

**Untangling the impact of
mutations in the coding region of
chikungunya virus nsP1**

Kate Amy Loveday

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

The University of Leeds
School of Molecular and Cellular Biology

December 2022

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

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

© 2022 The University of Leeds and Kate Loveday.

Acknowledgements

Firstly, I would like to thank Dr Andrew Tuplin for his support and guidance throughout my PhD, and for keeping me on track when I would lose sight of the final picture. I am grateful to Prof. Andres Merits who kindly provided the CHIKV constructs, as well as the α -CHIKV nsP1 antibody, without which this project would not have been possible.

Thank you to all current and past members of the Tuplin group for providing support and advice across the years. Special thanks go to Veronica for always being there for a chat and/or rant, especially throughout the writing months. Thank you for pushing me to apply for jobs and for manifesting my life in Sheffield, it is without a doubt I'd be unemployed without your recruitment expertise. Thank you also to Fran, for understanding the frustrations with SHAPE and for becoming the group mother hen without anybody asking, we'd have fallen apart without you!

Thank you to Ellie who became my lab wife, I will miss you and your sliders but at least I'll stop losing all my baking equipment! Please never stop sending me Twilight related memes. Amelia, thank you for always providing an answer to the most random questions I would ask. Thank you all for making Friday afternoons a hoot with hangman and GeoGuesser. Thank you to Anya for always being my Cat-3 buddy and absolutely smashing the high-score on Snake, thank you for always stopping by to say hi as you walked past my desk. I cannot thank Suki enough for her continued support and friendship over the years, thank you for always reminding me to believe in myself and that there is in fact light at the end of this very long tunnel. I must also thank Connor, it's a good job we've only been friends for 6 months as I'm not sure I'd have achieved much over the last 4 years due to the near constant coffee and biscuit breaks we have. Thanks to Freddie for the white room tour of Leeds and never ending food recommendations.

You wanted a whole page but a paragraph will do - Joe, the amount of science advice and support you have provided over the last 4 years has been irreplaceable.

Thank you for reminding me why science *is* fun and for your never ending excitement when I could only see the flaws. Thank you for always being keen for the pub (even when we should've gone climbing) and eating an ungodly amount of food at Slap & Pickle. Thank you for providing me with a roof over my head whilst I finished my PhD, sorry I always had an excuse to avoid doing any renovation work!

There are even more people outside of the lab to thank: I must thank Deborah and Beckie for always being at the other end of a text, and for agreeing to not ask me questions about my PhD this year - chin chin! Thank you to everybody from the kayaking gang, but most importantly to Matt and Kate. Thank you both for the daily walks during lockdown and for making me laugh and smile on days I didn't want to. Thank you to Kate for bravely living with me whilst I have finished my PhD, I'm unbelievably grateful for all the cooked meals and chocolate - I promise I'll eat it one day! Thank you goes to Murray and Harry for always listening to my rambling voicenotes, thank you to Murray for the nightly gratitudes and to Harry for the beige dinners.

Everybody at Alderley Park Lighthouse labs, you were all so welcoming and friendly. Thank you for showing me a different side to science and for giving me the confidence I needed to grow professionally. Thank you to Stefan and Naomi for getting me out of being an "assistant project manager". Thank you to the automation girls for providing an excuse for a natter. Special thanks to Bev (and Joel) for making my time in Manchester more fun than I'd ever imagined.

Greg - I simply cannot thank you enough. Your unwavering support over the last 5 years has been invaluable. Thank you for keeping me level headed and reminding me to control the controllables. Sorry for always calling you when I'd be crying over nothing. You deserve a medal for your patience.

Finally, thank you to my family. Grandma, I'm not sure you've really understood what I've been doing the last 5 years but you've been proud nonetheless! The biggest thank you has to go to my parents, without your eternal support my whole journey at Leeds would have been impossible.

Abstract

Chikungunya virus (CHIKV) is a single-stranded, positive-sense RNA virus of the *alphavirus* genus and is the known causative agent of chikungunya fever. CHIKV is an arbovirus and transmitted between humans and *Aedes* spp. mosquitoes. Increasing global temperatures has resulted in the expansion of the mosquitoes geographical range. As a consequence the epidemic incident rate of CHIKV is increasing and it is now recognised as a potential threat to global health. There are currently no approved specific therapeutics or vaccines against CHIKV, so a greater understanding into the CHIKV lifecycle is essential. Preliminary data from our group demonstrated that the presence of a second methionine (M²⁴) in the non-structural protein 1 (nsP1) coding region enhanced CHIKV replication and hypothesised that this residue functioned as an alternative start codon to initiate translation of truncated nsP1.

