing between the start codons (Fig 5.11). In order to analyse the mechanism by which AUG2 functions as a start codon, a stop codon (UAG) was introduced into the terminal loop of SL85 (Fig 5.12A). The resulting mutant, *AUG1STOP,* retains the AUG1 sequence but presumably cannot express full length nsP1 due to termination after 4 aa. This mutant can be compared to *AUG1mut*, which lacks the AUG1 sequence altogether. Comparison of these mutants allows the mechanism of AUG2 initiation to be determined as AUG1-dependent or independent, Secondarily, the stop codon disrupts base-pairing of the predicted pseudoknot Pk2 in a manner that allows for restoration of base-pairing by compensatory mutation. Thus, the *AUG1STOP* mutant can be compared to the *AUG1STOP-Pk2-Comp* mutant to assess the role of the putative pseudoknot Pk2 during viral replication. While the introduction of a stop codon limits the conclusions that can be drawn about the formation of Pk2, an interesting replication phenotype was observed. *AUG1STOP* and *AUG1STOP-Pk2-Comp* replicate to wild-type titres in C6/36 cells, in spite of the introduced stop codon ($p \geq 0.05$) (Fig 5.12B)ii)). The introduction of a stop codon presumably prevents production of full-length nsP1, which therefore appears to be dispensable to CHIKV replication in mosquito cells. However, *AUG1mut*, which lacks the authentic start codon AUG1, is incapable of virus replication in C6/36 cells. This was previously assumed to be due to dependence on full-length nsP1 for CHIKV replication in mosquito cells. In light of the ability of *AUG1STOP* to replicate in the absence of full-length nsP1, it appears that the sequence of AUG1 is essential for viral replication in mosquito cells but not the functionality of AUG1 as a start codon in producing full-length nsP1.

Unlike in C6/36 cells, the introduction of a stop codon significantly reduced CHIKV replication in Huh7 cells, as may be expected (Fig 5.12B)i)). *AUG1STOP* and *AUG1STOP-Pk2-Comp* exhibited significantly reduced virus replication relative to WT in Huh7 cells ($p \leq 0.01$). *AUG1mut* displayed a complete lack of infectious virus release, as was observed in C6/36 cells. These data suggested that the sequence and function of AUG1 are necessary for CHIKV replication in mammalian cells. Interestingly, *AUG1STOP-Pk2-Comp* did not rescue *AUG1STOP* replication, indicating that restoring base-pairing in Pk2 had no effect on virus replication in the context of the stop codon.

A)i) *AUG1STOP* A)ii) *AUG1STOP-Pk2-Comp* A)iii) *AUG1mut*



B)

Mutant	Description
AUG1STOP	Stop codon introduced between AUG1 and AUG2 Disruption of Pk2 base-pairing
AUG1STOP-Pk2-Comp	Stop codon introduced between AUG1 and AUG2 Restoration of Pk2 base-pairing by compensatory mutation downstream
AUG1MUT	AUG1 start codon removed by substitution Met>Ala Pk2 base-pairing intact

C)i) C)ii)

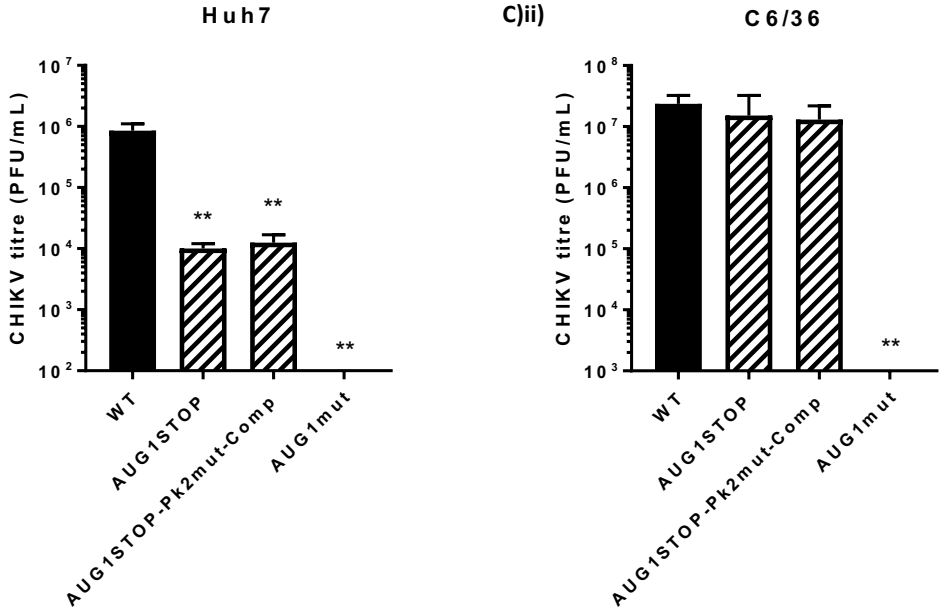


Figure 5.12: Introduction of a stop codon between the authentic and alternative start sites is tolerated in mosquitoes. (A) Schematic representation of **(i)** *AUG1STOP,* **(ii)** *AUG1STOP-Pk2-Comp* and **(iii)** *AUG1mut* displaying mutations (red). *AUG1STOP* and *AUG1STOP-Pk2-Comp* contain a stop codon (red box) within SL85 between the two in-frame AUG start codons, AUG1

and AUG2 (grey arrows). Dotted lines represent potential base-pairing in Pk2; black lines represent base-pairing, red lines represent disrupted pairing. *AUG1mut* contains a non-synonymous (Met>Ala) mutation which replaces the first AUG start codon with a GCG (yellow arrow). **(B)** Description of each mutant. **(C)** CHIKV titres following MOI=1 infection with virus of wild-type RNA sequence (black bar) and mutants (hatched bars) over 24 hours in **(i)** Huh7 and **(ii)** C6/36 cells (n=3). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: $p \leq 0.05$ (*), ≤ 0.01 (**). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

5.2.5 Alternative UUG start codon

In addition to reverse genetic studies demonstrating the importance of the authentic and alternative start sites AUG1 and AUG2, selection pressure acting on this region was revealed during passaging of CHIKV in mammalian cells. Serial passaging of WT CHIKV and mutants in Huh7 cells revealed a preference for UUG at the alternative Met^{24} start site. A single point mutation was present in the P10 sequence of all constructs passaged in mammalian cells, including WT and each of the mutants *SL47mut, SL85mut, SL102mut, SL165mut* and *SL194mut*, indicating a mammalian-specific cell culture adaptation (Fig 5.13A). This non-synonymous $A_{146}>U_{146}$ substitution has several effects. Firstly, primary RNA sequence is altered. Secondly, RNA structure is altered through the replacement of an A_{146}:U_{79} base-pair with a U_{146}/U_{79} bulge (Fig 5.13B). Thirdly, the nsP1 protein sequence is altered by $Met^{24}>Leu^{24}$ substitution, potentially affecting the protein structure. Finally, AUG>UUG alters the alternative start codon AUG2, likely reducing expression of the N-terminally truncated form of nsP1.

UUG encodes leucine but is also known to function as an alternative start codon, albeit less efficiently than AUG (445). The selection of the point mutant AUG>UUG in 7 separate passage series, including twice in the wild-type virus, suggests that less efficient expression of the truncated form is desirable in mammalian cells. The AUG>UUG mutation was fixed in each stem-loop mutant at a different passage (Fig 5.14). *SL194mut* and WT virus fixed the UUG substitution extremely quickly, within 1-2 passages, whereas *SL85mut* acquired the substitution much later, at P9.

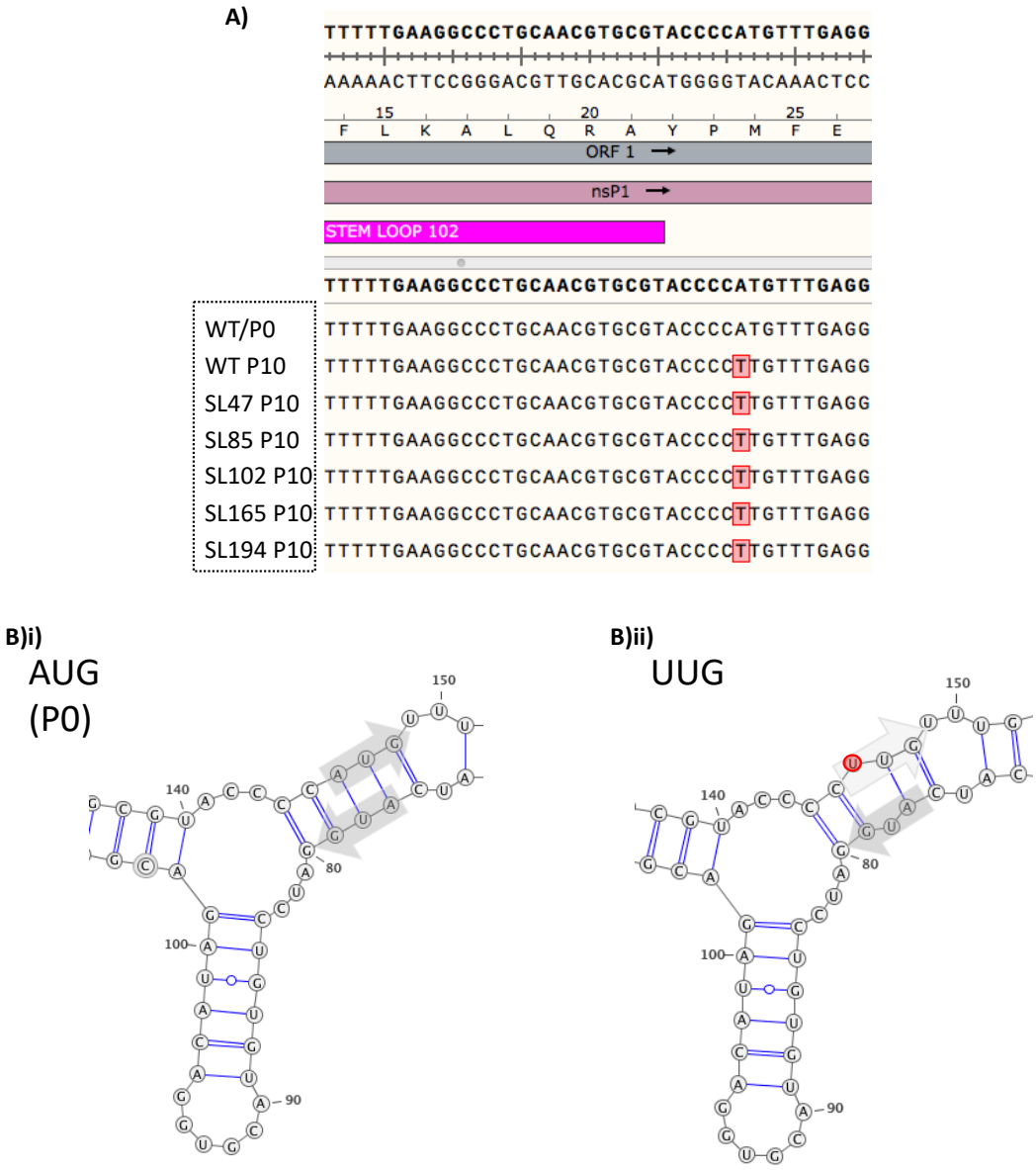


Fig 5.13: The alternative AUG start codon is invariably substituted with UUG during mammalian cell passage. (A) Presence of the AUG>UUG mutation (red box) for each stem-loop mutant at passage 10 (P10) in Huh7 cells. The P10 sequence for each passaged mutant is aligned below the WT P0 sequence containing AUG. **(B)** Schematic representation of UNAfold-predicted structures for **(i)** WT (P0) sequence containing base-paired AUG:AUG and **(ii)** P10 sequence containing UUG. The AUG1 start codon of nsP1 is denoted by a grey arrow and the A_{146}>U_{146} mutation is circled in red.

Stem-loop	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
WT(1)	×	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
WT(2)	×	×	×	×	×	✓	✓	✓	✓	✓	✓
SL47	×	×	×	×	×	×	✓	✓	✓	✓	✓
SL85	×	×	×	×	×	×	×	×	×	✓	✓
SL102	×	×	×	×	×	×	×	✓	✓	✓	✓
SL165	×	×	×	×	×	×	✓	✓	✓	✓	✓
SL194	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 5.14: Acquisition of AUG>UUG point mutation during passage. Acquisition of AUG>UUG point mutation in Huh7 cells for each stem-loop over 10 passages (P1-P10) for wild-type (WT) and stem-loop (SL) mutants. Ticks represent A_{146}>U_{146} mutation (UUG) detectable at that passage, crosses represent WT (AUG) sequence at that passage. WT and WT(2) represent independent passages of WT virus, since *SL47mut* and *SL194mut* were passaged alongside WT(1) prior to passage of SL85-SL165 alongside WT(2), providing two WT passages.

The strong selection pressure for the AUG>UUG substitution suggests that there may be a fitness benefit associated with UUG in mammalian cell culture. In order to examine the effect of this substitution, one-step growth assays were carried out for P0 and P10 virus derived from the WT in Huh7 and C6/36 cells (Fig 5.15A). While the P10 virus containing UUG exhibited an increased titre compared to P0 virus in Huh7 cells, this effect was not statistically significant. However, the same virus stock exhibited an ~3 log reduction in CHIKV titre in C6/36 cells compared to its P0 counterpart ($p \leq 0.005$). It may be expected that other mammalian-specific mutations had accumulated in the structural proteins of Huh7-passaged CHIKV over 10 passages, explaining the reduction in viral replication in mosquito cells by a reduction in infectivity or packaging efficiency. In order to determine whether the AUG>UUG substitution was responsible for decreasing CHIKV replication in mosquito cells, the $A_{146}>U_{146}$ mutation was incorporated into the CHIKV_IC construct to produce *UUGmut*. Huh7 and C6/36 cells were infected with *UUGmut* and WT virus at MOI=1 (Fig 5.13B). *UUGmut* significantly decreased CHIKV replication in C6/36 cells ($p \leq 0.0001$) while producing a small but non-significant increase in CHIKV titre in Huh7 cells. The detriment to CHIKV replication in *Ae. albopictus* cells explains the preference for AUG at the second start site in wild-type isolates. The replacement of the alternative start site with a less efficient codon may suggest a preference for altered regulation in mammalian cells, downregulating but not eliminating the truncation product. The selection of the codon UUG in multiple independent isolates, rather than the other viable leucine start codon CUG, indicates an advantage in mammalian cells which may be due to improved stability or flexibility of the local RNA structure, formation of alternative RNA structures including long-distance interactions or interaction with mammalian-specific binding factors. It cannot be ruled out that the selection of UUG is due to a tissue culture-specific benefit. Further analysis of AUG:AUG base-pairing is required to determine the effect of this interaction on translation and replication.

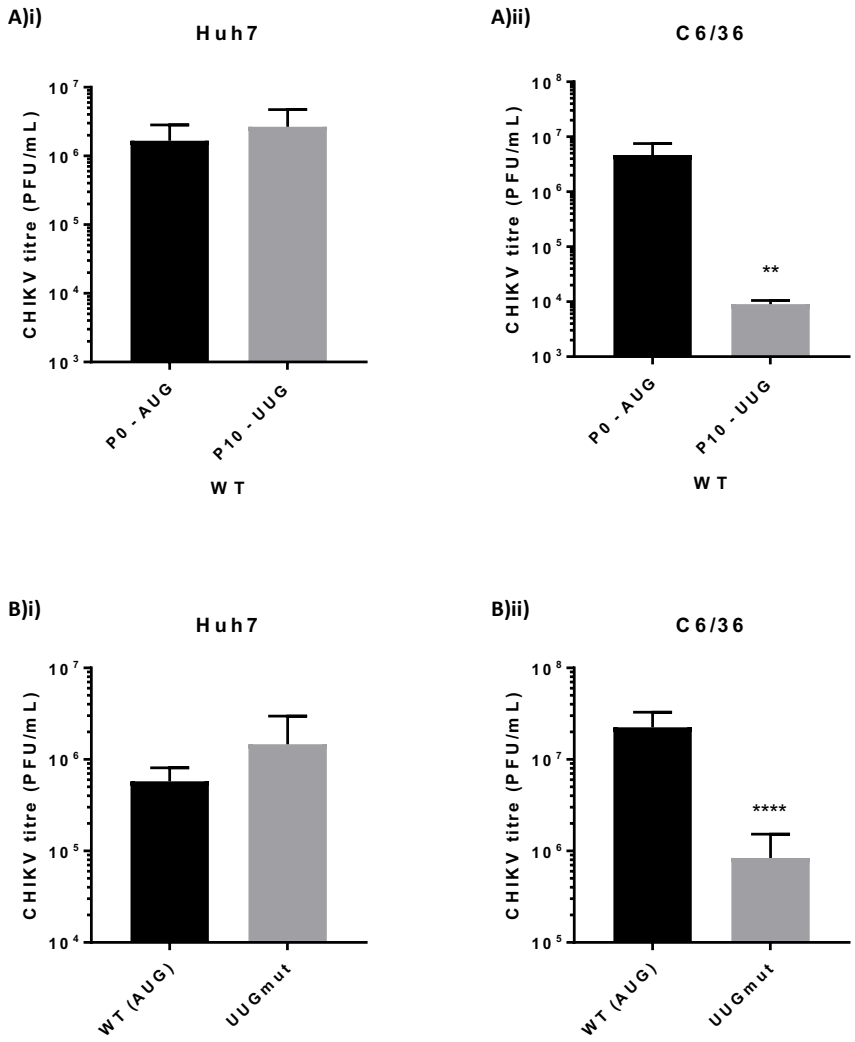


Figure 5.15: Substitution of AUG>UUG at the alternative start site is not tolerated in mosquitoes. (A) Replication phenotype of P0 WT CHIKV (black) compared to P10 WT-derived CHIKV containing the UUG mutation (grey) following infection of **i)** Huh7 and **ii)** C6/36 cells over 24 hours (n=3). **(B)** Replication phenotype of WT virus (black) compared to *UUGmut* virus (grey), synthesised by site-directed mutagenesis, following infection of **(i)** Huh7 and **(ii)** C6/36 cells over 24 hours (n=3). * represents statistical significance for each mutant compared to wild-type under a two-tailed Student's *t*-test: p≤0.05 (*), ≤0.01 (**), ≤0.001 (***), ≤0.0001 (****). Data shown is the mean of three independent biological replicates, with the error bar representing the standard deviation of the mean.

5.3 Discussion

5.3.1 Secondary structures in the 5′ of the CHIKV genome are essential to genome replication

Combined destabilisation of SL47, SL85, SL102, SL165, SL194 and SL246 abolished CHIKV genome replication in human- and *Ae. albopictus*-derived cells, demonstrating that RNA structures within the 5′ 300 nt of the CHIKV genome are essential for CHIKV replication in both host species (Figs 5.3 & 5.4). A total of 17 mutations in the 5′ 300 nt were sufficient to prevent production of infectious virus (Fig 5.2). A similar result was observed by Michel and colleagues upon extensive silent mutation of the 5′ 300 nt of VEEV, whereby 95 mutations prevented production of infectious virus in BHK-21 cells. Interestingly, the same study found that 97 synonymous mutations predicted to disrupt the secondary structure in the 5′ 300 nt of SINV did not completely abolish virus replication, although viral titre was reduced by ~3-4 logs in both BHK-21 and $C_7$10 cells (376). In contrast, an earlier study in SINV had demonstrated that 44 synonymous mutations in the same region resulted in barely detectable viral titres following infection of BHK-21 cells and no infectious virus recovery from mosquito cells (357). Overall, it can be concluded that the structured region in the 5′ 300 nt of alphavirus genomes plays a critical role in viral replication and is essential in CHIKV.

Clearly, the identity and location of substitutions during mutagenesis plays a major role in replication phenotype when disrupting alphavirus RNA structures. Far fewer mutations were necessary in CHIKV than in VEEV and SINV to abolish viral replication. Furthermore, 44 mutations resulted in abolition of SINV replication in one study, whereas 97 mutations were not sufficient in a later study. This may complicate the interpretation of previous findings and explain disagreement between phenotypic studies of the 51 nt CSE (352,376). In future studies, consideration of RNA sequence and structure, codon usage and dinucleotide frequency during design of targeted mutations may reduce variability between studies. It is important to note that four of the mutations present in the CHIKV *CombmutA* were in SL47, which has no homologue in VEEV or SINV. Therefore, the disruption of secondary structure in this region of CHIKV is not as directly comparable as between VEEV and SINV.

As discussed in Chapter 3, the replication-competent sub-genomic replicon CHIKV_Rep cannot distinguish between early stages of genome replication, such as translation of the ORF-1 non-structural polyprotein and initiation of transcription by the RdRp. The lack of genome replication for *CombmutA* (Fig 5.4) and *CombmutB* (Fig 5.6B) observed during assays of CHIKV_Rep could be the result of impaired translation. The translation phenotype of *CombmutA* could be determined using the CHIKV_Rep(GDD>GAA) construct, which undergoes initial translation to express the reporter in ORF-1 but expresses non-catalytic RdRp, preventing subsequent transcription events (Fig 3.13). Although individual stem-loop disruptions were shown not to impact translation (Fig 3.14), *CombmutA* and *CombmutB* contain more extensive changes to RNA structure and sequence. Further study is necessary to formally discount translational inhibition as the cause of impaired CHIKV replication for these mutants. Alternative methods of distinguishing the effects of translation and replication of the viral genome include *trans*-replicase studies. Translation and replication are uncoupled in such assays by the expression of the non-structural proteins in *trans* from an alternative promoter. The mutated genome acts as a template and expresses a reporter in place of viral proteins. In this way, the ability of the mutant sequence to act as a template for replication can be assayed in isolation. This approach has been successfully employed in the study of CHIKV in both human and mosquito cells (446). The impact of RNA stem-loops on CHIKV translation could also be further assessed by Western blot analysis of nsP expression or through polysome profiling to quantify the association of ribosomes with viral RNA.