A panel of mutants were introduced at M²⁴ in the full-length infectious clone, sub-genomic replicon and *trans*-replicase assay constructs. Attenuated virus replication was observed across the mutant panel in both mammalian and mosquito cells. Further experiments demonstrated that mutations did not inhibit replication in the context of nsP1 functioning in the replicase complex, but rather in the context of the RNA template. Previous studies have clearly demonstrated the importance of RNA structures within the 5' conserved sequence element. To understand the impact to RNA secondary structures across the mutant panel, selective 2' hydroxyl acetylation analysed by primer extension (SHAPE) was used to inform the modelling of RNA structures within the 5' region of CHIKV. The SHAPE-constrained minimum free energy models confirmed that these structures were disrupted to varying extents across the mutant panel, except M²⁴S which had high structural similarity with wild-type CHIKV. In order to distinguish a specific mechanism for this disrupted region, M²⁴A was passaged for 10 rounds in human and mosquito cells, and alternately between the two cell lines. Extracted cellular RNA was analysed by Next Generation Sequencing, and substitutions were

observed at nucleotide ^C147^U following passaging in human cells and alternate passaging between human and mosquito cell lines. Pseudo-escape mutations were observed at nucleotide ^A77^G, which caused an amino acid change of the ORF-1 start codon to a valine. Attempts to confirm whether these mutations were sufficient to restore virus replication to wild-type levels remained inconclusive. However, it was demonstrated that these mutations were sufficient to restore wild-type base pairing in RNA structures.

In parallel, this project aimed to determine the molecular structure of CHIKV nsP1 through X-ray crystallography. Recombinant full-length CHIKV nsP1 was successfully expressed and purified through a bacterial expression system and nickel affinity chromatography however contamination products were consistently present following size exclusion chromatography, and therefore the protein was not taken forward for X-ray crystallography trials.

Contents

| | |
|--|------------|
| Acknowledgements | ii |
| Abstract | iv |
| List of Figures | x |
| List of Tables | xiv |
| Abbreviations | xv |
| 1 Introduction | 1 |
| 1.1 Chikungunya virus | 1 |
| 1.1.1 Classification of Alphaviruses | 1 |
| 1.1.2 Epidemiology of CHIKV | 3 |
| 1.1.3 CHIKV Phylogeny | 6 |
| 1.1.4 Transmission of CHIKV | 7 |
| 1.1.5 CHIKV infection in humans | 11 |
| 1.1.6 Alphavirus infection in <i>Aedes</i> spp. mosquitoes | 13 |
| 1.1.7 Therapeutics and preventative measures against CHIKV | 13 |
| 1.2 CHIKV Molecular Biology | 14 |
| 1.2.1 CHIKV Virion Structure | 14 |
| 1.2.2 CHIKV Genome Organisation | 15 |
| 1.2.3 Function of non-structural proteins | 16 |
| 1.2.4 Alphavirus replication | 21 |
| 1.3 RNA structures | 26 |
| 1.3.1 Formation of RNA structures | 26 |
| 1.3.2 Predicting and visualising RNA structures | 27 |
| 1.3.3 Importance of RNA structures | 29 |
| 1.3.4 Importance of viral RNA structures | 31 |
| 1.3.5 Alphavirus RNA structures | 34 |
| 1.4 Aims of the project | 42 |
| 2 Materials and Methods | 44 |
| 2.1 DNA/RNA | 44 |
| 2.1.1 Transformation of bacterial cells | 44 |

CONTENTS

| | | |
|--------|---|----|
| 2.1.2 | Endonuclease digestion | 45 |
| 2.1.3 | Agarose TAE gel electrophoresis | 45 |
| 2.1.4 | Extraction of DNA | 45 |
| 2.1.5 | Ligation of DNA | 46 |
| 2.1.6 | Phenol-chloroform extraction and ethanol precipitation of DNA | 46 |
| 2.1.7 | Oligonucleotide primers | 47 |
| 2.1.8 | DNA sequencing | 47 |
| 2.1.9 | <i>In vitro</i> transcription | 47 |
| 2.1.10 | Lithium chloride purification of RNA | 48 |
| 2.1.11 | Phenol-chloroform extraction of RNA | 48 |
| 2.1.12 | Column purification of RNA | 49 |
| 2.1.13 | RNA agarose gel electrophoresis | 49 |
| 2.2 | Selective 2' Hydroxyl Acetylation Analysed by Primer Extension (SHAPE) | 49 |
| 2.2.1 | Template preparation | 49 |
| 2.2.2 | RNA folding and NMIA treatment | 50 |
| 2.2.3 | Primer extension | 50 |
| 2.2.4 | Ladder preparation | 51 |
| 2.2.5 | Capillary Sequencing of SHAPE reactions | 51 |
| 2.2.6 | Data analysis | 51 |
| 2.3 | Cell culture methods | 52 |
| 2.3.1 | Continuous cell culture | 52 |
| 2.3.2 | Transfection of cells | 52 |
| 2.4 | Virological methods | 54 |
| 2.4.1 | CHIKV generation | 54 |
| 2.4.2 | Infectious centre assay | 55 |
| 2.4.3 | Virus infection | 55 |
| 2.4.4 | CHIKV titre determination by plaque assay | 55 |
| 2.4.5 | Passaging | 56 |
| 2.4.6 | RNA Extraction | 56 |
| 2.4.7 | Next Generation Sequencing | 57 |
| 2.5 | Protein production | 59 |

CONTENTS

| | | |
|----------|---|-----------|
| 2.5.1 | Protein expression | 59 |
| 2.5.2 | Protein purification | 60 |
| 2.5.3 | Nickel affinity chromatography | 60 |
| 2.5.4 | Dialysis and cleavage of fusion tag | 61 |
| 2.5.5 | Size exclusion chromatography | 61 |
| 2.5.6 | Circular dichroism | 62 |
| 2.5.7 | Sparse matrix screening - X-ray crystallography trials . . . | 62 |
| 2.6 | Protein analysis | 62 |
| 2.6.1 | SDS Polyacrylamide Gel Electrophoresis (SDS PAGE) | 62 |
| 2.6.2 | Protein Visualisation by Coomassie | 63 |
| 2.6.3 | Western Blot Analysis | 63 |
| 3 | Investigating the importance of M²⁴ across the CHIKV lifecycle | 64 |
| 3.1 | Introduction | 64 |
| 3.1.1 | The function of nsP1 | 64 |
| 3.1.2 | Project background and preliminary results | 67 |
| 3.1.3 | Aims | 70 |
| 3.2 | Results | 70 |
| 3.2.1 | Phenotype of M ²⁴ mutations in infectious virus | 70 |
| 3.2.2 | No truncated nsP1 expression observed during CHIKV replication in BHK-21 or Huh7 cells | 79 |
| 3.2.3 | Introduction of mutants at M ²⁴ does not impact CHIKV RNA levels during infection | 81 |
| 3.2.4 | Mutations at M ²⁴ do not inhibit CHIKV genome replication | 83 |
| 3.2.5 | Substitutions at M ²⁴ do not inhibit the function of nsP1 . | 85 |
| 3.3 | Discussion | 90 |
| 4 | SHAPE-ing M²⁴ mutations in the CHIKV genome | 95 |
| 4.1 | Introduction | 95 |
| 4.1.1 | CHIKV replication elements | 95 |
| 4.1.2 | Aims | 101 |
| 4.2 | Results | 101 |
| 4.2.1 | SHAPE analysis of the M ²⁴ mutant panel | 101 |

CONTENTS

| | | |
|----------|---|------------|
| 4.2.2 | Monitoring of pseudo-escape mutations through serial passaging of M ²⁴ A CHIKV | 112 |
| 4.2.3 | Reverse genetic confirmation of CHIKV M ²⁴ A pseudo-escape mutants | 128 |
| 4.2.4 | M ²⁴ V CHIKV replication competed with wild-type CHIKV replication | 130 |
| 4.2.5 | SNPs detected during passaging of mutant CHIKV result in RNA secondary structure changes | 135 |
| 4.3 | Discussion | 139 |
| 4.3.1 | Mutations in the M ²⁴ codon disrupted RNA secondary structures | 139 |
| 4.3.2 | Passaging of M ²⁴ A CHIKV leads to pseudo-escape mutants to rescue wild-type replication | 141 |
| 5 | Purification of recombinant CHIKV nsP1 | 147 |
| 5.1 | Introduction | 147 |
| 5.1.1 | Aims | 148 |
| 5.2 | Results | 149 |
| 5.3 | Discussion | 163 |
| 6 | Discussion | 168 |
| 6.1 | Concluding remarks | 172 |
| 7 | Appendix | 174 |
| | References | 195 |