Synonymous mutations alter codon usage and dinucleotide frequency, as discussed in Chapter 3. Thus, the replication phenotypes observed for the combined mutants may be the result of reduced codon efficiency or enhanced immune recognition of the genome rather than disruption of RNA structure. However, there was no net increase in CpG frequency in any of the combined mutants and the net increase in UpA frequency did not exceed one pair (Fig 5.16). Therefore, dinucleotide frequency is not a likely explanation for the abolition in viral replication observed for *CombmutA* and *CombmutB*. Similarly, codon usage cannot account for abolition of replication as changes to codon usage did not reduce net codon efficiency more than 15% for any of the combined mutants (Fig 5.17).

	CpG	UpA
Combinedmut	0	0.5
CombA	0	0.2
CombB	0	1

Figure 5.16: Changes in CpG and UpA dinucleotide frequency in combined mutants. Net change in CpG and UpA dinucleotide frequencies in each of the combined mutants, expressed as the net number of dinucleotide pairs introduced by the mutations displayed (Figs 5.2 & 5.5).

	Percentage change in codon efficiency	
	Human	*D. melanogaster*
Combinedmut	-6.6	-1.0
CombA	-3.0	-0.1
CombB	-14.8	-3.0

Figure 5.17: Changes in codon efficiency in combined mutants. Net change in codon efficiency for each of the combined mutants in human- and *D. melanogaster*-derived cells. Changes to codon efficiency are displayed as a relative percentage decrease in codon efficiency in each organism upon mutation. Each percentage is an average of the change induced by all substitutions per mutant. Standard deviation for each average was equal to or less than 10%. Codon efficiency is measured as codon frequency in the genome of each organism, as drawn from the codon frequency tool from Genscript (405). SL47 is in the untranslated region, thus codon usage is null.

5.3.2 Secondary structures exhibit synergy and functional redundancy

CombmutA contained disruptive mutations in each stem-loop in the region of interest: SL47, SL85, SL102, SL165, SL194 and SL246 (Fig 5.2). *CombmutB* and *CombmutC* contained mutations in only those stem-loops which enhanced replication in either human- or *Ae. albopictus*-derived cells respectively (Fig 5.5). These mutants served two purposes during the study. Firstly, in cell lines where the stem-loops had exhibited functionality, individual and combined mutant phenotypes were compared to reveal interactions between stem-loops. In the case of no functional interactions between stem-loops, an additive effect would be expected, representing the sum of the individual replication phenotypes. In the case of functional interactions between stem-loops, a synergistic effect would be expected, whereby the combined mutant replication phenotype exceeds the sum of individual phenotypes. Secondly, in cell lines where the individual stem-loop mutants exhibited no replication phenotype, functional redundancy could be determined by the presence of a replication phenotype for the combined mutant.

Genome replication was not detectable following transfection of Huh7 and C6/36 cells with *in vitro* transcribed *CombmutA* CHIKV_Rep RNA (Fig 5.4). The abolition of genome replication for *CombmutA* in Huh7 cells suggested that essential synergistic interactions occur between stem-loops, since replication was not abolished in assays of any individual stem-loop mutant (Chapter 3). The predicted additive phenotype for *CombmutA*may be expected to be relatively large due to the number of structures disrupted. However, genome replication for *CombmutA* was measured 24-hours post-transfection, when the additive phenotype was predicted to be no more than a ~1.5 log decrease in genome replication relative to WT (Fig 3.9A-F). In light of the ~3-4 log decrease in genome replication to below detectable levels, synergistic interaction between the stem-loops in Huh7 cells was concluded. As might be expected, *CombmutB* was similarly incapable of genome replication, indicating that SL246 was not sufficient to restore genome replication (Fig 5.6B). This was not surprising, given the lack of phenotype for SL246 in mammalian cells when individually disrupted (Fig 3.9F). The synergistic interaction occurring within the 5′ 300 nt of the genome in mammalian cells involves two or more of the stem-loops which had exhibited individual phenotypes: SL47, SL85, SL102, SL165 and SL194.

In addition to synergy between stem-loops in mammalian cells, the abolition of genome replication in C6/36 cells for *CombmutA* suggested synergistic interactions in the mosquito host. Since *CombmutA* cannot form SL47 or SL246, each of which exhibited an individual replication phenotype in C6/36 cells, it was hypothesised that the essential synergistic interaction may occur between these stem-loops (Fig 3.12A & F). Interestingly, *CombmutB* was also incapable of genome replication in C6/36 cells, demonstrating that the reintroduction of SL246 was not sufficient to restore genome replication (Fig 5.6B). Given that SL47 is the only stem-loop disrupted in *CombmutB* which had demonstrated an individual phenotype, the replication phenotype of *CombmutB* in C6/36 cells should be equal to that of *SL47mut* (Fig 3.12). This was not the case – *SL47mut* exhibits a ~1 log decrease in genome replication while *CombmutB* exhibits a ~2.5 log decrease in replication to below detectable levels. These data suggest that SL85, SL102, SL165 and SL194, or some combination of these stem-loops, play an essential role in CHIKV genome replication in C6/36 cells. Since no phenotype was previously detected during individual disruption of these stem-loops, functional redundancy was concluded to exist between one or more of these structures in mosquito cells. Host-specific functional redundancy has been demonstrated in DENV, as a means of mitigating the effect of host-switching (393).

In contrast to *CombmutB, CombmutC* infectious virus was detectable following infection of Huh7 and C6/36 cells (Fig 5.6C). The ability of *CombmutC* to replicate in C6/36 cells precludes the hypothesis that a synergistic interaction between SL47 and SL246 is essential for viral replication in the mosquito host. A small but significant synergistic interaction did occur, whereby the phenotype exceeds the additive phenotype by ~0.5 logs, but this was not essential to CHIKV replication. An even greater synergistic effect was observed between SL47 and SL246 in Huh7 cells. Since *SL246mut* displayed no replication phenotype in Huh7 cells during previous assays, the replication phenotype of *CombmutC* should be equal to that of *SL47mut*. This was not the case – *SL47mut* exhibited a ~1 log decrease in viral titre relative to WT in Huh7 cells while *CombmutC* exhibited a ~2.5 log decrease. Therefore, a synergistic interaction between SL47 and SL246 in Huh7 cells was concluded to occur. In a similar pattern to SL85-194 in mosquito cells, host-specific functional redundancy occurred for SL246 in mammalian cells. The results from *CombmutC* assays highlight the benefits of pairwise disruptions in revealing functional redundancy and synergy of stem-loops in CHIKV replication. Further pairwise and multiple

disruptions should be designed to dissect the minimum set of stem-loops required to prevent or restore CHIKV replication e.g. reintroduction of SL47 may restore replication in *CombmutB* and pairwise disruption of SL85/102 and SL165/194 in *CombmutC* may abolish replication. A mutant restoring SL47 in *CombmutB* would enable investigation of functional redundancy in SL85-194.

As discussed earlier, the genetic stability of attenuation is a critical factor in the consideration of vaccine candidates. *CombmutC* is attenuated in both human and mosquito cell lines but not incapable of replication in either, making it a possible component in an attenuated vaccine strain. Single- and multi-host passaging will be necessary to determine the stability of the combined mutations. Passaging may also reveal potential routes of reversion, whether in the proximal stem-loops SL85-194 or in interacting proteins which may be attenuated in turn during vaccine development.

5.3.3 Alternative conformations of SL85

In Chapter 3, SL85 was demonstrated to be essential for efficient CHIKV genome replication in human-derived cells (Fig 3.5). In Chapter 4, structure-led reverse genetic analysis, destabilising SL85 and then restoring base-pairing with compensatory substitutions, demonstrated that SL85 functions through a structure-dependent mechanism (Figs 4.2A & 4.3A). However, SHAPE reactivity data for SL85 suggests that this region of the genome may be structurally dynamic and able to form alternative interactions (Figs 5.7 & 5.8A).

The formation of alternative RNA structures is well established in several positive-sense RNA viruses, as a means of temporal and spatial control of viral replication (343,447). Conformational switching between alternative structures may involve local rearrangement or modulation of long-range RNA:RNA interactions. Local conformational changes are exemplified by riboswitches, such as the tertiary L-box motif of *Bacillus subtilis lysC* mRNA which undergoes local structural rearrangement in the presence of lysine (448). Two possible local alternative conformations around SL85 were investigated during this study, pseudoknot-1 (Pk1) and pseudoknot-2 (Pk2) (Fig 5.8). Pk1 was predicted to form between the terminal loop of SL85 and the upstream start codon AUG1. Therefore, mutations were designed in close proximity to the start codon which disrupted and

restored base-pairing in the pseudoknot. The sequence surrounding a start codon is known to impact the efficiency of translation (449). While the formation of a tertiary structure involving AUG1 may be selected to regulate translation, base-pairing of the Kozak sequence presented two major difficulties for reverse genetic analysis. Firstly, of the three mutants assayed, only those with a mutated Kozak sequence exhibited a reduction in viral titre, suggesting reduced translational efficiency (Fig 5.9). The translation phenotypes of *Pk1$_S$mut* and *Pk1-Comp* could be determined using the CHIKV_Rep(GDD>GAA) construct or a *trans*-replicase system in order to confirm reduced translation. Secondly, the C_{76}>G_{76} mutation present in both *Pk1$_S$mut* and *Pk1-Comp* disrupts a C:G base-pair between AUG1 and AUG2 (Fig 5.9). The potential for competing base-pairing between AUG1:AUG2 and AUG1:SL85 further complicated reverse genetic analysis of the region, as mutations designed to disrupt one interaction inevitably affected the other.

Reverse genetic analysis indicated that Pk1 is unlikely to form. *Pk1-Comp* did not rescue viral replication relative to *Pk1$_S$mut* and in fact exhibited a more severe phenotype. Restoring base-pairing in Pk1 did not rescue replication, suggesting that the phenotype observed for *Pk1$_S$mut* was not due to the disruption of the pseudoknot Pk1. Additionally, *Pk1$_L$mut* replicated to wild-type titres, suggesting that the mutated nucleotides were not involved in base-pairing as proposed in Pk1. However, only two of the seven predicted base-pairs in Pk1 were fully disrupted by the mutations made in *Pk1$_L$mut,* since C_{91}>U_{91} produces a G_{79}:U_{91} non-Watson-Crick pair. Therefore, *Pk1$_L$mut* may still form the pseudoknot Pk1. Due to the restrictions imposed by synonymous mutation, the range of alternative mutations was extremely limited. Non-synonymous mutants could be designed followed by *in vitro* SHAPE mapping to probe for disruption and restoration of tertiary structure. Biochemical determination of RNA structure would be particularly useful as it allows more extensive mutation of the loop region as well as mutation of the start codon and Kozak sequence. Interestingly, *Pk1$_S$mut* replicated to wild-type titres in C6/36 cells, suggesting less stringent restrictions on Kozak sequence in mosquito cells.

Pk2 was predicted to form between the terminal loop of SL85 and a downstream sequence in SL102 (Fig 5.10), therefore the Kozak sequence and disruption of AUG:AUG pairing was not an issue during design of synonymous mutants. However, the range of potential silent mutations in the terminal loop region was limited. Only two nucleotides predicted to base-pair in Pk2 could be synonymously mutated: G_{88} and C_{91}. C_{91}>U_{91} was the only synonymous mutation that could be

made at nt 91 and this had been made in *Pk1$_L$mut*. Additionally, C_{91}:G_{79} in Pk1 corresponds to C_{91}:G_{139} in Pk2, meaning that in both cases, C>U at this position produced a G:U non-Watson-Crick base-pair as opposed to full disruption. It was not possible to design a synonymous structural mutant of the terminal loop which disrupted Pk2 but not Pk1. In fact, all synonymous mutants predicted to disrupt base-pairing of the terminal loop of SL85 disrupted Pk1 to a greater degree than Pk2. Thus, the wild-type level of replication displayed by *Pk1$_L$mut* also demonstrated that Pk2 is unlikely to form. However, two silent mutations in *Pk2$_S$mut* were sufficient to reduce viral replication by ~2 logs in Huh7 cells, indicating a function for this region of SL102 during viral replication. *AUG1STOP* represents a non-synonymous mutant of the terminal loop which disrupts Pk2 base-pairing but not that of Pk1 (Fig 5.12A)i)). The compensatory mutant *AUG1STOP-Pk2-Comp* contains the mutations of both *AUG1STOP* and *Pk2$_S$mut,* designed to restore base-pairing in Pk2 (Fig 5.12A)ii)). Although the mutants are non-synonymous, a comparison of *AUG1STOP* and *AUG1STOP-Pk2-Comp* indicated that restoring base-pairing in Pk2 did not rescue wild-type levels of replication. Interestingly, no additive effect was observed for *AUG1STOP-Pk2-Comp;* the replication defect was equal to that of either of the component mutants *Pk2$_S$mut* and *AUG1STOP*. As described for Pk1, SHAPE mapping of Pk2 may provide structural data without the restrictions imposed by phenotypic analysis. Equally, *trans*-replicase studies could reveal the importance of each putative pseudoknot in the template.

Alternative structures may form involving SL85 aside from Pk1 and Pk2. As discussed above, conformational switching between alternative structures may involve the formation of long-range RNA:RNA interactions. One example is SL9266 of HCV, which takes part in multiple, mutually exclusive long-range interactions with structures in the 5′ and 3′ non-coding regions (430). Sequences and structures from the 3′ UTR could be screened for likely interactions with the terminal loop of SL85, experimentally or *in silico.* Circularisation of alphaviral genomes is thought to occur by indirect PABP binding, as occurs with cellular mRNA. However, it is possible that direct interactions occur between the 5′ and 3′ termini of the CHIKV genome resulting in circularisation *in vivo.* This may be determined by *in vivo* SHAPE mapping. Alternatively, *in vitro* SHAPE mapping on a 5′ fragment of the CHIKV genome lacking the 3′ terminus may demonstrate altered reactivity in SL85. SHAPE studies of the CHIKV 5′ terminus to date have been carried out by NMIA treatment of full-length genomic RNA and probed only in the region of interest (13,396).

5.3.4 Alternative start codons

CHIKV replication required the AUG1 start codon in both Huh7 cells and C6/36 cells (Figs 5.12 & 5.18). This suggested a requirement for full length nsP1 during the viral lifecycle. However, the introduction of a stop codon between AUG1 and AUG2 was tolerated during CHIKV replication in C6/36 cells. Presumably, introduction of a stop codon prevents generation of full-length nsP1, producing a short polypeptide and the truncated form of nsP1 (Fig 5.19). The ability of *AUG1STOP* to replicate to wild-type titres in C6/36 cells may be explained by rapid reversion, which could be investigated in the manner described in Chapter 4 for passage mutants. In this case, rapid reversion must have occurred independently within 24 hours in *AUG1STOP* and *AUG1STOP-Pk2-Comp*. A second hypothesis to explain the tolerance for a stop codon between AUG1 and AUG2 in C6/36 cells was that the sequence of AUG1 and not functionality as a start codon is necessary for viral replication. AUG1 or base-pairing of AUG1:AUG2 may represent a binding motif for a *trans*-activating factor. Insertion of a stop codon immediately following AUG1 may provide further insight, by preventing translation elongation. Alternatively, the 4 aa peptide which would be translated from AUG1 prior to termination in *AUG1STOP* may be sufficient to enact a function of the N-terminus of nsP1 essential to replication. This could be determined by supplementing *AUG1mut* with the 4 aa peptide in *trans* to determine rescue of viral replication.

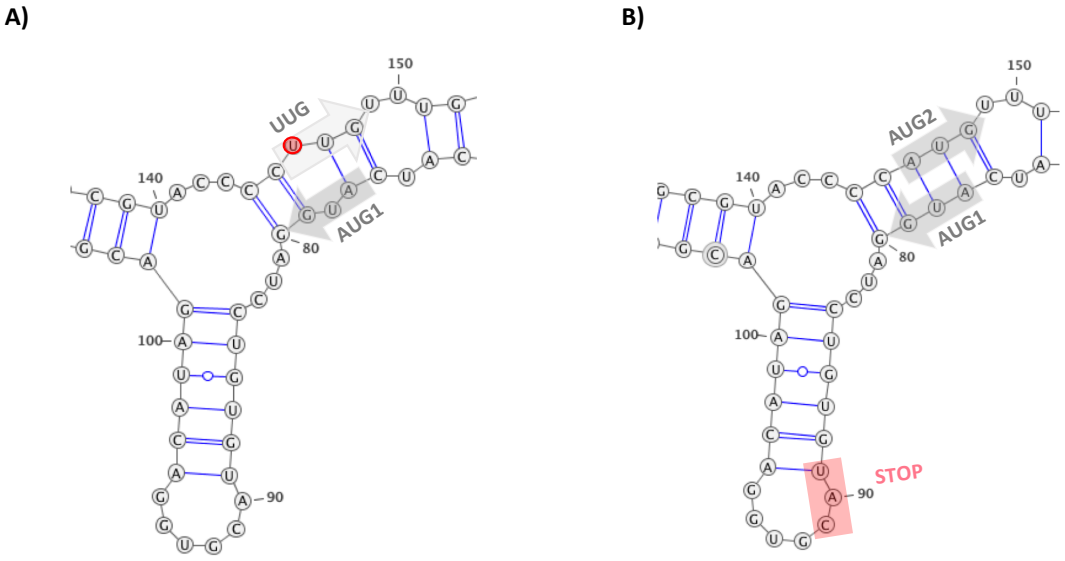


Figure 5.18: Different requirements for alternative translation initiation in *nsp1* in mosquitoes and humans. Mutations tolerated in **(A)** Huh7 and **(B)** C6/36 cells. Each mutant is capable of wild-type levels of replication in the given cell line. Start codons are denoted by a grey arrow and labelled according to position and type of start codon: AUG codons AUG1 and AUG2, and an alternative UUG codon. AUG>UUG mutation is denoted by a red circle. The introduced stop codon is denoted by a red box and labelled STOP.

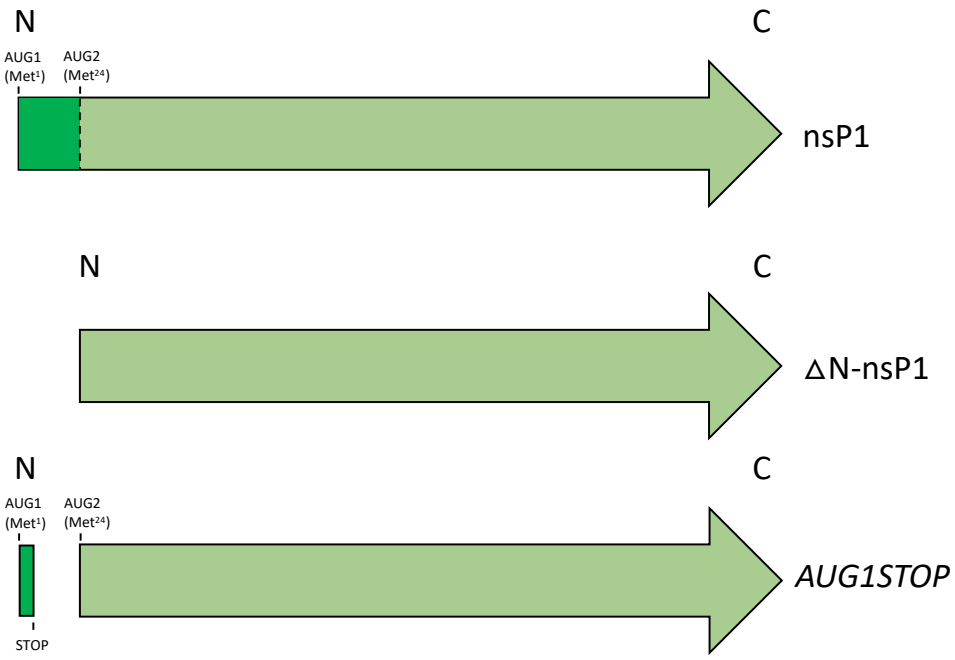


Figure 5.19: Truncation products resulting from introduction of a stop codon between the authentic and alternative start sites of nsP1. Schematic representation of CHIKV non-structural protein 1, showing wild-type nsP1 relative to the N-terminal truncated mutant (ΔN-nsP1) and *AUG1STOP*. The N-terminus (N) and C-terminus (C) are labelled, along with each start codon AUG1 (Methionine^{1}) and AUG2 (Methionine^{24}). The N-terminal sequence is denoted in darker green in wild-type nsP1. The stop codon introduced into *AUG1STOP* is labelled STOP.

A final hypothesis explaining tolerance of the stop codon was drawn from the work of Hemmings-Mieszczak and colleagues who demonstrated that initiation, elongation and termination of a short peptide was required for translation of a downstream open reading frame in cauliflower mosaic virus (CaMV) (450). "Toeprinting" of ribosomal initiation and elongation inhibition was used to demonstrate that AUG recognition and peptide release of a short upstream open reading frame was necessary for ribosomal shunting across a stable secondary structure and initiation at a downstream alternative start site. The upstream open reading frame was found to drive ribosomal shunting efficiently at only 3 aa in length. In order to explore this possibility in CHIKV replication, the sequence between AUG1 and the stop codon could be replaced with a random 9 nt (3 aa) sequence, as the sequence of the peptide was shown not to impact the process in the context of CaMV. Alternatively, deep sequencing methods for analysis of ribosomal distribution and accumulation on vRNA such as ribosome profiling (Riboseq) could be carried out to determine the relative usage of the alternative start site in human and mosquito cells.

Interestingly, efficient CHIKV replication requires an AUG codon at the second start site in C6/36 cells, while UUG is preferred at this site in Huh7 cells (Fig 5.15). The UUG codon was selected in multiple independent passages in Huh7 cells (Fig 5.14), whereas AUG was present at this site in every sequence recovered from C6/36 cells. AUG>UUG mutation represents Met>Leu at the amino acid level, thus the lack of tolerance for UUG in C6/36 cells may be due to non-synonymous mutation of nsP1 rather than reduced alternative translation efficiency. Investigation could be carried out by western blot to determine the presence of ΔN-nsP1 in C6/36 cells infected with *UUGmut*, alongside phenotypic analysis of a range of non-synonymous mutations at Met^{24}. This study is currently being undertaken by K. Loveday (University of Leeds).

Aside from alteration of protein sequence, UUG weakened the strength of the second start site (445) and reduced the number of hydrogen bonds present between AUG1:AUG2 (Fig 5.18A). A weaker context for translation initiation may downregulate production of ΔN-nsP1, preventing accumulation of the truncated product in the cell. The cellular location and function of ΔN-nsP1 is unknown at present, although the truncated nsP1 may be cytosolic if altered tertiary structure prevents anchoring of the replicase polyprotein to the plasma membrane. The preference for UUG may represent mammalian cell-specific culture adaptation which would not represent a selective

advantage during *in vivo* infection of a human host. It must be noted, however, that CUG (451) and GUG (452) have been demonstrated to act as effective alternative start codons in a range of viruses. UUG was selected specifically, perhaps because it represents the weakest translational context.

5.3.5 Conclusion

The secondary structure of the 5′ 300 nt of the CHIKV genome is essential for viral genome replication in human and mosquito hosts. Synergistic and functionally redundant interactions occur between stem-loops. There is also the possibility that SL85 exists in a dynamic tertiary interaction. Two proximal sites were investigated for potential pseudoknot formation with no conclusive result, suggesting that SL85 may be involved in long-range RNA:RNA interactions. The requirements for function of an alternative start codon vary between mosquito and human cells.

Chapter 6. Final Discussion

CHIKV is a major public health threat, re-emerging in recent years to establish local transmission on five continents (453). There are no currently available vaccines or direct-acting anti-viral therapeutics. A greater understanding of the CHIKV replication cycle is essential, as much of what is known about the replication cycle is assumed from related but divergent alphaviruses such as SINV and VEEV. Two regions of 5′ structure have been studied in SFV, SINV and VEEV in the context of viral genome replication. The 5′ CSE, conserved as SL3 in CHIKV, forms part of the core promoter for negative and positive strand synthesis in mammalian and mosquito cells. The importance of the 51 nt CSE during SINV and VEEV replication in mosquitoes has been repeatedly demonstrated, but there are conflicting reports of functionality in mammalian cells (13,350). This study represents the first investigation of functional elements within the 5′ UTR and nsP1-coding region in CHIKV.

Chapter 3 set out to determine the phenotypic importance of each stem-loop SL47-SL246 in human and mosquito cells at multiple stages of the CHIKV replication cycle. None of the individual stem-loops examined were essential for replication in either host. Rather, each stem-loop enhanced viral genome replication. Disruption of the stem-loops had no impact on translation of the viral genome, acting instead at the level of RNA synthesis. The results of the sub-genomic replicon assays in Chapter 3 demonstrate that the stem-loops impact early replication events, subsequent to initial translation of the genome and prior to the translation of the sub-genomic 26S RNA. Synthesis of positive- and negative-sense genomic RNA and sub-genomic RNA occurs during this phase of the viral replication cycle, as well as potential innate immune restriction. The phenotypes observed may be due to direct disruption of RNA synthesis mechanisms such as RNA-binding protein association or RNA:RNA interactions. Pull-down assays could be carried out to compare binding of wild-type and stem-loop mutant RNAs. The differential association of RNA-binding proteins, either cellular or viral in origin, with the stem-loop mutants could provide further insight into the mechanisms by which they function.

Alternatively, disruption of RNA stem-loops may render viral RNA vulnerable to recognition by innate immune mechanisms such as RIG-I or Mda5 binding. Similar pull-down studies may reveal

an association of innate immune factors with mutant RNAs, suggesting a role in evasion of host immunity. Knock-down or knock-out studies of known interactors, followed by measurement of cytokines and chemokines released during infection with mutant RNAs, could determine the extent and nature of immune modulation by the structures.

In order to determine which stage of RNA synthesis is affected by disruption of the stem-loops, further study could focus on RNA synthesis in *CombA*-electroporated cells. *CombA* -electroporation does not result in or sub-genomic RNA synthesis or production of infectious virus, as demonstrated during sub-genomic replicon and virus studies. However, *CombA* transfection may still result in negative-strand intermediate production. The qRT-PCR method of detecting positive- and negative-sense genomic RNA used in Chapter 3 could detect production of genomic RNA and differentiate between stages of RNA synthesis which depend on the stem-loops.

SL47, a novel stem-loop in the 5′ UTR which is highly conserved among the Old World alphaviruses, enhanced genome replication in human- and *Ae. albopictus*-derived cells (Fig 6.1). SL47 has no homologue in SINV or VEEV and these viruses lack the necessary region of the UTR entirely. This study is the first to examine the presence and function of SL47 in any alphavirus. Given the severe reduction in CHIKV replication in the absence of SL47, similar studies in ONNV and RRV may reveal novel targets for prevention of replication in other Old World alphaviruses.

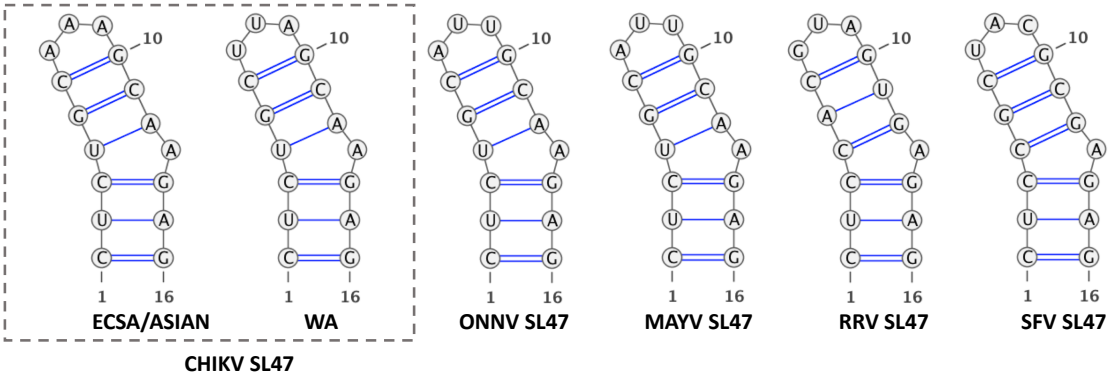


Figure 6.1: SL47 UNAfold predicted structures in several Old World alphaviruses. UNAfold predicted structure of SL47 for CHIKV ECSA/Asian/WA strains, O'nyong'nyong virus (ONNV), Mayaro virus (MAYV), Ross River virus (RRV) and Semliki Forest virus (SFV).

In contrast to SL47, SL85-SL246 enhanced genome replication in a host-dependent manner. SL85 and SL102 were mapped as distinct stem-loops in CHIKV, as opposed to the single elongated stem-loop mapped in SINV and VEEV (13). During Chapter 3, SL85 and SL102 were demonstrated to individually enhance CHIKV replication in mammalian cells. Deletion of the elongated stem-loop in SINV had no impact on viral replication in either host, except in the context of combined deletion of the 51 nt CSE (357). This suggests that the structural divergence between the SINV and CHIKV stem-loops represents functional divergence.

SL165 and SL194, representing the 51 nt CSE of CHIKV, individually enhanced CHIKV replication in mammalian cells but were dispensable in mosquito cells. This is in direct contrast to studies of the 51 nt CSE in SINV and VEEV which have consistently demonstrated the importance of the stem-loops for replication in mosquito cells (350,375). In addition, studies of the 51 nt CSE in VEEV have demonstrated that individual disruption of stem-loops in the 51 nt CSE is well tolerated in mammalian cells (376). Thus, the structural conservation of the 51 nt CSE in CHIKV may not mirror functional conservation. Alternatively, there may be functional redundancy between SL165 and SL194 in CHIKV replication in mosquito cells. Pairwise deletion of the stem-loops may severely inhibit CHIKV replication in mosquitoes, producing a similar phenotype to that seen in other alphaviruses.

A stem-loop downstream of the 51 nt CSE, SL246, was investigated for the first time during this study. SL246 was demonstrated to enhance replication in mosquito cells and was dispensable to replication in mammalian cells, representing the first mosquito-specific replication enhancer in alphaviruses, although the effect size is small. Phylogenetic analysis and *in silico* thermodynamic folding predictions suggest that SL246 homologues are present in a range of alphaviruses including the insect-specific EILV. Further investigation is required to determine whether SL246 has conserved functions within the *Alphavirus* genus. Interestingly, the host-specific phenotypes of stem-loops in *nsp1* were not temperature-dependent. Mammalian cells cultured at 28°C exhibited the same phenotypes as at 37°C, with the partial exception of SL85. Therefore, the basis of host-dependent functionality is likely due to differences in protein milieu. Comparative pull-downs in mosquito and human cells for wild-type and structural mutants may determine differential binding to the 5′ structures.

This study was limited to the European CHIKV vector, *Ae. albopictus*. As previously described, *Ae. aegypti* is a significant vector across the world, including North and South America. The functionality of the stem-loops investigated in this study remains to be seen in *Ae. aegypti* and could be determined by similar studies in *Ae. aegypti*-derived A20 cells. Both A20 and C6/36 cells have defective RNAi pathways. Alternative immunocompetent cell lines such as Aag2 and U4.4 could be optimised. Furthermore, infection and passaging studies would benefit from the use of *in vivo* mosquito models. Infection within an individual mosquito is far more complicated than can be measured using cell culture systems, in particular the bottlenecks of quasispecies diversity which occur between the midgut and haemocoel. Similarly, mouse models, 3D vascularised epidermal cultures or primary human muscle cell lines may provide further insight into the dependence of CHIKV on stem-loops during replication in mammalian tissues.

The sequence and structure requirements for stem-loop functionality were explored in Chapter 4 via phenotypic studies and passaging of mutants to examine potential routes of reversion. Each of the stem-loop mutants investigated could be rescued by compensatory mutations which restored base-pairing in the stem, indicating that the sequence of the stems was not essential for functionality. Furthermore, extensive mutations in the single-stranded bulges and terminal loop of SL102 and SL246 had no impact on CHIKV replication. Mutations in the terminal loop of SL194 produced a 10-fold decrease in viral genome replication, while terminal loop mutations of SL165 and SL194 reduced virus replication 100-fold. This is in agreement with the 100-fold decrease in SINV replication upon similar mutations in the 51 nt CSE (375).

Passaging of the stem-loop mutants resulted in direct reversion to the wild-type sequence, with a single exception in which a compensatory mutation became fixed. In contrast, studies in SINV and VEEV often produced pseudorevertants in interacting proteins such as nsP2 and nsP3. These studies passaged extensively mutated constructs (>40 mutations) or deletion mutants which prevented direct reversion. Extensive mutation (17 mutations) of CHIKV during Chapter 5 abolished replication, precluding this form of study. However, *CombmutC* displayed a low level of replication which may be sufficient to establish passaging studies. Pull-downs carried out by M. Müller (University of Leeds) demonstrated strong co-precipitation of nsP3 with a fragment

comprising the 5′ 300 nt of the CHIKV genome. Therefore, passaging of *CombmutC* could be carried out and any revertant viruses sequenced to probe for pseudoreversion, particularly in nsP3. A common theme during passaging was the incremental restoration of the stem-loops beginning with the apical region. Sequences capable of reforming the apical loop and stem of SL47, SL102 and SL194 were detected during early passages before full structure reformation. This suggests that partial structure restoration is preferable to sequence restoration at the base of the stem. Interestingly, stem-loops which had proven dispensable to replication in mosquito cells were conserved during passaging in C6/36 cells and conversely for SL246 in mammalian cells.

Although a single sequence for each mutant virus was presented for the passages 0-10 in Chapter 4, it is important to note that each passage contained a quasispecies. Each round of replication produces a staggering number of genomic variants of differing frequencies. The sequences presented in Chapter 4 are composites, determined by the most common nucleotide at each position, as measured by Sanger sequencing of a sample of the quasispecies population. The consensus sequence produced from this data is an average of the population and does not represent a single genomic sequence. More information about the makeup of the quasispecies could be gained using next-generation sequencing (NGS) methods. NGS methods exhibit higher sensitivity to detect low-frequency variants and higher numbers of reads, resulting in greater sequencing depth and more information about the prevalence of certain mutations within the quasispecies. NGS methods could also reveal the comparative complexity of quasispecies in mammalian and mosquito cells, which may account for the differences in relative reversion rates.

The potential for synergistic interactions and functional redundancy between the stem-loops was explored in Chapter 5. Synonymous site mutations which disrupted the stem-loops SL47, SL85, SL102, SL165 and SL194 in combination abolished viral genome replication in human and *Ae. albopictus* cells. This was particularly interesting given that SL85-SL194 had no individual impact on replication in mosquito cells. Therefore, it is likely that these stem-loops are functionally redundant in mosquito cells but not in mammalian cells. Host-specific redundancy has been observed in another arbovirus, DENV. Two stem-loops in the 3′ UTR of DENV, SL-I and SL-II, each impact viral replication in mosquito cells but are functionally redundant in the human host (393). Similarly, SL47 and SL246 exhibit synergistic interactions during infection of mammalian

cells while SL246 has no individual impact on replication in the same cells. Multiple and pairwise mutations will be necessary to untangle the minimum set of structures required for CHIKV replication in both hosts. The effect of extensive synonymous site mutation in the 3′ 500nt of alphaviral genomes is particularly contentious. One study in SINV introduced 43 synonymous mutations into the 5′ 300 nt which abolished replication in mosquito cells but had no impact on viral replication in mammalian cells (350). A second study in SINV mutated the homologues of SL85-SL194 and found a 50-fold decrease in viral replication in mammalian cells and a 5-fold decrease in negative strand synthesis *in vitro* (357). Similar mutations in VEEV abolished replication in both hosts (376). Thus, the conclusions made during this study are limited to CHIKV.

Chapter 5 also explored the potential for tertiary structure formation around SL85. Two proximal sequences with complementarity to the terminal loop of SL85 were probed for pseudoknot formation by reverse genetic analysis. Neither the upstream or downstream sequences could be rescued by double mutation, suggesting that the putative pseudoknots do not form. The analysis was complicated by the presence of the AUG1 initiation site within the upstream sequence and limitations of synonymous mutation within the terminal loop of SL85. SHAPE analysis of more extensively mutated pairings would more conclusively determine the potential for pseudoknot formation. Alternatively, long-range interactions may occur with the terminal loop of SL85, which could be investigated by SHAPE on fragments of the genome and pairs of fragments such as the 5′ and 3′ termini. In addition, *trans* replicase assays could determine the ability of the pseudoknot mutants to act as templates when the Kozak context is mutated.

Finally, single-host passaging revealed host-specific preferences around the alternative start site AUG2. UUG was selected at the second start site in all mammalian-passaged strains, which was poorly tolerated by mosquito cells. Conversely, mosquito cells tolerated the insertion of a stop codon between AUG1 and AUG2, which was not tolerated in mammalian cells. Therefore, the requirements for alternative translation differ between human and mosquito cells. Further work is needed to elucidate the different mechanisms of alternative translation. Studies of the shunting mechanism in cauliflower mosaic virus utilised initiation and elongation inhibitors in combination with ribosomal toeprinting to demonstrate the requirement for initiation at the first start site in

promoting translation from the second (450). A similar approach in CHIKV may determine whether shunting occurs across SL85 and SL102 during translation in mosquito cells.

Further work is needed to determine the mechanisms by which stem-loops in the 5′ region of CHIKV enhance replication. In particular, interacting partners such as nsP2 and nsP3 which have been identified in preliminary studies could be serially truncated to identify specific RNA binding sites. Conversely, the ability of non-structural proteins to bind to disrupted stem-loop mutants during EMSA could identify the sequence and structure requirements for interaction. Further study of the structure of the CHIKV genome would also provide insight into the mechanism of CHIKV replication, particularly exploration of tertiary structure by NMR or hydroxyl radical footprinting, investigation of dynamic changes over the course of the viral replication cycle and isolation of genomic RNA from different intracellular locations. The secondary structure map of the CHIKV 5′ 300nt provides a foundation for a deeper understanding of the mechanisms by which CHIKV utilises RNA structure to direct genomic replication.

In addition to broadening knowledge of CHIKV and alphaviral replication, the stem-loops characterised within this work represent targets for attenuation during vaccine design. RNA stem-loops within the CHIKV genome could be mutated, alongside independent targets such as envelope proteins, to produce a heavily attenuated live CHIKV vaccine. The data presented in Chapter 5 indicates that combinations of stem-loop mutants would be most effective in attenuating the virus and maintaining genetically stable mutations. Alternatively, the stem-loops may represent targets for drug design as has been achieved for HIV-1 RNA secondary structures with the drug NSC (323). Drug targeting may provide a specific therapeutic for CHIKV-induced disease which would greatly reduce the suffering associated with infection.

6.1 Conclusion

The secondary structure of the 5′ 300 nt of the CHIKV genome is essential for genome replication in human- and *Ae. albopictus*-derived cells. The novel stem-loop SL47 in the 5′ UTR functions in a host-independent manner while individual stem-loops in *nsp1* function in a host-dependent manner. Base-pairing within the stems represents the greatest determinant of replication enhancement, with a high tolerance for sequence substitution. Further research is necessary to

determine the mechanisms of action and tertiary interactions of the stem-loops during CHIKV replication.

Bibliography

1. Yactayo S, Staples JE, Millot V, Cibrelus L, Ramon-Pardo P. Epidemiology of chikungunya in the americas. J Infect Dis. 2016;214(Suppl 5):S441–5.
2. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Dempsey DM, Dutilh BE, et al. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). Arch Virol [Internet]. 2019;164(9):2417–29.
3. Weston JH, Welsh MD, McLoughlin MF, Todd D. Salmon pancreas disease virus, an alphavirus infecting farmed Atlantic salmon, Salmo salar L. Virology. 1999;256(2):188–95.
4. La Linn M, Gardner J, Warrilow D, Darnell GA, McMahon CR, Field I, et al. Arbovirus of Marine Mammals: a New Alphavirus Isolated from the Elephant Seal Louse, Lepidophthirus macrorhini. J Virol. 2001;75(9):4103–9.
5. Roundy CM, Azar SR, Rossi SL, Weaver SC, Vasilakis N. Insect-Specific Viruses: A Historical Overview and Recent Developments [Internet]. 1st ed. Vol. 98, Advances in Virus Research. Elsevier Inc.; 2017. 119–146 p.
6. Forrester NL, Palacios G, Tesh RB, Savji N, Guzman H, Sherman M, et al. Genome-Scale Phylogeny of the Alphavirus Genus Suggests a Marine Origin. J Virol. 2012;86(5):2729–38.
7. Vazeille M, Moutailler S, Coudrier D, Rousseaux C, Khun H, Huerre M, et al. Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus. PLoS One. 2007;2(11).
8. Zacks MA, Paessler S. Encephalitic alphaviruses. Vet Microbiol. 2010;140(3–4):281–6.
9. Doherty RL, Standfast HA, Domrow R, Wetters EJ, Whitehead RH, Carley JG. Studies of the epidemiology of arthropod-borne virus infections at Mitchell River Mission, Cape York Peninsula, North Queensland IV. Arbovirus infections of mosquitoes and mammals, 1967-1969. Trans R Soc Trop Med Hyg. 1971;65(4):504–13.
10. Stephenson EB, Peel AJ, Reid SA, Jansen CC, McCallum H. The non-human reservoirs of Ross River virus: A systematic review of the evidence. Parasites and Vectors. 2018;11(1):1–13.
11. Hahn CS, Lustig S, Strauss EG, Strauss JH. Western equine encephalitis virus is a recombinant virus. Proc Natl Acad Sci U S A. 1988;85(16):5997–6001.
12. Strauss JH, Strauss EG. The Alphaviruses : Gene Expression, Replication, and Evolution. Microbiol Rev. 1994;58(3):491–562.
13. Kutchko KM, Madden EA, Morrison C, Plante KS, Sanders W, Vincent HA, et al. Structural divergence creates new functional features in alphavirus genomes. Nucleic Acids Res. 2018;46(7):3657–70.
14. Nasar F, Palacios G, Gorchakov R V., Guzman H, Travassos Da Rosa AP, Savji N, et al. Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. Proc Natl Acad Sci U S A. 2012;109(36):14622–7.
15. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB, Estrada A, et al. Discovery of a novel alphavirus related to Eilat virus. J Gen Virol. 2017;98(1):43–9.
16. Torii S, Orba Y, Hang'ombe BM, Mweene AS, Wada Y, Anindita PD, et al. Discovery of Mwinilunga alphavirus: A novel alphavirus in Culex mosquitoes in Zambia. Virus Res [Internet]. 2018;250(April):31–6.