List of Figures

| | | |
|------|--|----|
| 1.1 | Phylogenic tree of the <i>alphavirus</i> genus | 3 |
| 1.2 | Geographical distribution of CHIKV | 6 |
| 1.3 | CHIKV phylogenic tree | 7 |
| 1.4 | A schematic of the transmission cycle of CHIKV | 8 |
| 1.5 | Alphavirus virion structure | 15 |
| 1.6 | CHIKV genome schematic | 16 |
| 1.7 | Alphavirus replication in mammalian cells | 21 |
| 1.8 | Schematic representation of RNA structures | 27 |
| 1.9 | Examples of viral RNA structures important for replication | 33 |
| 1.10 | SHAPE-constrained thermodynamic folding prediction of the CHIKV 5' region | 35 |
| 1.11 | Schematic of the 5' 7-methylguanosine mRNA cap structure | 37 |
| 3.1 | Capping mechanism of nsP1. | 66 |
| 3.2 | Sequence alignment of nsP1 across the alphaviruses. | 68 |
| 3.3 | Wild-type and M ²⁴ mutation viral titres following transfection of CHIKV RNA into BHK-21 and Huh7 cells. | 72 |
| 3.4 | Infectious centre assay of M ²⁴ mutations. | 73 |
| 3.5 | Viral titres of wild-type and M ²⁴ mutations following infection at MOI = 0.001 | 75 |
| 3.6 | Wild-type and M ²⁴ mutation viral titres following transfection of CHIKV RNA in C6/36 cells. | 77 |
| 3.7 | Viral titres of wild-type and M ²⁴ mutations following infection at MOI = 0.1 | 78 |
| 3.8 | Western blot analysis of CHIKV infected cell lysates | 80 |
| 3.9 | Ratio of positive to negative strand RNA levels in wild-type and M ²⁴ mutant viral infections | 82 |
| 3.10 | Firefly luciferase sub-genomic replicon | 83 |
| 3.11 | Phenotype of M ²⁴ mutants in CHIKV sub-genomic replicon | 84 |

LIST OF FIGURES

| | | |
|------|---|-----|
| 3.12 | <i>Trans</i> -replicase assay schematic | 86 |
| 3.13 | <i>Trans</i> -replicase assay development | 88 |
| 3.14 | The phenotype of M ²⁴ A in the <i>trans</i> -replicase assay | 89 |
| 4.1 | SHAPE-constrained thermodynamic folding prediction of the 5' 51 nt CSE in alphaviruses | 97 |
| 4.2 | SHAPE-constrained thermodynamic folding prediction of CHIKV 5' 300 nts. | 98 |
| 4.3 | SHAPE-constrained thermodynamic folding prediction of CHIKV SL102 and 5' CSE | 100 |
| 4.4 | Schematic of the SHAPE protocol | 103 |
| 4.5 | SHAPE constrained thermodynamic model of wild-type CHIKV . | 106 |
| 4.6 | SHAPE constrained thermodynamic model of M ²⁴ A CHIKV . . . | 107 |
| 4.7 | SHAPE constrained thermodynamic model of M ²⁴ Q CHIKV . . . | 108 |
| 4.8 | SHAPE constrained thermodynamic model of M ²⁴ S CHIKV . . . | 109 |
| 4.9 | SHAPE constrained thermodynamic model of M ²⁴ LI CHIKV . . . | 110 |
| 4.10 | SHAPE constrained thermodynamic model of M ²⁴ LII CHIKV . . | 111 |
| 4.11 | Schematic representation of the amplicons generated during the library preparation of CHIKV for NGS analysis | 114 |
| 4.12 | Confirmation that the primer pairs designed generate amplicon products of the correct length | 115 |
| 4.13 | Schematic of the library preparation workflow for NGS | 116 |
| 4.14 | Genome coverage plot of CHIKV RNA | 117 |
| 4.15 | Viral titres of wild-type and M ²⁴ A CHIKV throughout passaging in Huh7 cells | 120 |
| 4.16 | Viral titres of wild-type and M ²⁴ A CHIKV throughout passaging in Huh7 and C6/36 cells alternately | 123 |
| 4.17 | Viral titres of wild-type and M ²⁴ CHIKV throughout passaging in C6/36 cells | 126 |
| 4.18 | Titre of escape mutations following transfection of BHK-21 cells . | 130 |
| 4.19 | Titre of virus generated across passaging the competition assay of mutant CHIKV against wild-type CHIKV | 131 |