17. Hermanns K, Marklewitz M, Zirkel F, Overheul GJ, Page RA, Loaiza JR, et al. Agua Salud alphavirus defines a novel lineage of insect-specific alphaviruses discovered in the New World. J Gen Virol [Internet]. 2019;jgv001344.
18. Tschá MK, Suzukawa AA, Gräf T, Piancini LDS, da Silva AM, Faoro H, et al. Identification of a novel alphavirus related to the encephalitis complexes circulating in southern Brazil. Emerg Microbes Infect. 2019;8(1):920–33.
19. Weaver SC, Forrester NL. Chikungunya: Evolutionary history and recent epidemic spread. Antiviral Res [Internet]. 2015;120:32–9.
20. Arankalle VA, Shrivastava S, Cherian S, Gunjikar RS, Walimbe AM, Jadhav SM, et al. Genetic divergence of Chikungunya viruses in India (1963-2006) with special reference to the 2005-2006 explosive epidemic. J Gen Virol. 2007;88(7):1967–76.
21. Teo T-H, Her Z, Tan JJL, Lum F-M, Lee WWL, Chan Y-H, et al. Caribbean and La Réunion Chikungunya Virus Isolates Differ in Their Capacity To Induce Proinflammatory Th1 and NK Cell Responses and Acute Joint Pathology. J Virol. 2015;89(15):7955–69.
22. Chiam CW, Chan YF, Ong KC, Wong KT, Sam IC. Neurovirulence comparison of chikungunya virus isolates of the Asian and East/Central/South African genotypes from Malaysia. J Gen Virol. 2015;96(11):3243–54.
23. Wahid B, Ali A, Rafique S, Idrees M. Global expansion of chikungunya virus: mapping the 64-year history. Int J Infect Dis. 2017;58:69–76.
24. Langsjoen RM, Haller SL, Roy CJ, Vinet-Oliphant H, Bergren NA, Erasmus JH, et al. Chikungunya virus strains show lineage-specific variations in virulence and cross-protective ability in murine and nonhuman primate models. MBio. 2018;9(2):1–13.
25. Galatas B, Ly S, Duong V, Baisley K, Nguon K, Chan S, et al. Long-Lasting Immune Protection and Other Epidemiological Findings after Chikungunya Emergence in a Cambodian Rural Community, April 2012. PLoS Negl Trop Dis. 2016;10(1):1–10.
26. Auerswald H, Boussioux C, In S, Mao S, Ong S, Huy R, et al. Broad and long-lasting immune protection against various Chikungunya genotypes demonstrated by participants in a cross-sectional study in a Cambodian rural community article. Emerg Microbes Infect [Internet]. 2018;7(1).
27. Chen R, Wang E, Tsetsarkin K a., Weaver SC. Chikungunya Virus 3' Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces. PLoS Pathog. 2013;9(8).
28. Schneider A de B, Ochsenreiter R, Hostager R, Hofacker IL, Janies D, Wolfinger MT. Updated Phylogeny of Chikungunya Virus Suggests Lineage-Specific RNA Architecture. Viruses. 2019;11(9):1–16.
29. Ross BYRW. The Newala Epidemic. J Hyg (Lond). 1956;54(2):177–91.
30. Lumsden WHR. An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953 II. General description and epidemiology. Trans R Soc Trop Med Hyg. 1955;49(1):33–57.
31. Zeller H, Van Bortel W, Sudre B. Chikungunya: Its history in Africa and Asia and its spread to new regions in 2013-2014. J Infect Dis. 2016;214(Suppl 5):S436–40.
32. Weinbren MP, Haddow AJ, Williams MC. The occurrence of chikungunya virus in Uganda I. Isolation from mosquitoes. Trans R Soc Trop Med Hyg. 1958;52(3):253–62.
33. McCrae AWR, Henderson BE, Kirya BG, Sempala SDK. Chikungunya virus in the entebbe area of uganda: Isolations and epidemiology. Trans R Soc Trop Med Hyg. 1971;65(2):152–68.

34. Jupp PG, McIntosh BM. Aedes furcifer and other mosquitoes as vectors of chikungunya virus at Mica, northeastern Transvaal, South Africa. J Am Mosq Control Assoc. 1990;
35. Diallo M, Thonnon J, Traore-Lamizana M, Fontenille D. Vectors of Chikungunya virus in Senegal: Current data and transmission cycles. Am J Trop Med Hyg. 1999;60(2):281–6.
36. Kading RC, Borland EM, Cranfield M, Powers AM. Prevalence of antibodies to alphaviruses and flaviviruses in free-ranging game animals and nonhuman primates in the greater Congo basin. J Wildl Dis. 2013;49(3):587–99.
37. Althouse BM, Guerbois M, Cummings DAT, Diop OM, Faye O, Faye A, et al. Role of monkeys in the sylvatic cycle of chikungunya virus in Senegal. Nat Commun. 2018;9(1):1–10.
38. Hammon WMD, Rudnick A, Sather GE. Viruses associated with epidemic hemorrhage fevers of the Philippines and Thailand. Science (80-). 1960;
39. Volk SM, Chen R, Tsetsarkin KA, Adams AP, Garcia TI, Sall AA, et al. Genome-Scale Phylogenetic Analyses of Chikungunya Virus Reveal Independent Emergences of Recent Epidemics and Various Evolutionary Rates. J Virol. 2010;84(13):6497–504.
40. Banerjee K. A note on antibodies to chikungunya virus in human sera collected in Madras state in 1956. Indian J Med Res. 1965;
41. Chatterjee SN, Chakravarty MS, Chakravarty SK, Ray S, Sarkar JK. Survey of antibodies against chikungunya virus in the sera collected in Calcutta during 1964 and 1965. Indian J Med Res. 1967;
42. Pastorino B, Muyembe-Tamfum JJ, Bessaud M, Tock F, Tolou H, Durand JP, et al. Epidemic resurgence of Chikungunya virus in Democratic Republic of the Congo: Identification of a new central African strain. J Med Virol. 2004;74(2):277–82.
43. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006;3(7):1058–70.
44. Renault P, Solet JL, Sissoko D, Balleydier E, Larrieu S, Filleul L, et al. A major epidemic of chikungunya virus infection on Réunion Island, France, 2005-2006. Am J Trop Med Hyg. 2007;77(4):727–31.
45. Tsetsarkin K a., Vanlandingham DL, McGee CE, Higgs S. A single mutation in Chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog. 2007;3(12):1895–906.
46. Sergon K, Njuguna C, Kalani R, Ofula V, Onyango C, Konongoi LS, et al. Seroprevalence of Chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. Am J Trop Med Hyg. 2008;78(2):333–7.
47. Peyrefitte CN, Rousset D, Pastorino BAM, Pouillot R, Bessaud M, Tock F, et al. Chikungunya virus, Cameroon, 2006. Emerg Infect Dis. 2007;13(5):768–71.
48. Leroy EM, Nkoghe D, Ollomo B, Nze-Nkogue C, Becquart P, Grard G, et al. Concurrent chikungunya and dengue virus infections during simultaneous outbreaks, Gabon, 2007. Emerg Infect Dis. 2009;15(4):591–3.
49. Hertz JT, Munishi OM, Ooi EE, Howe S, Lim WY, Chow A, et al. Chikungunya and dengue fever among hospitalized febrile patients in Northern Tanzania. Am J Trop Med Hyg. 2012;86(1):171–7.
50. Lanciotti RS, Valadere AM. Transcontinental movement of Asian Genotype Chikungunya virus. Emerg Infect Dis. 2014;20(8):1400–2.
51. Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MUG, de Oliveira LF,

et al. Emergence and potential for spread of Chikungunya virus in Brazil. BMC Med [Internet]. 2015;13(1).
52. Teixeira MG, Andrade AMS, Da Costa MCN, Castro JSM, Oliveira FLS, Goes CSB, et al. East/central/South African genotype chikungunya virus, Brazil, 2014. Emerg Infect Dis. 2015;21(5):906–8.
53. White SK, Mavian C, Salemi M, Morris JG, Elbadry MA, Okech BA, et al. A new "American" subgroup of African-lineage Chikungunya virus detected in and isolated from mosquitoes collected in Haiti, 2016. PLoS One. 2018;13(5):1–11.
54. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli A, Panning M, et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet. 2007;370(9602):1840–6.
55. Grandadam M, Caro V, Plumet S, Thiberge JM, Souares Y, Failloux AB, et al. Chikungunya virus, southeastern France. Emerg Infect Dis [Internet]. 2011;17(5):910–3.
56. Bonilauri P, Bellini R, Calzolari M, Angelini R, Venturi L, Fallacara F, et al. Chikungunya virus in Aedes albopictus, Italy. Emerg Infect Dis. 2008;14(5):852–4.
57. Lindh E, Argentini C, Remoli ME, Fortuna C, Faggioni G, Benedetti E, et al. The Italian 2017 outbreak chikungunya virus belongs to an emerging aedes albopictus-adapted virus cluster introduced from the Indian subcontinent. Open Forum Infect Dis. 2019;6(1).
58. Delisle E, Rousseau C, Broche B, Leparc-Goffart I, L'Ambert G, Cochet A, et al. Chikungunya outbreak in Montpellier, France, September to October 2014. Euro Surveill Eur Commun Dis Bull. 2015;20(17).
59. C. C, M. G-G, F. F, C. J, M. A-C, G. G, et al. Preliminary report of an autochthonous chikungunya outbreak in France, July to September 2017. Eurosurveillance [Internet]. 2017;22(39):5–10.
60. Vega-Rua A, Zouache K, Caro V, Diancourt L, Delaunay P, Grandadam M, et al. High Efficiency of Temperate Aedes albopictus to Transmit Chikungunya and Dengue Viruses in the Southeast of France. PLoS One. 2013;8(3):1–8.
61. Trentini F, Poletti P, Baldacchino F, Drago A, Montarsi F, Capelli G, et al. The containment of potential outbreaks triggered by imported Chikungunya cases in Italy: A cost utility epidemiological assessment of vector control measures. Sci Rep. 2018;8(1):1–9.
62. Campbell LP, Luther C, Moo-Llanes D, Ramsey JM, Danis-Lozano R, Peterson AT. Climate change influences on global distributions of dengue and chikungunya virus vectors. Philos Trans R Soc B Biol Sci. 2015;370(1665):1–9.
63. Medlock JM, Vaux AG, Cull B, Schaffner F, Gillingham E, Pfluger V, et al. Detection of the invasive mosquito species Aedes albopictus in southern England. Lancet Infect Dis. 2017;17(2):140.
64. Mavalankar D, Shastri P, Bandyopadhyay T, Parmar J, Ramani K V. Increased mortality rate associated with chikungunya epidemic, Ahmedabad, India. Emerg Infect Dis. 2008;
65. Renault P, Josseran L, Pierre V. Chikungunya-related fatality rates, Mauritius, India, and Reunion Island. Emerg Infect Dis. 2008;14(8):1327.
66. Beesoon S, Funkhouser E, Kotea N, Spielman A, Robich RM. Chikungunya fever, Mauritius, 2006. Emerg Infect Dis. 2008;14(2):337–8.
67. CDC. CDC CHIKV geographic distribution [Internet]. [cited 2019 Dec 19]. Available from: https://www.cdc.gov/chikungunya/geo/index.html
68. MapChart.net. MapChart tool [Internet]. [cited 2019 Nov 4]. Available from:

https://mapchart.net/
69. Sourisseau M, Schilte C, Casartelli N, Trouillet C, Guivel-Benhassine F, Rudnicka D, et al. Characterization of reemerging chikungunya virus. PLoS Pathog. 2007;3(6):0804–17.
70. Labadie K, Grand R Le, Roques P, Labadie K, Larcher T, Joubert C, et al. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages Find the latest version : Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. J Clin Invest. 2010;120(3):894–906.
71. Chua HH, Abdul Rashid K, Law WC, Hamizah A, Chem YK, Khairul AH, et al. A fatal case of chikungunya virus infection with liver involvement. Med J Malaysia. 2010;65(1):83–4.
72. Sissoko D, Moendandze A, Malvy D, Giry C, Ezzedine K, Solet JL, et al. Seroprevalence and risk factors of Chikungunya virus infection in Mayotte, Indian Ocean, 2005-2006: A population- based survey. PLoS One. 2008;3(8):2005–6.
73. Appassakij H, Khuntikij P, Kemapunmanus M, Wutthanarungsan R, Silpapojakul K. Viremic profiles in asymptomatic and symptomatic chikungunya fever: A blood transfusion threat? Transfusion. 2013;53(10 PART 2):2567–74.
74. Simmons G, Brès V, Lu K, Liss NM, Brambilla DJ, Ryff KR, et al. High incidence of chikungunya virus and frequency of viremic blood donations during epidemic, Puerto Rico, USA, 2014. Emerg Infect Dis. 2016;22(7):1221–8.
75. Borgherini G, Poubeau P, Staikowsky F, Lory M, Moullec NL, Becquart JP, et al. Outbreak of Chikungunya on Reunion Island: Early Clinical and Laboratory Features in 157 Adult Patients. Clin Infect Dis. 2007;44(11):1401–7.
76. Economopoulou A, Dominguez M, Helynck B, Sissoko D, Wichmann O, Quenel P, et al. Atypical Chikungunya virus infections: Clinical manifestations, mortality and risk factors for severe disease during the 2005-2006 outbreak on Réunion. Epidemiol Infect. 2009;137(4):534–41.
77. Lemant J, Boisson V, Winer A, Thibault L, André H, Tixier F, et al. Serious acute chikungunya virus infection requiring intensive care during the reunion island outbreak in 2005-2006. Crit Care Med. 2008;36(9):2536–41.
78. Tandale B V., Sathe PS, Arankalle VA, Wadia RS, Kulkarni R, Shah S V., et al. Systemic involvements and fatalities during Chikungunya epidemic in India, 2006. J Clin Virol. 2009;46(2):145–9.
79. Gérardin P, Barau G, Michault A, Bintner M, Randrianaivo H, Choker G, et al. Multidisciplinary prospective study of mother-to-child chikungunya virus infections on the island of La Réunion. PLoS Med. 2008;5(3):0413–23.
80. Ng LFP, Chow A, Sun YJ, Kwek DJC, Lim PL, Dimatatac F, et al. IL-1β, IL-6, and RANTES as biomarkers of Chikungunya severity. PLoS One. 2009;4(1):1–8.
81. Borgherini G, Poubeau P, Jossaume A, Gouix A, Cotte L, Michault A, et al. Persistent Arthralgia Associated with Chikungunya Virus: A Study of 88 Adult Patients on Reunion Island. Clin Infect Dis. 2008;47(4):469–75.
82. Hoarau J-J, Jaffar Bandjee M-C, Krejbich Trotot P, Das T, Li-Pat-Yuen G, Dassa B, et al. Persistent Chronic Inflammation and Infection by Chikungunya Arthritogenic Alphavirus in Spite of a Robust Host Immune Response. J Immunol. 2010;184(10):5914–27.
83. Moro ML, Grilli E, Corvetta A, Silvi G, Angelini R, Mascella F, et al. Long-term chikungunya infection clinical manifestations after an outbreak in Italy: A prognostic cohort study. J Infect [Internet]. 2012;65(2):165–72.

84. Schilte C, Staikovsky F, Couderc T, Madec Y, Carpentier F, Kassab S, et al. Chikungunya Virus-associated Long-term Arthralgia: A 36-month Prospective Longitudinal Study. PLoS Negl Trop Dis. 2013;7(3).
85. de Araujo Eyer-Silva W, de Barros Pinto Neto H, da Silva GAR, de Almeida Ferry FR. A case of Chikungunya virus disease presenting with remarkable acute arthritis of a previously damaged finger joint. Rev Soc Bras Med Trop. 2016;49(6):790–2.
86. Chang AY, Martins KAO, Encinales L, Reid SP, Acuña M, Encinales C, et al. Chikungunya Arthritis Mechanisms in the Americas: A Cross-Sectional Analysis of Chikungunya Arthritis Patients Twenty-Two Months After Infection Demonstrating No Detectable Viral Persistence in Synovial Fluid. Arthritis Rheumatol. 2018;70(4):585–93.
87. Schwartz O, Albert ML. Biology and pathogenesis of chikungunya virus. Nat Rev Microbiol. 2010;8(7):491–500.
88. Tanabe ISB, Tanabe ELL, Santos EC, Martins W V., Araújo IMTC, Cavalcante MCA, et al. Cellular and Molecular Immune Response to Chikungunya Virus Infection. Front Cell Infect Microbiol. 2018;8(October):345.
89. Thiberville SD, Moyen N, Dupuis-Maguiraga L, Nougairede A, Gould EA, Roques P, et al. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. Antiviral Res [Internet]. 2013;99(3):345–70.
90. Huits R, De Kort J, Van Den Berg R, Chong L, Tsoumanis A, Eggermont K, et al. Chikungunya virus infection in Aruba: Diagnosis, clinical features and predictors of post-chikungunya chronic polyarthralgia. PLoS One. 2018;13(4):1–21.
91. Silva LA, Dermody TS. Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. J Clin Invest [Internet]. 2017;127(3):737–49.
92. Weaver SC, Scott TW, Lorenz LH, Repik PM. Detection of eastern equine encephalomyelitis virus deposition in Culiseta melanura following ingestion of radiolabeled virus in blood meals. Am J Trop Med Hyg. 1991;
93. Wong HV, Chan YF, Sam IC, Sulaiman WYW, Vythilingam I. Chikungunya virus infection of Aedes mosquitoes. Methods Mol Biol. 2016;1426:119–28.
94. Dubrulle M, Mousson L, Moutailier S, Vazeille M, Failloux AB. Chikungunya virus and Aedes mosquitoes: Saliva is infectious as soon as two days after oral infection. PLoS One. 2009;4(6).
95. Li YG, Siripanyaphinyo U, Tumkosit U, Noranate N, A-Nuegoonpipat A, Tao R, et al. Chikungunya virus induces a more moderate cytopathic effect in mosquito cells than in mammalian cells. Intervirology. 2012;56(1):6–12.
96. Martin E, Moutailler S, Madec Y, Failloux AB. Differential responses of the mosquito aedes albopictus from the indian ocean region to two chikungunya isolates. BMC Ecol. 2010;10.
97. Vogels CBF, Göertz GP, Pijlman GP, Koenraadt CJM. Vector competence of European mosquitoes for West Nile virus. Emerg Microbes Infect [Internet]. 2017;6(11):1–13.
98. Lim EXY, Lee WS, Madzokere ET, Herrero LJ. Mosquitoes as suitable vectors for alphaviruses. Viruses. 2018;10(2):1–17.
99. Mulwa F, Lutomiah J, Chepkorir E, Okello S, Eyase F, Tigoi C, et al. Vector competence of Aedes bromeliae and Aedes vitattus mosquito populations from Kenya for chikungunya virus. PLoS Negl Trop Dis. 2018;12(10):e0006746.
100. Harrington LC, Fleisher A, Ruiz-Moreno D, Vermeylen F, Wa C V., Poulson RL, et al.

Heterogeneous Feeding Patterns of the Dengue Vector, Aedes aegypti, on Individual Human Hosts in Rural Thailand. PLoS Negl Trop Dis. 2014;8(8).
101. Savage HM, Niebylski ML, Smith GC, Mitchell CJ, Craig GB. Host-feeding patterns of Aedes albopictus (Diptera: Culicidae) at a temperate North American site. J Med Entomol. 1993;30(1):27–34.
102. Vega-Rúa A, Zouache K, Girod R, Failloux A-B, Lourenço-de-Oliveira R. High level of vector competence of Aedes aegypti and Aedes albopictus from ten American countries as a crucial factor in the spread of Chikungunya virus. J Virol [Internet]. 2014;88(11):6294–306.
103. Benedict MQ, Levine RS, Hawley WA, Lounibos LP. Spread of the tiger: Global risk of invasion by the mosquito Aedes albopictus. Vector-Borne Zoonotic Dis. 2007;7(1):76–85.
104. Sota T, Mogi M. Interspecific variation in desiccation survival time of Aedes (Stegomyia) mosquito eggs is correlated with habitat and egg size. Oecologia. 1992;90(3):353–8.
105. Vazeille M, Mousson L, Martin E, Failloux AB. Orally co-infected Aedes albopictus from La Reunion Island, Indian Ocean, can deliver both dengue and chikungunya infectious viral particles in their saliva. PLoS Negl Trop Dis. 2010;4(6):1–5.
106. Rückert C, Weger-Lucarelli J, Garcia-Luna SM, Young MC, Byas AD, Murrieta RA, et al. Impact of simultaneous exposure to arboviruses on infection and transmission by Aedes aegypti mosquitoes. Nat Commun. 2017;8(May):1–9.
107. Carrillo-Hernández MY, Ruiz-Saenz J, Villamizar LJ, Gómez-Rangel SY, Martínez-Gutierrez M. Co-circulation and simultaneous co-infection of dengue, chikungunya, and zika viruses in patients with febrile syndrome at the Colombian-Venezuelan border. BMC Infect Dis. 2018;18(1):1–12.
108. Tsetsarkin K a., McGee CE, Volk SM, Vanlandingham DL, Weaver SC, Higgs S. Epistatic roles of E2 glycoprotein mutations in adaption of Chikungunya virus to Aedes albopictus and Ae. Aegypti mosquitoes. PLoS One. 2009;4(8).
109. Tsetsarkin K a., Weaver SC. Sequential adaptive mutations enhance efficient vector switching by chikungunya virus and its epidemic emergence. PLoS Pathog. 2011;7(12).
110. Powers AM. Vaccine and therapeutic options to control chikungunya virus. Clin Microbiol Rev. 2018;31(1):1–29.
111. Levitt NH, Ramsburg HH, Hasty SE, Repik PM, Cole FE, Lupton HW. Development of an attenuated strain of chikungunya virus for use in vaccine production. Vaccine. 1986;4(3):157–62.
112. McClain DJ, Pittman PR, Ramsburg HH, Nelson GO, Rossi CA, Mangiafico JA, et al. Immunologic Interference from Sequential Administration of Live Attenuated Alphavirus Vaccines. J Infect Dis. 1998;177(3):634–41.
113. Edelman R, Tacket CO, Wasserman SS, Bodison SA, Perry JG, Mangiafico JA. Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218. Am J Trop Med Hyg. 2000;62(6):681–5.
114. Gorchakov R, Wang E, Leal G, Forrester NL, Plante K, Rossi SL, et al. Attenuation of Chikungunya Virus Vaccine Strain 181/Clone 25 Is Determined by Two Amino Acid Substitutions in the E2 Envelope Glycoprotein. J Virol. 2012;86(11):6084–96.
115. Hallengard D, Kakoulidou M, Lulla A, Kummerer BM, Johansson DX, Mutso M, et al. Novel Attenuated Chikungunya Vaccine Candidates Elicit Protective Immunity in C57BL/6 mice. J Virol. 2014;88(5):2858–66.
116. Goo L, Dowd KA, Lin TY, Mascola JR, Graham BS, Ledgerwood JE, et al. A virus-like

particle vaccine elicits broad neutralizing antibody responses in humans to all chikungunya virus genotypes. J Infect Dis. 2016;214(10):1487–91.
117. Muthumani K, Lankaraman KM, Laddy DJ, Sundaram SG, Chung CW, Sako E, et al. Immunogenicity of novel consensus-based DNA vaccines against Chikungunya virus. Vaccine. 2008;26(40):5128–34.
118. Khan M, Dhanwani R, Rao PVL, Parida M. Subunit vaccine formulations based on recombinant envelope proteins of Chikungunya virus elicit balanced Th1/Th2 response and virus-neutralizing antibodies in mice. Virus Res [Internet]. 2012;167(2):236–46.
119. Erasmus JH, Auguste AJ, Kaelber JT, Luo H, Rossi SL, Fenton K, et al. A chikungunya fever vaccine utilizing an insect-specific virus platform. Nat Med. 2017;23(2):192–9.
120. Plante K, Wang E, Partidos CD, Weger J, Gorchakov R, Tsetsarkin K, et al. Novel chikungunya vaccine candidate with an ires-based attenuation and host range alteration mechanism. PLoS Pathog. 2011;7(7).
121. Roy CJ, Adams AP, Wang E, Plante K, Gorchakov R, Seymour RL, et al. Chikungunya vaccine candidate is highly attenuated and protects nonhuman primates against telemetrically monitored disease following a single dose. J Infect Dis. 2014;209(12):1891–9.
122. R.M. L, S.L. H, C.J. R, H. V-O, N.A. B, J.H. E, et al. Chikungunya virus strains show lineage-specific variations in virulence and cross-protective ability in murine and nonhuman primate models. MBio [Internet]. 2018;9(2):1–13.
123. Reisinger EC, Tschismarov R, Beubler E, Wiedermann U, Firbas C, Loebermann M, et al. Immunogenicity, safety, and tolerability of the measles-vectored chikungunya virus vaccine MV-CHIK: a double-blind, randomised, placebo-controlled and active-controlled phase 2 trial. Lancet [Internet]. 2018;392(10165):2718–27.
124. Lani R, Hassandarvish P, Chiam CW, Moghaddam E, Chu JJH, Rausalu K, et al. Antiviral activity of silymarin against chikungunya virus. Sci Rep [Internet]. 2015;5:1–10.
125. Wintachai P, Thuaud F, Basmadjian C, Roytrakul S, Ubol S, Désaubry L, et al. Assessment of flavaglines as potential chikungunya virus entry inhibitors. Microbiol Immunol. 2015;59(3):129–41.
126. Ho YJ, Wang YM, Lu JW, Wu TY, Lin LI, Kuo SC, et al. Suramin inhibits chikungunya virus entry and transmission. PLoS One. 2015;10(7).
127. Mounce BC, Cesaro T, Carrau L, Vallet T, Vignuzzi M. Curcumin inhibits Zika and chikungunya virus infection by inhibiting cell binding. Antiviral Res [Internet]. 2017;142:148–57.
128. Briolant S, Garin D, Scaramozzino N, Jouan A, Crance JM. In vitro inhibition of Chikungunya and Semliki Forest viruses replication by antiviral compounds: Synergistic effect of interferon-α and ribavirin combination. Antiviral Res. 2004;61(2):111–7.
129. Ravichandran R, Manian M. Ribavirin therapy for Chikungunya arthritis. J Infect Dev Ctries. 2008;2(2):140–2.
130. Gallegos KM, Drusano GL, D′argenio DZ, Brown AN. Chikungunya Virus: In Vitro Response to Combination Therapy with Ribavirin and Interferon Alfa 2a. J Infect Dis. 2016;214(8):1192–7.
131. Smith SA, Silva LA, Fox JM, Flyak AI, Kose N, Sapparapu G, et al. Isolation and characterization of broad and ultrapotent human monoclonal antibodies with therapeutic activity against chikungunya virus. Cell Host Microbe. 2015;18(1):86–95.
132. Broeckel R, Fox JM, Haese N, Kreklywich CN, Sukulpovi-Petty S, Legasse A, et al.

Therapeutic administration of a recombinant human monoclonal antibody reduces the severity of chikungunya virus disease in rhesus macaques. PLoS Negl Trop Dis. 2017;11(6):1–25.
133. Organisation WH. Global Strategy for Dengue Prevention and Control, 2012–2020 WHO, Geneva 2012. 2012.
134. Lupi E, Hatz C, Schlagenhauf P. The efficacy of repellents against Aedes, Anopheles, Culex and Ixodes spp. - A literature review. Travel Med Infect Dis [Internet]. 2013;11(6):374–411.
135. Wilson AL, Dhiman RC, Kitron U, Scott TW, van den Berg H, Lindsay SW. Benefit of Insecticide-Treated Nets, Curtains and Screening on Vector Borne Diseases, Excluding Malaria: A Systematic Review and Meta-analysis. PLoS Negl Trop Dis. 2014;8(10).
136. Vanlerberghe V, Toledo ME, Rodríguez M, Gomez D, Baly A, Benitez JR, et al. Community involvement in dengue vector control: Cluster randomised trial. BMJ. 2009;338(7709):1477–9.
137. George L, Lenhart A, Toledo J, Lazaro A, Han WW, Velayudhan R, et al. Community-Effectiveness of Temephos for Dengue Vector Control: A Systematic Literature Review. PLoS Negl Trop Dis. 2015;9(9).
138. Maoz D, Ward T, Samuel M, Müller P, Runge-Ranzinger S, Toledo J, et al. Community effectiveness of pyriproxyfen as a dengue vector control method: A systematic review. PLoS Negl Trop Dis. 2017;11(7).
139. Samuel M, Maoz D, Manrique P, Ward T, Runge-Ranzinger S, Toledo J, et al. Community effectiveness of indoor spraying as a dengue vector control method: A systematic review. PLoS Negl Trop Dis. 2017;11(8):1–13.
140. Barrera R, Acevedo V, Felix GE, Hemme RR, Vazquez J, Munoz JL, et al. Impact of autocidal gravid ovitraps on chikungunya virus incidence in Aedes Aegypti (Diptera: Culicidae) in Areas With and Without Traps. J Med Entomol. 2018;54(2):387–95.
141. Lorenzi OD, Major C, Acevedo V, Perez-Padilla J, Rivera A, Biggerstaff BJ, et al. Reduced incidence of Chikungunya virus infection in communities with ongoing aedes aegypti mosquito trap intervention studies — Salinas and Guayama, Puerto Rico, November 2015–february 2016. Morb Mortal Wkly Rep. 2016;65(18):479–80.
142. Ranson H, Burhani J, Lumjuan N, Black, IV WC. Insecticide resistance in dengue vectors. TropIKA [Internet]. 2010;0–0.
143. Helinski MEH, Parker AG, Knols BGJ. Radiation biology of mosquitoes. Malar J. 2009;8(SUPPL. 2):1–13.
144. Chambers EW, Hapairai L, Peel BA, Bossin H, Dobson SL. Male mating competitiveness of a Wolbachia-introgressed Aedes polynesiensis strain under semi-field conditions. PLoS Negl Trop Dis. 2011;5(8):1–6.
145. Zheng X, Zhang D, Li Y, Yang C, Wu Y, Liang X, et al. Incompatible and sterile insect techniques combined eliminate mosquitoes. Nature. 2019;572(7767):56–61.
146. Carvalho DO, McKemey AR, Garziera L, Lacroix R, Donnelly CA, Alphey L, et al. Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes. PLoS Negl Trop Dis. 2015;9(7):1–15.
147. Evans BR, Kotsakiozi P, Costa-da-Silva AL, Ioshino RS, Garziera L, Pedrosa MC, et al. Transgenic Aedes aegypti Mosquitoes Transfer Genes into a Natural Population. Sci Rep [Internet]. 2019;9(1):1–6.
148. Higashi N, Matsumoto A, Tabata K, Nagatomo Y. Electron microscope study of

development of Chikungunya virus in green monkey kidney stable (VERO) cells. Virology. 1967;33(1):55–69.
149. Voss JE, Vaney MC, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, et al. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature. 2010;468(7324):709–12.
150. Simizu B, Yamamoto K, Hashimoto K, Ogata T. Structural proteins of Chikungunya virus. J Virol. 1984;
151. Sun S, Xiang Y, Akahata W, Holdaway H, Pal P, Zhang X, et al. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. Elife. 2013;2013(2):1–27.
152. Yap ML, Klose T, Urakami A, Hasan SS, Akahata W, Rossmann MG. Structural studies of Chikungunya virus maturation. Proc Natl Acad Sci U S A. 2017;114(52):13703–7.
153. Khan AH, Morita K, del Carmen Parquet M, Hasebe F, Mathenge EGM, Igarashi A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. J Gen Virol. 2002;83(12):3075–84.
154. Strauss EG, Rice CM, Strauss JH. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. Proc Natl Acad Sci U S A. 1983;80(17):5271–5.
155. Melton J V., Ewart GD, Weir RC, Board PG, Lee E, Gage PW. Alphavirus 6K proteins form ion channels. J Biol Chem. 2002;277(49):46923–31.
156. Bakar FA, Ng LFP. Nonstructural proteins of alphavirus—potential targets for drug development. Viruses. 2018;10(2):1–15.
157. Moller-Tank S, Kondratowicz AS, Davey RA, Rennert PD, Maury W. Role of the Phosphatidylserine Receptor TIM-1 in Enveloped-Virus Entry. J Virol. 2013;87(15):8327–41.
158. Silva LA, Khomandiak S, Ashbrook AW, Weller R, Heise MT, Morrison TE, et al. A Single-Amino-Acid Polymorphism in Chikungunya Virus E2 Glycoprotein Influences Glycosaminoglycan Utilization. J Virol. 2014;88(5):2385–97.
159. Zhang R, Kim AS, Fox JM, Nair S, Basore K, Klimstra WB, et al. Mxra8 is a receptor for multiple arthritogenic alphaviruses. Nature. 2018;557(7706):570–4.
160. Basore K, Kim AS, Nelson CA, Zhang R, Smith BK, Uranga C, et al. Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. Cell [Internet]. 2019;177(7):1725-1737.e16.
161. Bernard E, Solignat M, Gay B, Chazal N, Higgs S, Devaux C, et al. Endocytosis of chikungunya virus into mammalian cells: Role of clathrin and early endosomal compartments. PLoS One. 2010;5(7).
162. Lee RCH, Hapuarachchi HC, Chen KC, Hussain KM, Chen H, Low SL, et al. Mosquito Cellular Factors and Functions in Mediating the Infectious entry of Chikungunya Virus. PLoS Negl Trop Dis. 2013;7(2).
163. Ruehle MD, Zhang H, Sheridan RM, Mitra S, Chen Y, Gonzalez RL, et al. A dynamic RNA loop in an IRES affects multiple steps of elongation factor-mediated translation initiation: (A) Schematic of the IGR IRES initiation factor-independent translation initiation mechanism. The IGR IRESs occupy the same binding sites as tRNAs i. Elife. 2015;4:1–24.
164. van Duijl-Richter MKS, Hoornweg TE, Rodenhuis-Zybert IA, Smit JM. Early events in chikungunya virus infection—from virus cell binding to membrane fusion. Viruses.

2015;7(7):3647–74.
165. Hoornweg TE, van Duijl-Richter MKS, Ayala Nuñez N V., Albulescu IC, van Hemert MJ, Smit JM. Dynamics of chikungunya virus cell entry unraveled by single virus tracking in living cells. J Virol [Internet]. 2016;90(9):JVI.03184-15.
166. Wengler G, Wengler G. Identification of a transfer of viral core protein to cellular ribosomes during the early stages of alphavirus infection. Virology. 1984;134(2):435–42.
167. Singh I, Helenius A. Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. J Virol [Internet]. 1992;66(12):7049–58.
168. Sokoloski KJ, Nease LM, May NA, Gebhart NN, Jones CE, Morrison TE, et al. Identification of Interactions between Sindbis Virus Capsid Protein and Cytoplasmic vRNA as Novel Virulence Determinants. PLoS Pathog. 2017;13(6):1–29.
169. Jones JE, Long KM, Whitmore AC, Sanders W, Thurlow LR, Brown JA, et al. Disruption of the opal stop codon attenuates chikungunya virus-induced arthritis and pathology. MBio. 2017;8(6):1–16.
170. Garcia-Moreno M, Sanz MA, Carrasco L. A Viral mRNA Motif at the 3′-Untranslated Region that Confers Translatability in a Cell-Specific Manner. Implications for Virus Evolution. Sci Rep [Internet]. 2016;6(July 2015):1–17.
171. Ahola T, Lampio A, Auvinen P, Kääriäinen L. Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity. EMBO J. 1999;18(11):3164–72.
172. Vasiljeva L, Valmu L, Kääriäinen L, Merits A. Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2. J Biol Chem. 2001;276(33):30786–93.
173. de Groot RJ, Hardy WR, Shirako Y, Strauss JH. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. EMBO J. 1990;9(8):2631–8.
174. Abraham R, Hauer D, McPherson RL, Utt A, Kirby IT, Cohen MS, et al. ADP-ribosyl-binding and hydrolase activities of the alphavirus nsP3 macrodomain are critical for initiation of virus replication. Proc Natl Acad Sci U S A. 2018;115(44):E10457–66.
175. Neuvonen M, Kazlauskas A, Martikainen M, Hinkkanen A, Ahola T, Saksela K. SH3 domain-mediated recruitment of host cell amphiphysins by alphavirus nsp3 promotes viral RNA replication. PLoS Pathog. 2011;7(11).
176. Kallio K, Hellström K, Jokitalo E. RNA Replication and Membrane Modification Require the Same Functions of Alphavirus Nonstructural Proteins. 2016;90(3):1687–92.
177. Meshram CD, Agback P, Shiliaev N, Urakova N, Mobley JA, Agback T, et al. Multiple Host Factors Interact with the Hypervariable Domain of Chikungunya Virus nsP3 and Determine Viral Replication in Cell-Specific Mode. J Virol. 2018;92(16):1–24.
178. Lee CY, Kam YW, Fric J, Malleret B, Koh EGL, Prakash C, et al. Chikungunya virus neutralization antigens and direct cell-to-cell transmission are revealed by human antibody-escape mutants. PLoS Pathog. 2011;7(12).
179. Jose J, Taylor AB, Kuhn RJ. Spatial and temporal analysis of alphavirus replication and assembly in mammalian and mosquito cells. MBio. 2017;8(1):1–16.
180. Rodriguez AK, Muñoz AL, Segura NA, Rangel HR, Bello F. Molecular characteristics and replication mechanism of dengue, zika and chikungunya arboviruses, and their treatments with natural extracts from plants: An updated review. EXCLI J. 2019;18:988–1006.

181. Rupp JC, Jundt N, Hardy RW. Requirement for the Amino-Terminal Domain of Sindbis Virus nsP4 during Virus Infection. J Virol. 2011;85(7):3449–60.
182. De I, Fata-Hartley C, Sawicki SG, Sawicki DL. Functional Analysis of nsP3 Phosphoprotein Mutants of Sindbis Virus. J Virol. 2003;77(24):13106–16.
183. Mayuri, Geders TW, Smith JL, Kuhn RJ. Role for Conserved Residues of Sindbis Virus Nonstructural Protein 2 Methyltransferase-Like Domain in Regulation of Minus-Strand Synthesis and Development of Cytopathic Infection. J Virol. 2008;82(15):7284–97.
184. Li ML, Stollar V. Identification of the amino acid sequence in Sindbis virus nsP4 that binds to the promoter for the synthesis of the subgenomic RNA. Proc Natl Acad Sci U S A. 2004;101(25):9429–34.
185. Thal MA, Wasik BR, Posto J, Hardy RW. Template requirements for recognition and copying by Sindbis virus RNA-dependent RNA polymerase. Virology. 2007;358(1):221–32.
186. Kääriäinen L, Takkinen K, Keränen S, Söderlund H. Replication of the genome of alphaviruses. J Cell Sci Suppl. 1987;
187. Shirako Y, Strauss JH. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. J Virol [Internet]. 1994;68(3):1874–85.
188. Shirako Y, Strauss JH. Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. Virology. 1990;177(1):54–64.
189. Lulla A, Lulla V, Tints K, Ahola T, Merits A. Molecular Determinants of Substrate Specificity for Semliki Forest Virus Nonstructural Protease. J Virol. 2006;80(11):5413–22.
190. Vasiljeva L, Merits A, Golubtsov A, Sizemskaja V, Kääriäinen L, Ahola T. Regulation of the Sequential Processing of Semliki Forest Virus Replicase Polyprotein. J Biol Chem. 2003;278(43):41636–45.
191. Sawicki DL, Sawicki SG. Short-lived minus-strand polymerase for Semliki Forest virus. J Virol [Internet]. 1980;34(1):108–18.
192. Shin G, Yost SA, Miller MT, Elrod EJ, Grakoui A, Marcotrigiano J. Structural and functional insights into alphavirus polyprotein processing and pathogenesis. Proc Natl Acad Sci U S A. 2012;109(41):16534–9.
193. Li M-L, Stollar V. Distinct Sites on the Sindbis Virus RNA-Dependent RNA Polymerase for Binding to the Promoters for the Synthesis of Genomic and Subgenomic RNA. J Virol. 2007;81(8):4371–3.
194. Sawicki DL, Sawicki SG. A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis. J Virol. 1993;67(6):3605–10.
195. Suopanki J, Sawicki DL, Sawicki SG, Kääriäinen L. Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. J Gen Virol. 1998;79(2):309–19.
196. LaStarza MW, Lemm JA, Rice CM. Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. J Virol [Internet]. 1994;68(9):5781–91.
197. Gao Y, Goonawardane N, Ward J, Tuplin A, Harris M. Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during Chikungunya virus genome replication and transcription. PLoS Pathog. 2019;15(1):1–28.
198. Salonen A, Vasiljeva L, Merits A, Magden J, Jokitalo E, Kaariainen L. Properly Folded Nonstructural Polyprotein Directs the Semliki Forest Virus Replication Complex to the

Endosomal Compartment. J Virol. 2003;77(3):1691–702.
199. Gorchakov R, Frolova E, Sawicki S, Atasheva S, Sawicki D, Frolov I. A New Role for ns Polyprotein Cleavage in Sindbis Virus Replication. J Virol. 2008;82(13):6218–31.
200. Scholte FEM, Tas A, Albulescu IC, Žusinaite E, Merits A, Snijder EJ, et al. Stress Granule Components G3BP1 and G3BP2 Play a Proviral Role Early in Chikungunya Virus Replication. J Virol. 2015;89(8):4457–69.
201. Sanz MA, García-Moreno M, Carrasco L. Inhibition of host protein synthesis by Sindbis virus: Correlation with viral RNA replication and release of nuclear proteins to the cytoplasm. Cell Microbiol. 2015;17(4):520–41.
202. Castelló A, Sanz MA, Molina S, Carrasco L. Translation of Sindbis virus 26S mRNA does not require intact eukariotic initiation factor 4G. J Mol Biol [Internet]. 2006;355(5):942–56.
203. Ventoso I, Sanz MA, Molina S, Berlanga JJ, Carrasco L, Esteban M. Translational resistance of late alphavirus mRNA to eIF2α phosphorylation: A strategy to overcome the antiviral effect of protein kinase PKR. Genes Dev. 2006;20(1):87–100.
204. Garoff H, Huylebroeck D, Robinson A, Tillman U, Liljestrom P. The signal sequence of the p62 protein of Semliki Forest virus is involved in initiation but not in completing chain translocation. J Cell Biol. 1990;111(3):867–76.
205. Anthony RP, Paredes AM, Brown DT. Disulfide bonds are essential for the stability of the sindbis virus envelope. Virology. 1992;190(1):330–6.
206. De Curtis I, Simons K. Dissection of Semliki Forest virus glycoprotein delivery from the trans-Golgi network to the cell surface in permeabilized BHK cells. Proc Natl Acad Sci U S A. 1988;85(21):8052–6.
207. Perera R, Owen KE, Tellinghuisen TL, Gorbalenya AE, Kuhn RJ. Alphavirus Nucleocapsid Protein Contains a Putative Coiled Coil -Helix Important for Core Assembly. J Virol. 2001;75(1):1–10.
208. Kim DY, Firth AE, Atasheva S, Frolova EI, Frolov I. Conservation of a Packaging Signal and the Viral Genome RNA Packaging Mechanism in Alphavirus Evolution. J Virol. 2011;85(16):8022–36.
209. Liu N, Brown DT. Phosphorylation and Dephosphorylation Events Play Critical Roles in Sindbis Virus Maturation. Virology. 1993;196(2):703–11.
210. Sharma R, Kesari P, Kumar P, Tomar S. Structure-function insights into chikungunya virus capsid protein: Small molecules targeting capsid hydrophobic pocket. Virology [Internet]. 2018;515(December 2017):223–34.
211. Ferreira D, Hernandez R, Horton M, Brown DT. Morphological variants of sindbis virus produced by a mutation in the capsid protein. Virology. 2003;307(1):54–66.
212. Kononchik JP, Nelson S, Hernandez R, Brown DT. Helical virus particles formed from morphological subunits of a membrane containing icosahedral virus. Virology [Internet]. 2009;385(2):285–93.
213. Snyder JE, Berrios CJ, Edwards TJ, Jose J, Perera R, Kuhn RJ. Probing the Early Temporal and Spatial Interaction of the Sindbis Virus Capsid and E2 Proteins with Reverse Genetics. J Virol. 2012;86(22):12372–83.
214. Gaedigk-Nitschko K, Schlesinger MJ. Site-directed mutations in sindbis virus E2 glycoprotein's cytoplasmic domain and the 6K protein lead to similar defects in virus assembly and budding. Virology. 1991;183(1):206–14.
215. Snyder JE, Kulcsar KA, Schultz KLW, Riley CP, Neary JT, Marr S, et al. Functional