LIST OF FIGURES

| | | |
|------|---|-----|
| 4.20 | NGS analysis of RNA extracted from cells co-transfected with wild-type and mutant CHIKV RNA | 134 |
| 4.21 | <i>In silico</i> minimum free energy RNA secondary structure prediction of M ¹ V & M ²⁴ A and M ²⁴ V CHIKV | 136 |
| 4.22 | SHAPE constrained thermodynamic model of M ¹ V & M ²⁴ A CHIKV | 137 |
| 4.23 | SHAPE constrained thermodynamic model of M ²⁴ V CHIKV . . . | 138 |
| 5.1 | Schematic representation of the 6xHis-SUMO nsP1 construct . . . | 148 |
| 5.2 | Schematic of the protein purification workflow for CHIKV nsP1 . | 150 |
| 5.3 | Primary and secondary HisTrap column fractions from the first purification. | 152 |
| 5.4 | Size exclusion chromatography of the first purification. | 153 |
| 5.5 | Circular dichroism spectrum of purified CHIKV nsP1 | 154 |
| 5.6 | Primary and secondary HisTrap column fractions from the second purification. | 156 |
| 5.7 | Size exclusion chromatography of the second purification | 158 |
| 5.8 | SDS-PAGE analysis of purified CHIKV nsP1 | 159 |
| 5.9 | HisTrap column fractions from the third purification. | 161 |
| 5.10 | Size exclusion chromatography of the third purification. | 163 |
| 5.11 | Crown-shaped ring structure of CHIKV nsP1 | 165 |
| 5.12 | Molecular structure of CHIKV replicase complex | 167 |
| 7.1 | Strand specific qPCR of WT and M ²⁴ mutations | 174 |
| 7.2 | Introduction of M ²⁴ A mutant in the <i>trans</i> -replicase assay | 175 |
| 7.3 | SHAPE data analysis workflow | 176 |
| 7.4 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of wild-type CHIKV | 177 |
| 7.5 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of M ²⁴ A CHIKV | 178 |
| 7.6 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of M ²⁴ Q CHIKV | 179 |
| 7.7 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of M ²⁴ S CHIKV | 180 |

LIST OF FIGURES

| | | |
|------|---|-----|
| 7.8 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of M ²⁴ LI CHIKV | 181 |
| 7.9 | <i>In-silico</i> minimum free energy RNA secondary structure prediction of M ²⁴ LII CHIKV | 182 |
| 7.10 | SHAPE map of CHIKV M ²⁴ mutant panel | 183 |

List of Tables

| | | |
|-----|---|-----|
| 2.1 | Next Generation Sequencing primers | 57 |
| 3.1 | Mutations introduced at M ²⁴ in CHIKV ORF-1. | 69 |
| 4.1 | Sequence observed at nucleotide 147 during passaging of M ²⁴ A CHIKV in Huh7 cells | 121 |
| 4.2 | Sequence observed at nucleotide 147 during passaging of M ²⁴ A CHIKV alternately between Huh7 and C6/36 cells | 124 |
| 4.3 | Sequence observed at nucleotide 77 during passaging of M ²⁴ A CHIKV in C6/36 cells | 128 |
| 4.4 | Summary of pseudo-escape mutants introduced into the CHIKV ICRES infectious clone | 129 |
| 7.1 | Mutations detected across the CHIKV genome following passaging of M ²⁴ in Huh7 cells | 184 |
| 7.2 | Mutations detected across the CHIKV genome following passaging of M ²⁴ alternately in Huh7 and C6/36 cells | 186 |
| 7.3 | Mutations detected across the CHIKV genome following passaging of M ²⁴ in C6/36 cells | 189 |