Characterization of the Alphavirus TF Protein. J Virol. 2013;87(15):8511–23.
216. Soonsawad P, Xing L, Milla E, Espinoza JM, Kawano M, Marko M, et al. Structural Evidence of Glycoprotein Assembly in Cellular Membrane Compartments prior to Alphavirus Budding. J Virol. 2010;84(21):11145–51.
217. Mancini EJ, Clarke M, Gowen BE, Rutten T, Fuller SD. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. Mol Cell. 2000;5(2):255–66.
218. Mukhopadhyay S, Zhang W, Gabler S, Chipman PR, Strauss EG, Strauss JH, et al. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. Structure. 2006;14(1):63–73.
219. Zhang R, Hryc CF, Cong Y, Liu X, Jakana J, Gorchakov R, et al. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. EMBO J [Internet]. 2011;30(18):3854–63.
220. Melancon P, Garoff H. Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. J Virol [Internet]. 1987;61(5):1301–9.
221. Garoff H, Frischauf AM, Simons K, Lehrach H, Delius H. The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. Proc Natl Acad Sci U S A. 1980;77(11 I):6376–80.
222. Choi HK, Tong L, Minor W, Dumas P, Boege U, Rossmann MG, et al. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature. 1991;354(6348):37–43.
223. Hong EM, Perera R, Kuhn RJ. Alphavirus Capsid Protein Helix I Controls a Checkpoint in Nucleocapsid Core Assembly. J Virol. 2006;80(18):8848–55.
224. Lulla V, Kim DY, Frolova EI, Frolov I. The Amino-Terminal Domain of Alphavirus Capsid Protein Is Dispensable for Viral Particle Assembly but Regulates RNA Encapsidation through Cooperative Functions of Its Subdomains. J Virol. 2013;87(22):12003–19.
225. Hahn CS, Strauss JH. Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. J Virol [Internet]. 1990;64(6):3069–73.
226. Aggarwal M, Sharma R, Kumar P, Parida M, Tomar S. Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. Sci Rep. 2015;5(April):1–12.
227. Thomas S, Rai J, John L, Schaefer S, Pützer BM, Herchenröder O. Chikungunya virus capsid protein contains nuclear import and export signals. Virol J. 2013;10:1–13.
228. von Bonsdorff CH, Harrison SC. Hexagonal glycoprotein arrays from Sindbis virus membranes. J Virol [Internet]. 1978;28(2):578–83.
229. Wengler G, Wengler G, Rey FA. The isolation of the ectodomain of the alphavirus E1 protein as a soluble hemagglutinin and its crystallization. Virology. 1999;257(2):472–82.
230. Julien Lescar, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Rey FA. The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell. 2001;105(1):137–48.
231. Smith TJ, Cheng RH, Olson NH, Peterson P, Chase E, Kuhn RJ, et al. Putative receptor binding sites on alphaviruses as visualized by cryoelectron microscopy. Proc Natl Acad Sci U S A. 1995;92(23):10648–52.
232. Schmidt MF, Bracha M, Schlesinger MJ. Evidence for covalent attachment of fatty acids to Sindbis virus glycoproteins. Proc Natl Acad Sci U S A. 1979;76(4):1687–91.

233. Wahlberg JM, Garoff H. Membrane fusion process of Semliki Forest virus I: Low pH-induced rearrangement in spike protein quaternary structure precedes virus penetration into cells. J Cell Biol. 1992;116(2):339–48.
234. Wahlberg JM, Bron R, Wilschut J, Garoff H. Membrane fusion of Semliki Forest virus involves homotrimers of the fusion protein. J Virol [Internet]. 1992;66(12):7309–18.
235. Lobigs M, Zhao HX, Garoff H. Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. J Virol [Internet]. 1990;64(9):4346–55.
236. Liljeström P, Garoff H. Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. J Virol [Internet]. 1991;65(1):147–54.
237. Gaedigk-Nitschko K, Ding M, Levy MA, Schlesinger MJ. Site-directed mutations in the sindbis virus 6K protein reveal sites for fatty acylation and the underacylated protein affects virus release and virion structure. Virology. 1990;175(1):282–91.
238. Liljeström P, Lusa S, Huylebroeck D, Garoff H. In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. J Virol [Internet]. 1991;65(8):4107–13.
239. McInerney GM, Smit JM, Liljeström P, Wilschut J. Semliki Forest virus produced in the absence of the 6K protein has an altered spike structure as revealed by decreased membrane fusion capacity. Virology. 2004;325(2):200–6.
240. Kendra JA, Advani VM, Chen B, Briggs JW, Zhu J, Bress HJ, et al. Functional and structural characterization of the chikungunya virus translational recoding signals. J Biol Chem. 2018;293(45):17536–45.
241. Wang YF, Sawicki SG, Sawicki DL. Sindbis virus nsP1 functions in negative-strand RNA synthesis. J Virol [Internet]. 1991;65(2):985–8.
242. Cross RK. Identification of a unique guanine-7-methyltransferase in Semliki forest virus (SFV) infected cell extracts. Virology. 1983;130(2):452–63.
243. Laakkonen P, Hyvönen M, Peränen J, Kääriäinen L. Expression of Semliki forest virus nsP1-specific methyltransferase in insect cells and in Escherichia coli. J Virol. 1994;68(11):7418–25.
244. Ahola T, Kääriäinen L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. Proc Natl Acad Sci U S A [Internet]. 1995;92(2):507–11.
245. Ahola T, Laakkonen P, Vihinen H, Kääriäinen L. Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. J Virol [Internet]. 1997;71(1):392–7.
246. Mizumoto K, Kaziro Y, Lipmann F. Reaction mechanism of mRNA guanylyltransferase from rat liver: isolation and characterization of a guanylyl-enzyme intermediate. Proc Natl Acad Sci U S A. 1982;79(6 I):1693–7.
247. Vasiljeva L, Merits A, Auvinen P, Kääriäinen L. Identification of a novel function of the Alphavirus capping apparatus. J Biol Chem. 2000;275(23):17281–7.
248. Lampio A, Kilpeläinen I, Pesonen S, Karhi K, Auvinen P, Somerharju P, et al. Membrane binding mechanism of an RNA virus-capping enzyme. J Biol Chem. 2000;275(48):37853–9.
249. Spuul P, Salonen A, Merits A, Jokitalo E, Kaariainen L, Ahola T. Role of the Amphipathic Peptide of Semliki Forest Virus Replicase Protein nsP1 in Membrane

Association and Virus Replication. J Virol. 2007;81(2):872–83.
250. Kallio K, Hellström K, Jokitalo E, Ahola T. RNA Replication and Membrane Modification Require the Same Functions of Alphavirus Nonstructural Proteins. J Virol. 2016;90(3):1687–92.
251. Rupp JC, Sokoloski KJ, Gebhart NN, Hardy RW. Alphavirus RNA synthesis and nonstructural protein functions. J Gen Virol [Internet]. 2015;96(9):2483–500.
252. Laakkonen P, Ahola T, Kääriäinen L. The effects of palmitoylation on membrane association of Semliki Forest virus RNA capping enzyme. J Biol Chem. 1996;271(45):28567–71.
253. Shirako Y, Strauss EG, Strauss JH. Suppressor mutations that allow Sindbis virus RNA polymerase to function with nonaromatic amino acids at the N-terminus: Evidence for interaction between nsP1 and nsP4 in minus-strand RNA synthesis. Virology. 2000;276(1):148–60.
254. Fata CL, Sawicki SG, Sawicki DL. Modification of Asn374 of nsP1 Suppresses a Sindbis Virus nsP4 Minus-Strand Polymerase Mutant. J Virol. 2002;76(17):8641–9.
255. Birdwell CR, Strauss EG, Strauss JH. Replication of Sindbis virus. III. An electron microscopic study of virus maturation using the surface replica technique. Virology. 1973;56(2):429–38.
256. Martinez MG, Snapp E-L, Perumal GS, Macaluso FP, Kielian M. Imaging the Alphavirus Exit Pathway. J Virol. 2014;88(12):6922–33.
257. Laakkonen P, Auvinen P, Kujala P, Kääriäinen L. Alphavirus replicase protein NSP1 induces filopodia and rearrangement of actin filaments. J Virol [Internet]. 1998;72(12):10265–9.
258. Ahola T, Kujala P, Tuittila M, Blom T, Laakkonen P, Hinkkanen A, et al. Effects of Palmitoylation of Replicase Protein nsP1 on Alphavirus Infection. J Virol. 2000;74(15):6725–33.
259. Žusinaite E, Tints K, Kiiver K, Spuul P, Karo-Astover L, Merits A, et al. Mutations at the palmitoylation site of non-structural protein nsP1 of Semliki Forest virus attenuate virus replication and cause accumulation of compensatory mutations. J Gen Virol. 2007;88(7):1977–85.
260. Karo-Astover L, Šarova O, Merits A, Žusinaite E. The infection of mammalian and insect cells with SFV bearing nsP1 palmitoylation mutations. Virus Res. 2010;153(2):277–87.
261. Gomez De Cedrón M, Ehsani N, Mikkola ML, García JA, Kääriäinen L. RNA helicase activity of Semliki Forest virus replicase protein NSP2. FEBS Lett. 1999;448(1):19–22.
262. Karpe YA, Aher PP, Lole KS. NTPase and 5′-RNA triphosphatase activities of chikungunya virus nsP2 protein. PLoS One. 2011;6(7).
263. Gorbalenya AE, Blinov VM, Donchenko AP, Koonin E V. An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. J Mol Evol. 1989;28(3):256–68.
264. Russo AT, White MA, Watowich SJ. The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease. Structure. 2006;14(9):1449–58.
265. Ramakrishnan C, Kutumbarao NHV, Suhitha S, Velmurugan D. Structure–function relationship of Chikungunya nsP2 protease: A comparative study with papain. Chem Biol Drug Des. 2017;89(5):772–82.
266. Hardy WR, Strauss JH. Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in

trans. J Virol [Internet]. 1989;63(11):4653–64.
267. Strauss EG, De Groot RJ, Levinson R, Strauss JH. Identification of the active site residues in the nsP2 proteinase of sindbis virus. Virology. 1992;191(2):932–40.
268. Merits A, Vasiljeva L, Ahola T, Kääriäinen L, Auvinen P. Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein by nsP2 protease. J Gen Virol. 2001;82(4):765–73.
269. Frolova EI, Fayzulin RZ, Cook SH, Griffin DE, Rice CM, Frolov I. Roles of Nonstructural Protein nsP2 and Alpha/Beta Interferons in Determining the Outcome of Sindbis Virus Infection. J Virol. 2002;76(22):11254–64.
270. Tamm K, Merits A, Sarand I. Mutations in the nuclear localization signal of nsP2 influencing RNA synthesis, protein expression and cytotoxicity of Semliki Forest virus. J Gen Virol. 2008;89(3):676–86.
271. Bourai M, Lucas-Hourani M, Gad HH, Drosten C, Jacob Y, Tafforeau L, et al. Mapping of Chikungunya Virus Interactions with Host Proteins Identified nsP2 as a Highly Connected Viral Component. J Virol. 2012;86(6):3121–34.
272. Akhrymuk I, Kulemzin S V., Frolova EI. Evasion of the Innate Immune Response: the Old World Alphavirus nsP2 Protein Induces Rapid Degradation of Rpb1, a Catalytic Subunit of RNA Polymerase II. J Virol. 2012;86(13):7180–91.
273. Peränen J, Rikkonen M, Liljeström P, Kääriäinen L. Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. J Virol [Internet]. 1990;64(5):1888–96.
274. Cells V, Gorchakov R, Frolova E, Frolov I. Inhibition of Transcription and Translation in Sindbis. Society. 2005;79(15):9397–409.
275. Eckei L, Krieg S, Bütepage M, Lehmann A, Gross A, Lippok B, et al. The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases. Sci Rep [Internet]. 2017;7(February):1–18.
276. Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, Neuvonen M, et al. The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket. J Virol. 2009;83(13):6534–45.
277. Leung AKL, Vyas S, Rood JE, Bhutkar A, Sharp PA, Chang P. Poly(ADP-Ribose) Regulates Stress Responses and MicroRNA Activity in the Cytoplasm. Mol Cell. 2011;42(4):489–99.
278. McPherson RL, Abraham R, Sreekumar E, Ong SE, Cheng SJ, Baxter VK, et al. ADP-ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus replication and virulence. Proc Natl Acad Sci U S A. 2017;114(7):1666–71.
279. Strauss EG, Levinson R, Rice CM, Dalrymple J, Strauss JH. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: Sequence and comparison with those of other alphaviruses. Virology. 1988;
280. Meertens L, Hafirassou ML, Couderc T, Bonnet-Madin L, Kril V, Kümmerer BM, et al. FHL1 is a major host factor for chikungunya virus infection. Nature [Internet]. 2019;574(7777):259–63.
281. Mutso M, Morro AM, Smedberg C, Kasvandik S, Aquilimeba M, Teppor M, et al. Mutation of CD2AP and SH3KBP1 binding motif in alphavirus nsP3 hypervariable domain results in attenuated virus. Viruses. 2018;10(5):1–21.
282. Tossavainen H, Aitio O, Hellman M, Saksela K, Permi P. Structural basis of the high

affinity interaction between the Alphavirus nonstructural protein-3 (nsP3) and the SH3 domain of amphiphysin-2. J Biol Chem. 2016;
283. Fros JJ, Domeradzka NE, Baggen J, Geertsema C, Flipse J, Vlak JM, et al. Chikungunya virus nsP3 blocks stress granule assembly by recruitment of G3BP into cytoplasmic foci. J Virol [Internet]. 2012;86(19):10873–9.
284. Kim DY, Reynaud JM, Rasalouskaya A, Akhrymuk I, Mobley JA, Frolov I, et al. 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. PLoS Pathog. 2016;12(8):1–31.
285. Saxton-Shaw KD, Ledermann JP, Borland EM, Stovall JL, Mossel EC, Singh AJ, et al. O'nyong nyong Virus Molecular Determinants of Unique Vector Specificity Reside in Non-Structural Protein 3. PLoS Negl Trop Dis. 2013;7(1).
286. Chen MW, Tan YB, Zheng J, Zhao Y, Lim BT, Cornvik T, et al. Chikungunya virus nsP4 RNA-dependent RNA polymerase core domain displays detergent-sensitive primer extension and terminal adenylyltransferase activities. Antiviral Res [Internet]. 2017;143:38–47.
287. Rubach JK, Wasik BR, Rupp JC, Kuhn RJ, Hardy RW, Smith JL. Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. Virology. 2009;384(1):201–8.
288. Tomar S, Hardy RW, Smith JL, Kuhn RJ. Catalytic Core of Alphavirus Nonstructural Protein nsP4 Possesses Terminal Adenylyltransferase Activity. J Virol. 2006;80(20):9962–9.
289. Hardy RW, Rice CM. Requirements at the 3' End of the Sindbis Virus Genome for Efficient Synthesis of Minus-Strand RNA. J Virol. 2005;79(8):4630–9.
290. Shirako Y, Strauss JH. Requirement for an aromatic amino acid or histidine at the N terminus of sindbis virus RNA polymerase. J Virol. 1998;72(3):2310–5.
291. Rausch JW, Sztuba-Solinska J, Le Grice SFJ. Probing the structures of viral RNA regulatory elements with SHAPE and related methodologies. Front Microbiol. 2018;8(JAN):1–15.
292. Boerneke MA, Ehrhardt JE, Weeks KM. Physical and Functional Analysis of Viral RNA Genomes by SHAPE. Annu Rev Virol. 2019;6(1):93–117.
293. Mondal M, Mukherjee S, Halder S, Bhattacharyya D. Stacking geometry for non-canonical G:U wobble base pair containing dinucleotide sequences in RNA: Dispersion-corrected DFT-D study. Biopolymers. 2015;103(6):328–38.
294. Villordo SM, Gamarnik A V. Differential RNA Sequence Requirement for Dengue Virus Replication in Mosquito and Mammalian Cells. J Virol. 2013;87(16):9365–72.
295. Westhof E, Jaeger L. RNA pseudoknots. Curr Biol. 1992;2(6):300.
296. Firth AE, Brierley I. Non-canonical translation in RNA viruses. J Gen Virol. 2012;93(PART 7):1385–409.
297. Peselis A, Serganov A. Structure and function of pseudoknots involved in gene expression control. Wiley Interdiscip Rev RNA. 2014;5(6):803–22.
298. Filomatori C V., Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, Gamarnik A V. A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev. 2006;20(16):2238–49.
299. Dong H, Zhang B, Shi PY. Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. Virology. 2008;

300. Friebe P, Boudet J, Simorre J-P, Bartenschlager R. Kissing-Loop Interaction in the 3' End of the Hepatitis C Virus Genome Essential for RNA Replication. J Virol. 2005;79(1):380–92.
301. Pley HW, Flaherty KM, McKay DB. Three-dimensional structure of a hammerhead ribozyme. Nature. 1994;
302. Lebars I, Richard T, Di primo C, Toulmé JJ. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. Nucleic Acids Res. 2007;
303. Branfield L. Structural and biochemical analysis of protein/RNA interactions during the initiation of dengue virus genome replication. University of Leeds; 2018.
304. Myong S, Stevens BC, Ha T. Bridging Conformational Dynamics and Function Using Single-Molecule Spectroscopy. Structure. 2006.
305. Nickens DG, Hardy RW. Structural and functional analyses of stem-loop 1 of the Sindbis virus genome. Virology. 2008;370(1):158–72.
306. Malygin AA, Kossinova OA, Shatsky IN, Karpova GG. HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. Nucleic Acids Res. 2013;41(18):8706–14.
307. Biondi E, Burke DH. RNA structural analysis by enzymatic digestion. Methods Mol Biol. 2014;
308. Ding F, Lavender CA, Weeks KM, Dokholyan N V. Three-dimensional RNA structure refinement by hydroxyl radical probing. Nat Methods. 2012;
309. Peattie DA, Gilbert W. Chemical probes for higher-order structure in RNA. Proc Natl Acad Sci U S A. 1980;77(8):4679–82.
310. Wilkinson KA, Merino EJ, Weeks KM. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc. 2006;1(3):1610–6.
311. Berkhout B, Van Wamel JLB. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. J Virol. 1996;
312. Andino R, Rieckhof GE, Achacoso PL, Baltimore D. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA. EMBO J. 1993;12(9):3587–98.
313. Cahour A, Pletnev A, Vazeille-Falcoz M, Rosen L, Lai CJ. Growth-Restricted Dengue Virus Mutants Containing Deletions in the 5′ Noncoding Region of the RNA Genome. Virology. 1995;
314. Gerlach P, Malet H, Cusack S, Reguera J. Structural insights into bunyavirus replication and its regulation by the vRNA promoter. Cell. 2015;
315. Catchpole AP, Mingay LJ, Fodor E, Brownlee GG. Alternative base pairs attenuate influenza A virus when introduced into the duplex region of the conserved viral RNA promoter of either the NS or the PA gene. J Gen Virol. 2003;
316. Hsin WC, Chang CH, Chang CY, Peng WH, Chien CL, Chang MF, et al. Nucleocapsid protein-dependent assembly of the RNA packaging signal of Middle East respiratory syndrome coronavirus. J Biomed Sci. 2018;25(1):1–12.
317. Scott McBride M, Panganiban AT. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. J Virol. 1996;70(5):2963–73.
318. Bell NM, Kenyon JC, Balasubramanian S, Lever AML. Comparative structural effects of HIV-1 gag and nucleocapsid proteins in binding to and unwinding of the viral RNA

packaging signal. Biochemistry. 2012;51(15):3162–9.
319. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, et al. Structure of the HIV-1 RNA packaging signal. Science (80-). 2015;348(6237):917–21.
320. Hayashi T, Shioda T, Iwakura Y, Shibuta H. RNA packaging signal of human immunodeficiency virus type 1. Virology. 1992;188(2):590–9.
321. Pappalardo L, Kerwood DJ, Pelczer I, Borer PN. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. J Mol Biol. 1998;282(4):801–18.
322. Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. Implications for genome recognition. J Mol Biol. 2000;301(2):491–511.
323. Ingemarsdotter CK, Zeng J, Long Z, Lever AML, Kenyon JC. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. Retrovirology [Internet]. 2018;15(1):1–14.
324. White JP, Reineke LC, Lloyd RE. Poliovirus Switches to an eIF2-Independent Mode of Translation during Infection. J Virol. 2011;85(17):8884–93.
325. Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature. 1988;
326. Martínez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, Diaz-Toledano R. Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res. 2015;
327. Fernández IS, Bai XC, Murshudov G, Scheres SHW, Ramakrishnan V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. Cell. 2014;
328. Yamamoto H, Unbehaun A, Loerke J, Behrmann E, Collier M, Bürger J, et al. Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. Nat Struct Mol Biol. 2014;21(8):721–7.
329. Quade N, Boehringer D, Leibundgut M, Van Den Heuvel J, Ban N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. Nat Commun. 2015;6(May):1–9.
330. Terenin IM, Dmitriev SE, Andreev DE, Shatsky IN. Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. Nat Struct Mol Biol. 2008;15(8):836–41.
331. Kim JH, Park SM, Park JH, Keum SJ, Jang SK. eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. EMBO J [Internet]. 2011;30(12):2454–64.
332. Khromykh AA, Meka H, Guyatt KJ, Westaway EG. Essential Role of Cyclization Sequences in Flavivirus RNA Replication. J Virol. 2001;75(14):6719–28.
333. Tilgner M, Deas TS, Shi PY. The flavivirus-conserved penta-nucleotide in the 3′ stem-loop of the West Nile virus genome requires a specific sequence and structure for RNA synthesis, but not for viral translation. Virology. 2005;331(2):375–86.
334. Rohll JB, Percy N, Ley R, Evans DJ, Almond JW, Barclay WS. The 5'-untranslated regions of picornavirus RNAs contain independent functional domains essential for RNA replication and translation. J Virol [Internet]. 1994;68(7):4384–91.
335. Barton DJ, O'Donnell BJ, Flanegan JB. 5′ cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. EMBO J. 2001;20(6):1439–48.
336. Mason PW, Bezborodova S V., Henry TM. Identification and Characterization of a cis-Acting Replication Element (cre) Adjacent to the Internal Ribosome Entry Site of Foot-

and-Mouth Disease Virus. J Virol. 2002;
337. Andino R, Rieckhof GE, Baltimore D. A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA. Cell. 1990;63(2):369–80.
338. Burrill CP, Westesson O, Schulte MB, Strings VR, Segal M, Andino R. Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity. J Virol. 2013;87(21):11670–83.
339. Herold J, Andino R. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Mol Cell. 2001;7(3):581–91.
340. Vogt DA, Andino R. An RNA element at the 5′-end of the poliovirus genome functions as a general promoter for RNA synthesis. PLoS Pathog. 2010;6(6).
341. You S, Rice CM. 3' RNA Elements in Hepatitis C Virus Replication: Kissing Partners and Long Poly(U). J Virol. 2008;82(1):184–95.
342. Alvarez DE, De Lella Ezcurra AL, Fucito S, Gamarnik A V. Role of RNA structures present at the 3′UTR of dengue virus on translation, RNA synthesis, and viral replication. Virology. 2005;339(2):200–12.
343. Villordo SM, Alvarez DE, Gamarnik A V. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. Rna. 2010;16(12):2325–35.
344. Wang CC, Hsu YC, Wu HC, Wu HN. Insights into the coordinated interplay of the sHP hairpin and its co-existing and mutually-exclusive dengue virus terminal RNA elements for viral replication. Virology [Internet]. 2017;505(February):56–70.
345. Kuzembayeva M, Dilley K, Sardo L, Hu WS. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology [Internet]. 2014;454–455(1):362–70.
346. Spahn CMT, Kieft JS, Grassucci RA, Penczek PA, Zhou K, Doudna JA, et al. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. Science (80-). 2001;291(5510):1959–62.
347. Tuplin A. Diverse roles and interactions of RNA structures during the replication of positive-stranded RNA viruses of humans and animals. J Gen Virol. 2015;96(7):1497–503.
348. Kendra JA, de la Fuente C, Brahms A, Woodson C, Bell TM, Chen B, et al. Ablation of Programmed −1 Ribosomal Frameshifting in Venezuelan Equine Encephalitis Virus Results in Attenuated Neuropathogenicity. J Virol. 2017;91(3).
349. Sanz MÁ, Castelló A, Ventoso I, Berlanga JJ, Carrasco L. Dual mechanism for the translation of subgenomic mRNA from Sindbis virus in infected and uninfected cells. PLoS One. 2009;4(3).
350. Fayzulin R, Frolov I. 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. 2004;78(10):4953–64.
351. Hardy RW. The role of the 3′ terminus of the Sindbis virus genome in minus-strand initiation site selection. Virology. 2006;345(2):520–31.
352. Kulasegaran-Shylini R, Thiviyanathan V, Gorenstein DG, Frolov I. The 5'UTR-specific mutation in VEEV TC-83 genome has a strong effect on RNA replication and subgenomic RNA synthesis, but not on translation of the encoded proteins. Virology [Internet]. 2009;387(1):211–21.
353. Hyde JL, Gardner CL, Kimura T, White JP, Liu G, Trobaugh DW, et al. A viral RNA structural element alters host recognition of nonself RNA. Science (80-).

2014;343(6172):783–7.
354. Firth AE, Chung BYW, Fleeton MN, Atkins JF. Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. Virol J. 2008;5:1–19.
355. Chung BYW, Firth AE, Atkins JF. Frameshifting in Alphaviruses: A Diversity of 3′ Stimulatory Structures. J Mol Biol [Internet]. 2010;397(2):448–56.
356. Sanz MA, González Almela E, Carrasco L. Translation of Sindbis Subgenomic mRNA is Independent of eIF2, eIF2A and eIF2D. Sci Rep [Internet]. 2017;7(October 2016):1–14.
357. Frolov I, Hardy R, Rice CM. Cis-acting RNA elements at the 5′ end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. Rna. 2001;7(11):1638–51.
358. Kulasegaran-Shylini R, Atasheva S, Gorenstein DG, Frolov I. Structural and functional elements of the promoter encoded by the 5' untranslated region of the Venezuelan equine encephalitis virus genome. J Virol. 2009;83(17):8327–39.
359. Strauss JH, Strauss EG. The alphaviruses: Gene expression, replication, and evolution. Microbiol Rev. 1994;58(3):491–562.
360. Levis R, Schlesinger S, Huang H V. Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. J Virol [Internet]. 1990;64(4):1726–33.
361. Li ML, Lin YH, Stollar V. A cell-free system for the synthesis of Sindbis virus subgenomic RNA: Importance of the concentration of the initiating NTP. Virology. 2005;341(1):24–33.
362. Wielgosz MM, Raju R, Huang H V. Sequence Requirements for Sindbis Virus Subgenomic mRNA Promoter Function in Cultured Cells. J Virol. 2001;75(8):3509–19.
363. Hertz JM, Huang H V. Utilization of heterologous alphavirus junction sequences as promoters by Sindbis virus. J Virol. 1992;66(2):857–64.
364. Garneau NL, Sokoloski KJ, Opyrchal M, Neff CP, Wilusz CJ, Wilusz J. The 3' Untranslated Region of Sindbis Virus Represses Deadenylation of Viral Transcripts in Mosquito and Mammalian Cells. J Virol. 2008;82(2):880–92.
365. Dickson AM, Anderson JR, Barnhart MD, Sokoloski KJ, Oko L, Opyrchal M, et al. Dephosphorylation of HuR protein during alphavirus infection is associated with HuR relocalization to the cytoplasm. J Biol Chem. 2012;
366. Ou JH, Strauss EG, Strauss JH. Comparative studies of the 3′-terminal sequences of several alphavirus RNAs. Virology. 1981;
367. Kuhn RJ, Hong Z, Strauss JH. Mutagenesis of the 3' nontranslated region of Sindbis virus RNA. J Virol [Internet]. 1990;64(4):1465–76.
368. Raju R, Hajjou M, Hill KR, Botta V, Botta S. In vivo addition of poly(A) tail and AU-rich sequences to the 3' terminus of the sindbis virus RNA genome: A novel 3'-end repair pathway. J Virol. 1999;73(3):2410–9.
369. George J, Raju R. Alphavirus RNA Genome Repair and Evolution: Molecular Characterization of Infectious Sindbis Virus Isolates Lacking a Known Conserved Motif at the 3' End of the Genome. J Virol. 2000;74(20):9776–85.
370. James FD, Hietala KA, Eldar D, Guess TE, Cone C, Mundall N, et al. Efficient replication, and evolution of Sindbis virus genomes with non-canonical 3′ A/U-rich elements (NC3ARE) in neonatal mice. Virus Genes. 2007;35(3):651–62.
371. Reynaud JM, Kim DY, Atasheva S, Rasalouskaya A, White JP, Diamond MS, et al. IFIT1 Differentially Interferes with Translation and Replication of Alphavirus Genomes and Promotes Induction of Type I Interferon. PLoS Pathog. 2015;11(4):1–31.
372. Niesters HG, Strauss JH. Defined mutations in the 5' nontranslated sequence of Sindbis

virus RNA. J Virol [Internet]. 1990;64(9):4162–8.
373. Gorchakov R, Hardy R, Rice CM, Frolov I. Selection of functional 5' cis-acting elements promoting efficient sindbis virus genome replication. J Virol [Internet]. 2004;78(1):61–75.
374. Frey TK, Gard DL, Strauss JH. Biophysical studies on circle formation by Sindbis virus 49S RNA. J Mol Biol. 1979;132(1):1–18.
375. Niesters HG, Strauss JH. Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. J Virol. 1990;64(4):1639–47.
376. Michel G, Petrakova O, Atasheva S, Frolov I. Adaptation of Venezuelan equine encephalitis virus lacking 51-nt conserved sequence element to replication in mammalian and mosquito cells. Virology [Internet]. 2007;362(2):475–87.
377. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature. 1977;
378. Macpherson I, Stoker M. Polyoma transformation of hamster cell clones-an investigation of genetic factors affecting cell competence. Virology. 1962;
379. Nakabayashi H, Miyano K, Sato J, Yamane T, Taketa K. Growth of human hepatoma cell lines with differentiated functions in chemically defined medium. Cancer Res. 1982;
380. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H PW. In vitro cultivation of human tumors: Establishment of cell lines derived from a series of solid tumors. J Natl Cancer Inst. 1973;51(5):1417–23.
381. Condreay LD, Brown DT. Exclusion of superinfecting homologous virus by Sindbis virus-infected Aedes albopictus (mosquito) cells. J Virol. 1986;
382. Singh KR. Cell cultures derived from larvae of Aedes Albopictus (Skuse) and Aedes Aegypti (L.). Source Curr Sci. 1967;
383. Boylan BT, Moreira FR, Carlson TW, Bernard KA. Mosquito cell-derived West Nile virus replicon particles mimic arbovirus inoculum and have reduced spread in mice. PLoS Negl Trop Dis. 2017;
384. Pelletier J, Sonenberg N. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell. 1985;40(3):515–26.
385. Parkin NT, Cohen EA, Darveau A, Rosen C, Haseltine W, Sonenberg N. Mutational analysis of the 5′ non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. EMBO J. 1988;7(9):2831–7.
386. Lodeiro MF, Filomatori C V., Gamarnik A V. Structural and Functional Studies of the Promoter Element for Dengue Virus RNA Replication. J Virol. 2009;83(2):993–1008.
387. Rinnenthal J, Klinkert B, Narberhaus F, Schwalbe H. Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. Nucleic Acids Res. 2010;38(11):3834–47.
388. McDermott JJ, Civic B, Barkan A. Effects of RNA structure and salt concentration on the affinity and kinetics of interactions between pentatricopeptide repeat proteins and their RNA ligands. PLoS One. 2018;13(12):1–12.
389. Chen WJ, Wu HR, Chiou SS. E/NS1 modifications of dengue 2 virus after serial passages in mammalian and/or mosquito cells. Intervirology. 2003;46(5):289–95.
390. Coffey LL, Vasilakis N, Brault AC, Powers AM, Tripet FR, Weaver SC. Arbovirus evolution in vivo is constrained by host alternation. Proc Natl Acad Sci U S A. 2008;105(19):6970–5.
391. Cooper LA, Scott TW. Differential evolution of eastern equine encephalitis virus

populations in response to host cell type. Genetics. 2001;157(4):1403–12.
392. Greene IP, Wang E, Deardorff ER, Milleron R, Domingo E, Weaver SC. Effect of Alternating Passage on Adaptation of Sindbis Virus to Vertebrate and Invertebrate Cells. J Virol. 2005;79(22):14253–60.
393. Villordo SM, Filomatori C V., Sánchez-Vargas I, Blair CD, Gamarnik A V. Dengue Virus RNA Structure Specialization Facilitates Host Adaptation. PLoS Pathog. 2015;
394. Weaver SC, Brault AC, Kang W, Holland JJ. Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. J Virol [Internet]. 1999;73(5):4316–26.
395. Stapleford KA, Moratorio G, Henningsson R, Chen R, Matheus S, Enfissi A, et al. Whole-Genome Sequencing Analysis from the Chikungunya Virus Caribbean Outbreak Reveals Novel Evolutionary Genomic Elements. PLoS Negl Trop Dis. 2016;10(1).
396. Kendall C, Khalid H, Müller M, Banda DH, Kohl A, Merits A, et al. Structural and phenotypic analysis of Chikungunya virus RNA replication elements. Nucleic Acids Res. 2019;47(17):9296–312.
397. Roberts GC, Zothner C, Remenyi R, Merits A, Stonehouse NJ, Harris M. Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research. Sci Rep. 2017;
398. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, Nicol MQ, et al. Elevation of CpG frequencies in influenza a genome attenuates pathogenicity but enhances host response to infection. Elife. 2016;
399. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife. 2014;3:e04531.
400. Lobo FP, Mota BEF, Pena SDJ, Azevedo V, Macedo AM, Tauch A, et al. Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts. PLoS One. 2009;
401. Sørensen MA, Pedersen S. Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. J Mol Biol. 1991;222(2):265–80.
402. Nougairede A, de Fabritus L, Aubry F, Gould E a., Holmes EC, de Lamballerie X. Random Codon Re-encoding Induces Stable Reduction of Replicative Fitness of Chikungunya Virus in Primate and Mosquito Cells. PLoS Pathog. 2013;9(2).
403. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, et al. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc Natl Acad Sci U S A. 1988;85(5):1364–8.
404. Coffey LL, Beeharry Y, Bordería A V., Blanc H, Vignuzzi M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proc Natl Acad Sci U S A. 2011;108(38):16038–43.
405. GenScript. Codon usage table [Internet]. Available from: https://www.genscript.com/tools/codon-frequency-table
406. Murdock CC, Paaijmans KP, Bells AS, King JG, Hillyer JF, Read AF, et al. Complex effects of temperature on mosquito immune function. Proc R Soc B Biol Sci. 2012;279(1741):3357–66.
407. Ferreira PG, Tesla B, Cynthia E, Horácio A, Nahum A, Brindley MA, et al. Temperature Dramatically Shapes Mosquito Gene Expression with Consequences for Mosquito-Zika

Virus Interactions. bioRXiv. 2019;(Cm).
408. Mordecai EA, Cohen JM, Evans M V., Gudapati P, Johnson LR, Lippi CA, et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. PLoS Negl Trop Dis. 2017;11(4):1–18.
409. Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. Internal ribosome entry site within hepatitis C virus RNA. J Virol [Internet]. 1992;66(3):1476–83.
410. Guo L, Allen EM, Miller WA. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. Mol Cell. 2001;7(5):1103–9.
411. Pooggin MM, Ryabova LA, He X, Fütterer J, Hohn T. Mechanism of ribosome shunting in Rice tungro bacilliform pararetrovirus. Rna. 2006;12(5):841–50.
412. Schulze-Gahmen U, Hurley JH. Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. Proc Natl Acad Sci U S A. 2018;
413. Chavali SS, Bonn-Breach R, Wedekind JE. Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. J Biol Chem. 2019;
414. Butrapet S, Kinney RM, Huang CYH. Determining genetic stabilities of chimeric dengue vaccine candidates based on dengue 2 PDK-53 virus by sequencing and quantitative TaqMAMA. J Virol Methods. 2006;
415. Watters KE, Choudhary K, Aviran S, Lucks JB, Perry KL, Thompson JR. Probing of RNA structures in a positive sense RNA virus reveals selection pressures for structural elements. Nucleic Acids Res. 2018;
416. Abbink TEM, Berkhout B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 rna occludes the gag start codon. J Biol Chem. 2003;
417. Morley VJ, Noval MG, Chen R, Weaver SC, Vignuzzi M, Stapleford KA, et al. Chikungunya virus evolution following a large 3′UTR deletion results in host-specific molecular changes in protein-coding regions. Virus Evol. 2018;4(1):1–12.
418. Robinson M, Lilley R, Little S, Emtage JS, Yarranton G, Stephens P, et al. Codon usage can affect efficiency of translation of genes in Escherichia coli. Nucleic Acids Res. 1984;
419. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015;160(6):1111–24.
420. Gouy M, Gautier C. Codon usage in bacteria: Correlation with gene expressivity. Nucleic Acids Res. 1982;10(22):7055–74.
421. Bennetzen JL, Hall BD. Codon selection in yeast. J Biol Chem. 1982;257(6):3026–31.
422. Hastings KEM, Emerson CP. Codon usage in muscle genes and liver genes. J Mol Evol. 1983;19(3–4):214–8.
423. Limmer S, Reif B, Ott G, Arnold L, Sprinzl M. NMR evidence for helix geometry modifications by a G-U wobble base pair in the acceptor arm of E. coli tRNA Ala. FEBS Lett. 1996;
424. Domingo E, Sabo D, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. Cell. 1978;13(4):735–44.
425. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996;242(1):84–9.
426. Haller AA, Semler BL. Stem-Loop Structure Synergy in Binding Cellular Proteins to the 5′ Noncoding Region of Poliovirus RNA. Virology. 1995;206(2):923–34.
427. Smulevitch S, Bear J, Alicea C, Rosati M, Jalah R, Zolotukhin AS, et al. RTE and CTE mRNA export elements synergistically increase expression of unstable, Rev-dependent

HIV and SIV mRNAs. Retrovirology. 2006;3:1–9.
428. Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, et al. A Hepatitis C Virus cis-Acting Replication Element Forms a Long-Range RNA-RNA Interaction with Upstream RNA Sequences in NS5B. J Virol. 2008;82(18):9008–22.
429. Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green SM, Kern J, et al. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. Cell. 1990;60(4):685–93.
430. Tuplin A, Struthers M, Simmonds P, Evans DJ. A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. Nucleic Acids Res. 2012;40(14):6908–21.
431. Kieft JS, Zhou K, Grech A, Jubin R, Doudna JA. Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. Nat Struct Biol. 2002;9(5):370–4.
432. Su L, Chen L, Egli M, Berger JM, Rich A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. Nat Struct Biol. 1999;6(3):285–92.
433. Wyatt JR, Puglisi JD, Tinoco I. RNA pseudoknots. Stability and loop size requirements. J Mol Biol. 1990;214(2):455–70.
434. Gamarnik A V., Andino R. Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev. 1998;12(15):2293–304.
435. Stammler SN, Cao S, Chen SJ, Giedroc DP. A conserved RNA pseudoknot in a putative molecular switch domain of the 3′-untranslated region of coronaviruses is only marginally stable. Rna. 2011;17(9):1747–59.
436. Matsuda D, Dreher TW. The tRNA-like structure of Turnip yellow mosaic virus RNA is a 3′-translational enhancer. Virology. 2004;321(1):36–46.
437. Matsuda D, Yoshinari S, Dreher TW. eFF1A binding to aminoacylated viral RNA represses minus strand synthesis by TYMV RNA-dependent RNA polymerase. Virology. 2004;321(1):47–56.
438. Deiman BALM, Koenen AK, Verlaan PWG, Pleij CWA. Minimal template requirements for initiation of minus-strand synthesis in vitro by the RNA-dependent RNA polymerase of turnip yellow mosaic virus. J Virol. 1998;72(5):3965–72.
439. Jakubiec A, Notaise J, Tournier V, Hericourt F, Block MA, Drugeon G, et al. Assembly of Turnip Yellow Mosaic Virus Replication Complexes: Interaction between the Proteinase and Polymerase Domains of the Replication Proteins. J Virol. 2004;78(15):7945–57.
440. Staple DW, Butcher SE. Pseudoknots: RNA structures with diverse functions. PLoS Biol. 2005;3(6):0956–9.
441. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature. 1988;331(6153):280–3.
442. Wills NM, Gesteland RF, Atkins JF. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. Proc Natl Acad Sci U S A. 1991;88(16):6991–5.
443. Schmidt-Puchta W, Dominguez D, Lewetag D, Hohn T. Plant ribosome shunting in vitro. Nucleic Acids Res. 1997;25(14):2854–60.
444. Pooggin MM, Fütterer J, Skryabin KG, Hohn T. Ribosome shunt is essential for infectivity of cauliflower mosaic virus. Proc Natl Acad Sci U S A. 2001;98(3):886–91.
445. Heublein M, Ndi M, Vazquez-Calvo C, Vögtle FN, Ott M. Alternative Translation

Initiation at a UUG Codon Gives Rise to Two Functional Variants of the Mitochondrial Protein Kgd4. J Mol Biol. 2019;
446. Bartholomeeusen K, Utt A, Coppens S, Rausalu K, Vereecken K, Ariën KK, et al. A Chikungunya Virus trans -Replicase System Reveals the Importance of Delayed Nonstructural Polyprotein Processing for Efficient Replication Complex Formation in Mosquito Cells . J Virol. 2018;92(14).
447. Chase AJ, Daijogo S, Semler BL. Inhibition of Poliovirus-Induced Cleavage of Cellular Protein PCBP2 Reduces the Levels of Viral RNA Replication. J Virol. 2014;88(6):3192–201.
448. Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. 2003;17(21):2688–97.
449. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 1986;44(2):283–92.
450. Hemmings-Mieszczak M, Hohn T, Preiss T. Termination and Peptide Release at the Upstream Open Reading Frame Are Required for Downstream Translation on Synthetic Shunt-Competent mRNA Leaders. Mol Cell Biol. 2000;20(17):6212–23.
451. Shirako Y. Non-AUG translation initiation in a plant RNA virus: a forty-amino-acid extension is added to the N terminus of the soil-borne wheat mosaic virus capsid protein. J Virol [Internet]. 1998;72(2):1677–82.
452. Boeck R, Curran J, Matsuoka Y, Compans R, Kolakofsky D. The parainfluenza virus type 1 P/C gene uses a very efficient GUG codon to start its C' protein. J Virol. 1992;66(3):1765–8.
453. Waggoner JJ, Pinsky BA. How great is the threat of chikungunya virus? Expert Rev Anti Infect Ther. 2015;13(3):291–3.