Abbreviations

| | |
|-------------------------------|-----------------------------------|
| AdoHyc | S-adenosyl-L-homocysteine |
| AdoMet | S-adenosylmethionine |
| ADP | Adenosine diphosphate |
| <i>Ae.</i> | <i>Aedes</i> |
| <i>Ae. spp.</i> | <i>Aedes</i> species |
| ATP | Adenosine triphosphate |
| ATPSβ | ATP synthase β subunit |
| AUD | Alphavirus unique domain |
| cDNA | Complementary DNA |
| CHIKV | Chikungunya virus |
| CMC | Carboxyl-methyl cellulose |
| CME | Clathrin-mediated endocytosis |
| CPE | Cytopathic effect |
| Cryo-EM | Cryo-electron microscopy |
| CS | Cyclisation sequence |
| CSE | Conserved sequence element |
| DAR | Downstream AUG region |
| DENV | Dengue virus |
| DEPC | Diethyl pyrocarbonate |
| DMEM | Dulbecco's modified eagle's media |
| DNA | Dioxyribonucleic acid |
| DR | Direct repeat region |
| DSH | Downstream stable hairpin |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| ECSA | East/Central/South Africa |
| EEV | Easten equine encephalitis |
| EEEV | Easten equine encephalitis virus |
| EILV | Eilat virus |

| | |
|---------------|---|
| Fluc | Firefly luciferase |
| FMV | Fort Morgan virus |
| Gluc | <i>Gaussia</i> luciferase |
| GMP | Guanosine monophosphate |
| GTase | Guanyltransferase |
| GTP | Guanosine triphosphate |
| HIV-1 | Human immunodeficiency virus -1 |
| HJV | Highmans J virus |
| Hpi | Hours post infection |
| Hpt | Hours post transfection |
| HuR | Human-antibody R |
| ICA | Infectious centre assay |
| IFIT-1 | Interferon-induced protein with tetratricopeptide repeats 1 |
| IPTG | Isopropyl β -D-1 thiogalctopyranoside |
| IRE | Iron response element |
| IRES | Internal ribosome entry site |
| LAR II | Luciferase assay reagent II |
| LB | Luria-Bertani |
| LNA | Locked nucleic acid |
| lncRNA | Long non-coding RNA |
| MAYV | Mayaro virus |
| miRNA | Micro RNA |
| mRNA | Messenger RNA |
| MOI | Multiplicity of infection |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MTase | Methyl-transferase |
| Mxra8 | Mammalian matrix remodelling associated protein |
| ncRNA | Non-coding RNA |
| NGS | Next Generation Sequencing |
| NHP | Non-human primates |
| NMR | Nuclear magnetic resonance |
| nsP | Non-structural protein |
| NTase | Nucleoside triphosphatase |

| | |
|------------------|--|
| nts | Nucleotides |
| ONNV | O'nyong-nyong virus |
| ORF | Open reading frame |
| PARS | Parallel analysis of RNA structure |
| PCR | Polymerase chain reaction |
| PDB | Protein data base |
| PHB1 | Prohibitin-1 |
| PLB | Passive lysis buffer |
| PRF | Programmed ribosomal frameshifting |
| PVDF | Polyvinylidene fluoride |
| qPCR | Quantitative PCR |
| RER | Rough endoplasmic reticulum |
| RdRp | RNA dependent RNA polymerase |
| RNA | Ribonucleic acid |
| RRV | Ross river virus |
| RT-PCR | Reverse transcription PCR |
| SDV | Sleeping disease virus |
| SESV | Southern elephant seal virus |
| SF | Semliki forest |
| SFV | Semliki forest virus |
| SHAPE | 2' hydroxyl acetylation analysed by primer extension |
| SHAPE-MaP | SHAPE and mutational profiling |
| SINV | Sindbis virus |
| siRNA | Small interfering RNA |
| SL | Stem loop |
| SNP | Single nucleotide polymorphism |
| SPDV | Salmonid pancreatic disease virus |
| SRA | Steroid receptor RNA activator |
| TCR | Terminal codon readthrough |
| tRNA | Transfer RNA |
| TROV | Trocar virus |
| USA | United States of America |
| UTF | Untransfected control |

| | |
|-------------|--------------------------------------|
| UTR | Untranslated region |
| VEE | Venezuelan equine encephalitis |
| VEEV | Venezuelan equine encephalitis virus |
| VLP | Virus-like particle |
| WT | Wild-type |
| WEE | Western equine encephalitis |
| WEEV | Western equine encephalitis virus |

Chapter 1

Introduction

1.1 Chikungunya virus

1.1.1 Classification of Alphaviruses

Chikungunya virus (CHIKV) is a re-emerging arbovirus from the *Togaviridae* family and is a member of the *Alphavirus* genus. Alphaviruses are positive sense, single-stranded enveloped viruses. There are currently over 30 characterised alphaviruses (1) that are known to infect a broad range of vertebrate hosts, including humans, horses and birds. The majority of alphaviruses are arboviruses that are spread through infection of an arthropod vector (commonly mosquitoes). However, there are examples of alphaviruses that infect the *Salmonidae* family of fish (Salmonid pancreatic disease virus (SPDV) and sleeping disease virus (SDV)) and are presumed to not be transmitted via infected arthropods. It is instead suggested that transmission of SDV occurs by contact between naïve rainbow trout and the flesh of infected rainbow trout (2). The southern elephant seal virus (SESV) has been found to infect lice, further demonstrating that not all alphaviruses are transmitted through a mosquito vector (3). Eilat virus (EILV) was first isolated in 2012 from *Anopheles coustani* mosquitoes in Israel and was the first insect-specific alphavirus discovered (4). There are currently five recognised insect-specific alphaviruses that have no known vertebrate host (5).

Alphaviruses can be classified into 'New world' and 'Old world' categories through broad geographical distribution and disease onset. New world alphaviruses are associated with infection that leads to encephalitis (Venezuelan equine encephalitis virus (VEEV) and Western and Eastern equine encephalitis virus (WEEV and EEEV respectively)). New world alphaviruses are typically endemic in the

Americas (USA, Caribbean and Central/South America). The Old world viruses are more commonly associated with the onset of fever, rash and arthralgia (such as Sindbis virus (SINV), Semliki Forest virus (SFV), Ross River virus (RRV) and CHIKV). It is hypothesised that alphaviruses originated in the New world and were introduced to the Old world by migratory birds (6). The mortality rate of New world alphaviruses can be far greater than that of Old world alphaviruses, with the mortality reaching 75 - 90% in horses infected with EEEV (7). The mortality rate in humans is much lower than that observed in animals as >96% of infections are asymptomatic; 33% of individuals who develop neurological symptoms will die and the remainder will suffer severe neurological complications (8, 9).

Alphaviruses as a genus are geographically widespread and have been detected in every continent (bar Antarctica), but individual alphaviruses generally have limited geographical distribution. For example, Mayaro virus (MAYV) has only been detected within Latin America (10) and o'nyong-nyong virus (ONNV) has not been observed beyond Africa (11). It is believed that the specificity of virus-vector interactions determines the limited geographical distribution of certain alphaviruses. It is therefore important to monitor the increased distribution of mosquitoes due to climate change and global warming, as novel cases of transmission are being reported alongside the introduction of mosquitoes to a population.

Due to recent changes and increases in the geographical distribution of alphaviruses it has become less preferable to divide them into New world and Old world viruses. They can be grouped into 10 complexes based on antigenic similarities (4) (figure 1.1). The Western equine encephalitis (WEE) complex comprises both Old and New world viruses (New world - Aura virus and Old world - Whataroa virus and Sindbis virus (SINV)), as well as recombinant viruses (Buggy Creek, Highmans J (HJV) and Fort Morgan (FMV)). The Venezuelan equine encephalitis (VEE), Eastern encephalitis (EEV) and Trocara (TROV) complexes are solely comprised of New world viruses. Conversely, the Semliki forest (SF), Barmah Forest, Ndumu and Middleburg virus complexes are made up from Old world viruses. The aquatic alphaviruses comprise two complexes; SESV and SPDV. The discovery of

insect-specific alphaviruses (such as EILV) proposed a new alphavirus complex.