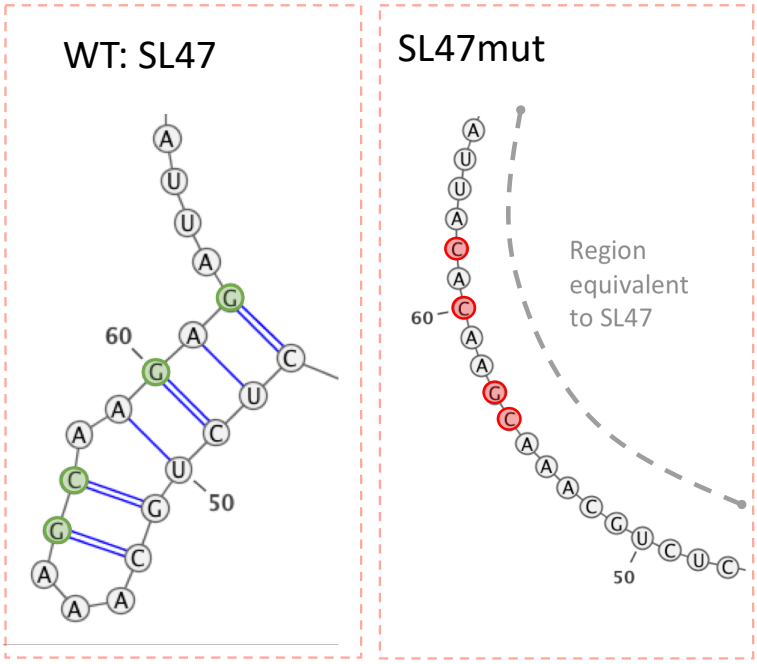
Appendix

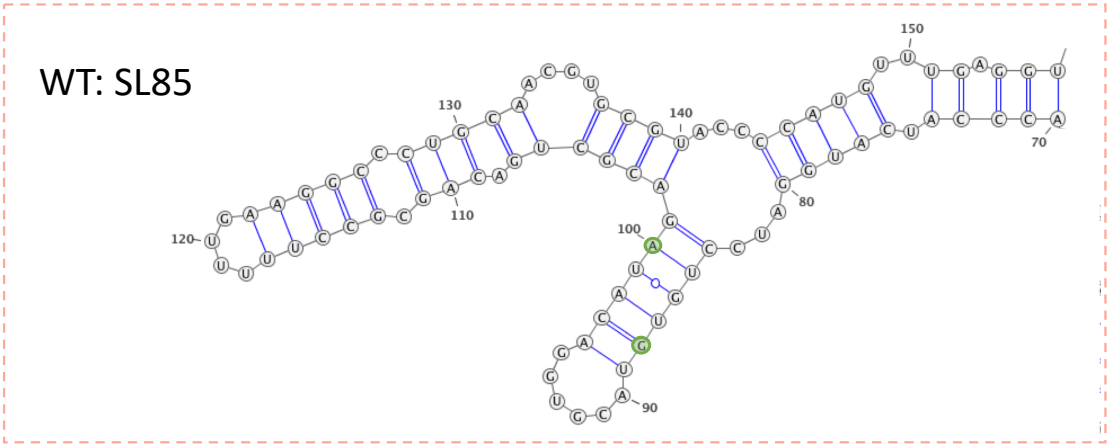
Primer name	Primer sequence
Sequencing Primers	
CHIKV_0_FWD	ATGGCTGCGTGAGACACAC
CHIKV_5_REV	CTACGTGTGTCTCACGCAG
CHIKV_729_REV	TTGCCTCGTCTACCTTCCG
CHIKV_1585_REV	GCTTCGCGAGTCAGTTCT
CHIKV_2401_REV	GACTGGTCTGTTGCATCCA
CHIKV_3123_REV	GAGGATAGGGACCAAGCTCT
CHIKV_3945_REV	GTGTCACCAGCAACACTGAG
CHIKV_4817_REV	TGCGTCTGCATCATCCAC
CHIKV_5590_REV	CGCTGTCTGTCAAGTCATCC
CHIKV_6389_REV	TCCTAATAGGGCTGGCAGC
CHIKV_7196_REV	CCTGTCACAGTATCGTGCAG
CHIKV_7946_REV	CTTGCTCACGAATACGACGG
CHIKV_8432_FWD	TGCTGGTGCCCACACTAT
CHIKV_8599_REV	CGCTGGTTGTTTCTGTCAG
CHIKV_9194_FWD	TTAAGGCCAAGAAGGGCGG
CHIKV_9995_FWD	CCCCTGACGAGCATCACAAA
CHIKV_10805_FWD	GACTCCCCGTCGTGTAGAT
CHIKV_11560_FWD	AAGGGAATAAGGGCGACAC
ORF2 Sequencing Primers	
CHIKV_7471_FWD	CGTCATAACTTTGTACGGCGG
CHIKV_8206_FWD	GGCAGACCGATCTTCGACAA
CHIKV_9084_FWD	GTCACAACAGTCCGGCAAC
CHIKV_9920_FWD	GACTCTTACCATGCTGCTGT
CHIKV_10765_FWD	GGCTGCCAAATAGCAACAA
Passaging Sequencing Primers	
CHIKV 1-20_FWD	ATGGCTGCGTGAGACACACG
CHIKV 1018_REV	CCGTGTCGGTAGTCTTGCAC
CHIKV 989_REV	CCGTCTGCGTGGTGGGTTAC
Walking Sequencing Primers	
CHIKV_301_FWD	CAGGAAGTACCACTGCGTCTGCCCG
CHIKV_1209_FWD	CCATGAAAAATTATCTGCTTCCCGTGGTCGCCCAAGCC
CHIKV_1284_REV	CCATGTCTTTCCGGCACTCCTTTGCCCACTTAC
CHIKV_2103_FWD	GCTGTAAGAAGGAAGAAGCCGCAGGACTGG
CHIKV_2186_REV	GCCCTTCATATGCGAATTCGTGGTAGGGCGG

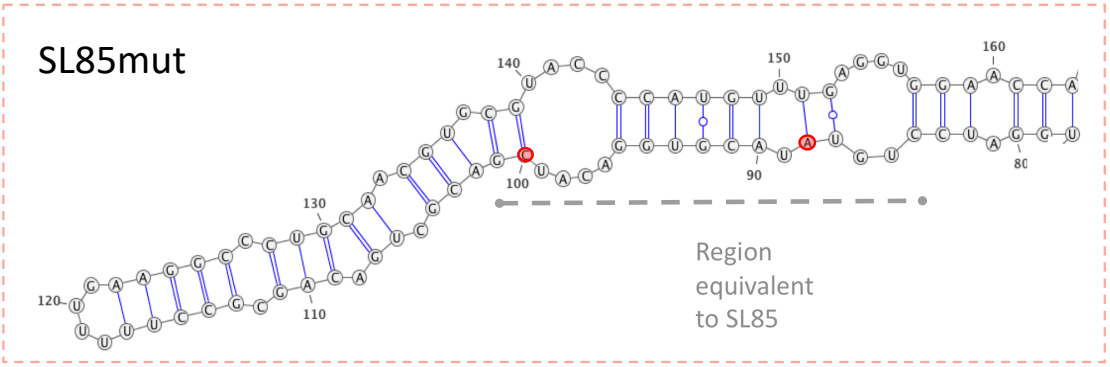
CHIKV_3002_FWD	GGAAACTTCAAAGCAACTATTAAGGAGTGGGAGG
CHIKV_3082_REV	GGTCATTTGGTGACTGCAGATGCCCGCC
CHIKV_3930_FWD	CGTTGAAACCACCATGTGTCACCAGCAACAC
CHIKV_4003_REV	CCTTCTGCCATTGTCAAAGTTGCTGAATAGG
CHIKV_4830_FWD	ACGCATCATCTCCCCCCAAAACTGTCCCG
CHIKV_4892_REV	CGCGTTCTGGAGTCATAGCGTAACGGC
CHIKV_5467_FWD	GCCGATCTCCTTCGGAGCATCAAGCG
CHIKV_5518_REV	GTCCCCAAATGTAATGGGGAACGTCTCGC
CHIKV_6341_REV	CCACGTTGAATACTGCTGAGTCCAAAGTGGG
Combination Mutagenesis Primers	
SL84_FWD	ATGGATCCTGTATACGTGGACATCGACGCTGACAG
SL84_REV	CTGTCAGCGTCGATGTCCACGTATACAGGATCCAT
SL102_FWD	GCCTTTTTGAAAGCCCTACAACGTGCATACCCCATGTTTG
SL102_REV	CAAACATGGGGTATGCACGTTGTAGGGCTTTCAAAAAGGC
SL165_FWD	CAAGGCAGGTAACACCGAATGACCATGCAAATGCTAGAGC
SL165_REV	GCTCTAGCATTTGCATGGTCATTCGGTGTTACCTGCCTTG
SL194_FWD	GCGTTCTCGCACCTCGCTATAAAACTAATAG
SL194_REV	CTATTAGTTTTATAGCGAGGTGCGAGAACGC
SL246_FWD	AAATTGACCCGGACTCTACAATACTGGATATCGGGAGTGCCCCAG
SL246_REV	CTGGGGCACTCCCGATATCCAGTATTGTAGAGTCCGGGTCAATTT
Compensatory Mutagenesis Primers	
SL84c_FWD	ATCATGGATCCGGTATACGTGGATATCGACGCTGAC
SL84c_REV	GTCAGCGTCGATATCCACGTATACCGGATCCATGAT
SL102c_FWD	CGTGGACATAGATGCTGATAGCGCTTTTTTGAAAGCCCTAC
SL102c_REV	GTAGGGCTTTCAAAAAAGCGCTATCAGCATCTATGTCCACG
SL165c_FWD	CAAGGCAGGTTACACCGAATGACCATGCCAATGCTAGAGC
SL165c_REV	GCTCTAGCATTGGCATGGTCATTCGGTGTAACCTGCCTTG
SL194c_FWD	CATGCTAATGCGAGGGCGTTCTCGCAC
SL194c_REV	GTGCGAGAACGCCCTCGCATTAGCATG
SL246c_FWD1	AAATTGACCCGGACTCGACTATTCTGGATATCGGGAGTGCGCCAG
SL246c_REV1	CTGGCGCACTCCCGATATCCAGAATAGTCGAGTCCGGGTCAATTT
SL246c_FWD2	AGGATGATGTCCGACAGGAAGTACC
SL246c_REV2	GGTACTTCCTGTCGGACATCATCCT
Terminal Loop Mutagenesis Primers	
194i_terminal_loop_FWD	CATGCTAATGCTAGAGCGTTTTCGCATCTAGCTATAAAAC
194i_terminal_loop_REV	GTTTTATAGCTAGATGCGAAAACGCTCTAGCATTAGCATG
SL165i_terminal_loop_FWD	GTGGAACCAAGGCAGGTCACTCCTAATGACCATGCTA
SL165i_terminal_loop_REV	TAGCATGGTCATTAGGAGTGACCTGCCTTGGTTCCAC
SL165ii_terminal_loop_FWD	GTGGAACCAAGGCAGGTCACGCCAAATGACCATGCTA

SL165ii_terminal_loop_REV	TAGCATGGTCATTTGGCGTGACCTGCCTTGGTTCCAC
SL165iii_terminalloopFWD	GTGGAACCAAGGCAGGTCACCCCGAATGACCATGCTA
SL165iii_terminal_loop_REV	TAGCATGGTCATTCGGGGTGACCTGCCTTGGTTCCAC
SL165iv_terminal_loop_FWD	GGAACCAAGGCAGGTAACACCGAACGACCATGCTAATGCTA
SL165iv_terminal_loop_REV	TAGCATTAGCATGGTCGTTCGGTGTTACCTGCCTTGGTTCC
Pseudoknot Mutagenesis Primers	
$Pk1_L$_mut_FWD	TAACCCATCATGGATCCTGTCTATGTCGACATAGACGCTGACAGC
$Pk1_L$_mut_REV	AAGGCGCTGTCAGCGTCTATGTCGACATAGACAGGATCCATGATG
$Pk1_S$_mut_FWD	CAAGAGATTAATAACCCAAGATGGATCCTGTGTACGTGGACATAG
$Pk1_S$_mut_REV	GTCTATGTCCACGTACACAGGATCCATCTTGGGTTATTAATCTCT
Pk1-Comp_FWD	TTAATAACCCATGATGGATCCTGTATATGTCGACATAGACGCTGAC
Pk1-Comp_REV	GTCAGCGTCTATGTCGACATATACAGGATCCATCATGGGTTATT
$Pk2_S$_mut_FWD	GGCCCTGCAACGTGCCTATCCCATGTTTGAGGTGG
$Pk2_S$_mut_REV	CCACCTCAAACATGGGATAGGCACGTTGCAGGGCC
AUG1STOP_FWD	CCCATCATGGATCCTGTATAGGTGGACATAGACGCTG
AUG1STOP_REV	CAGCGTCTATGTCCACCTATACAGGATCCATGATGGG
AUG1mut_FWD	GCAAGAGATTAATAACCCATCGCGGATCCTGTGTACGTG
AUG1mut_REV	CACGTACACAGGATCCGCGATGGGTTATTAATCTCTTGC
Passage mutant mutagenesis primers	
UUGmut_FWD	GCCCTGCAACGTGCGTACCCCTTGTTTGAGGTGGAACC
UUGmut_REV	GGTTCCACCTCAAACAAGGGGTACGCACGTTGCAGGGC
SL194mut(A199G)_FWD	CCATGCTAATGCTAGGGCGTTCTCGCACCTCGCTATAAAAC
SL194mut(A199G)_REV	GTTTTATAGCGAGGTGCGAGAACGCCCTAGCATTAGCATGG

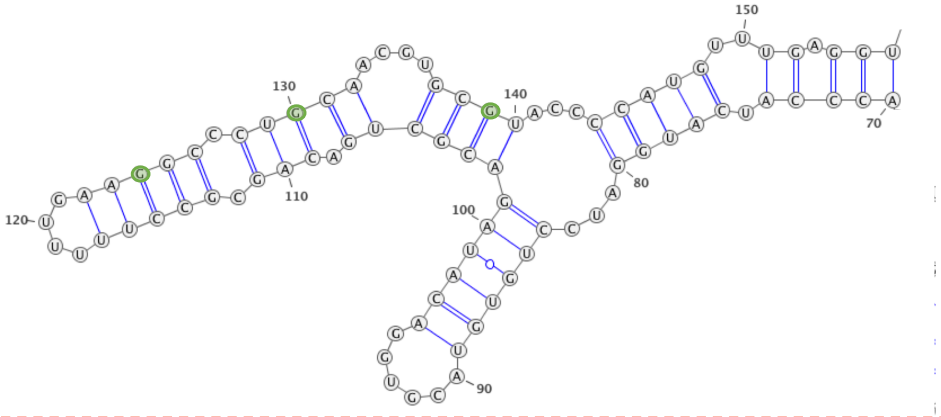
Table 1: Oligonucleotide primers


WT: SL47
60
50
SL47mut
60
50
Region
equivalent
to SL47

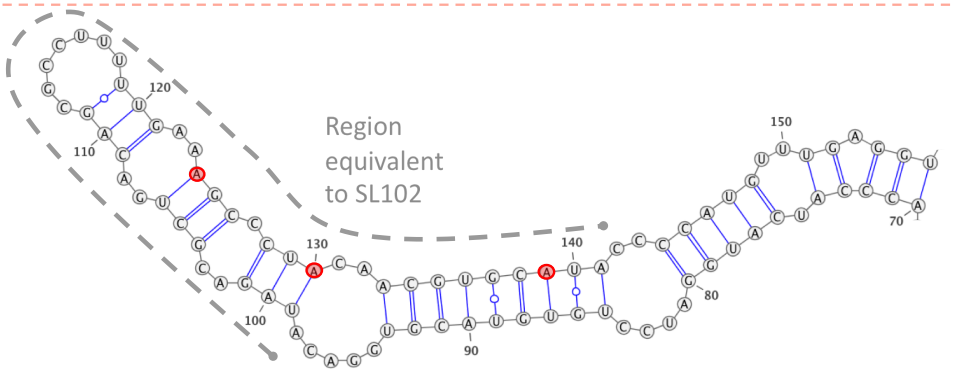

WT: SL85
120
130
110
140
100
90
150
70
80


SL85mut
120
130
110
140
100
150
90
160
80
Region
equivalent
to SL85

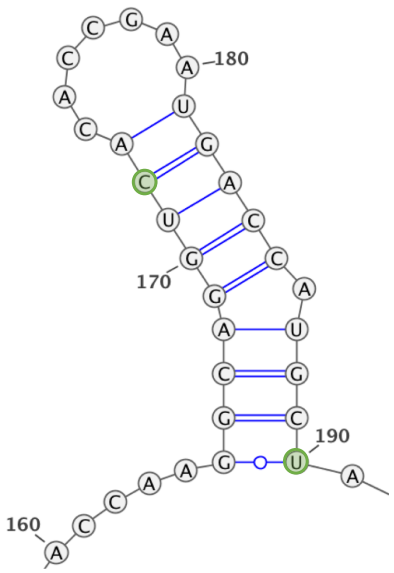
WT: SL102



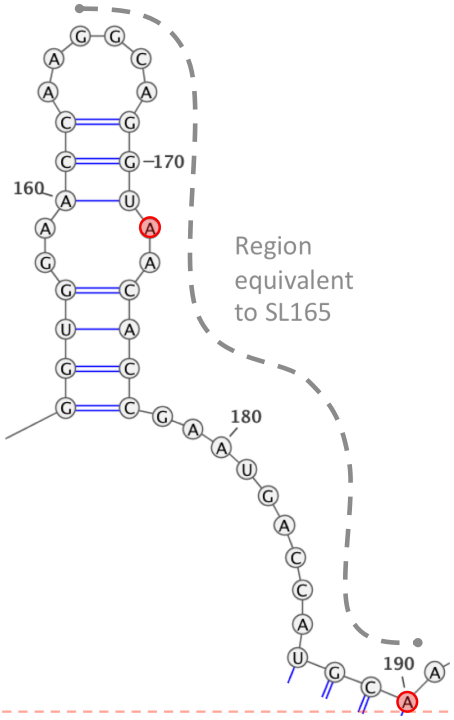
SL102mut



WT: SL165



SL165mut



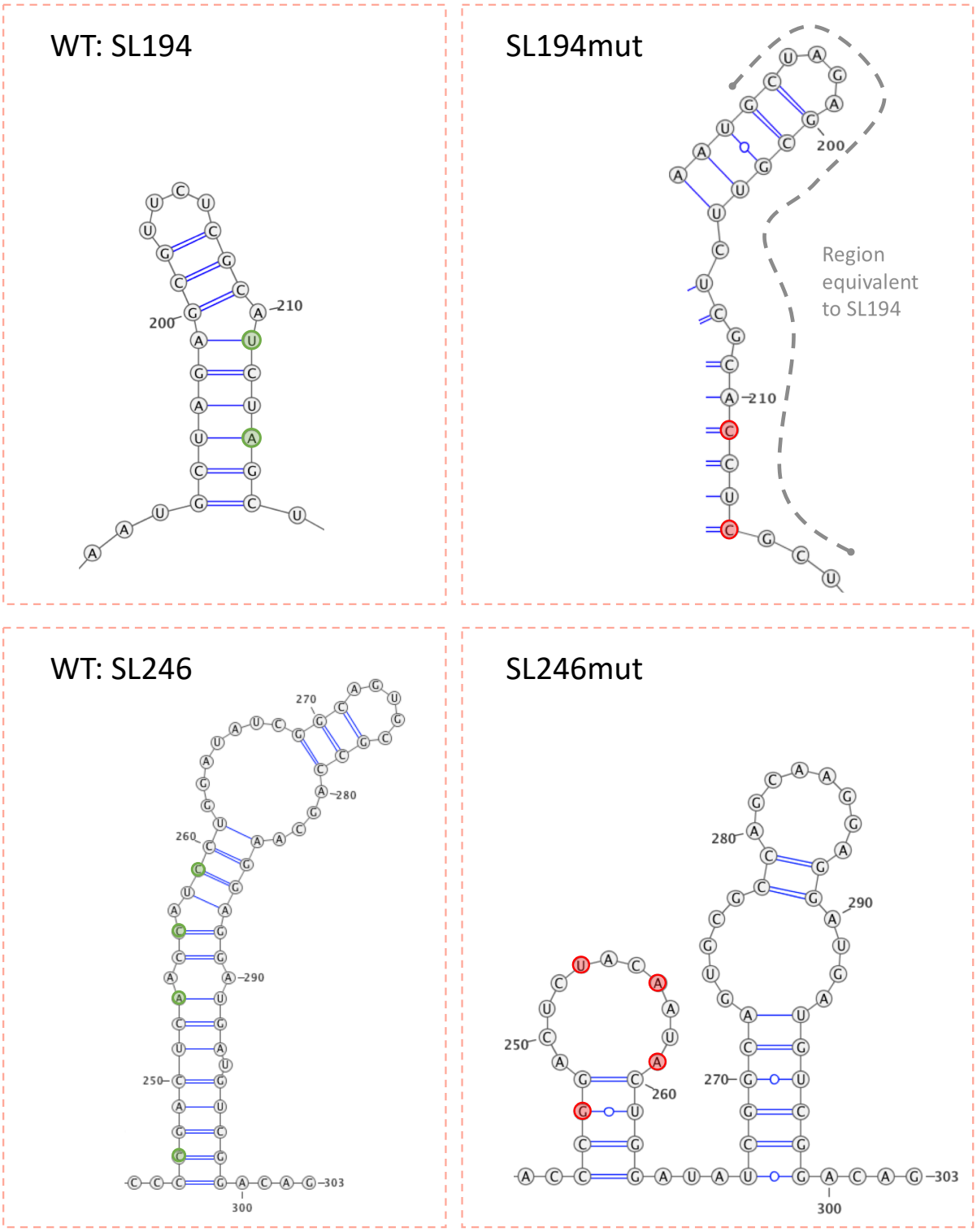


Table 2: UNAfold predictions of stem-loop mutants in the positive strand relative to WT