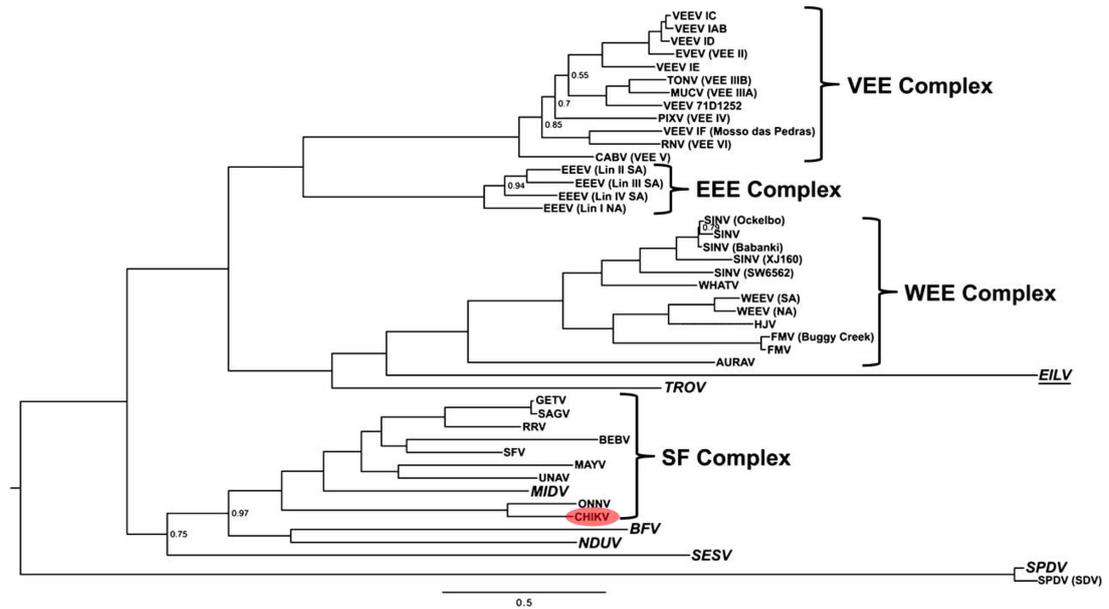


Figure 1.1: Phylogenetic tree of the *alphavirus* genus

A Bayesian phylogenetic tree based on the nucleotide sequences of ORF-2 across alphaviruses. A midpoint rooted tree is shown with all posterior probabilities <1 shown on major branches. CHIKV is highlighted in red. Virus abbreviations used as those used in Powers et al (12). Image adapted from Nasar et al (4).

1.1.2 Epidemiology of CHIKV

The first CHIKV epidemic was recorded in the Makonde plateau (an area found in modern-day Tanzania) in 1952 (13). Locals from this area began referring to the disease as *chikungunya*, translating as “that which bends up”, from the arthralgia that affects peoples ability to walk upright associated with CHIKV infection. Between 1960 - 1980 there were numerous sporadic outbreaks in West Africa (Senegal to Cameroon) and in multiple countries across Central and Southern Africa (such as; Democratic Republic of Congo, Nigeria, Angola, Uganda, Burundi, Guinea, Malawi, Central African republic and South Africa) (14, 15). African outbreaks most likely occur as spill over events from the sylvatic cycle of CHIKV between forest dwelling *Aedes* spp. mosquitoes and non-human primates.

CHIKV was first reported in Asia in 1954 in the Philippines, where there were also two epidemics in 1956 and 1986 (16). The Indian sub-continent observed its

first outbreak in 1963 in Kolkata which was followed by numerous other outbreaks across the country up until the Barsai epidemic in 1973 (14). There were no reported outbreaks across Africa and Asia from \approx 1980 - 2000.

One of the most severe CHIKV epidemic outbreaks began in Kenya in 2004, where 75% of the population in Lamu island were infected with CHIKV (15). The virus then spread across Comores, Seychelles and Mauritius islands. In 2005, CHIKV reached La Réunion where it rapidly spread across the island and to multiple other countries in the Indian Ocean. This outbreak was of particular importance due to the observation of more severe symptoms and for the first time CHIKV was associated with fatality - La Réunion observed mortality was 1/1000 cases (17, 18). By late 2006 151 districts of India had declared cases of chikungunya fever. There were no deaths reported from association with CHIKV across India. However, as with the La Réunion outbreak individuals suffered neurological complications, such as meningoencephalitis (16). Virus isolated from the La Réunion outbreak and the Indian outbreak was determined to be a strain that had diverged from the the East/Central/South Africa (ECSA) lineage (19). Cases were also noted in China and Southeast Asia (20, 21).

There had been no identified cases of CHIKV across India for over 30 years prior to the 2006 outbreak and the reduced immunity across the population may have been the cause for increased symptom severity and increased attack rate. The *Ae. albopictus* mosquito was the main vector in transmission of CHIKV across Réunion which was unusual as *Ae. aegypti* was previously understood to be the most common urban vector. This is due to female *Ae. aegypti* mosquitoes taking more blood meals than *Ae. albopictus* mosquitoes. A mutation in the E1 glycoprotein (A²²⁶V) was identified which was shown to permit increased replication rates within *Ae. albopictus* mosquitoes (22, 23), which may explain the increased attack rate of CHIKV during the Réunion and Indian epidemics as *Ae. albopictus* is more commonly seen across these areas.

There were reported cases of CHIKV across Europe and America that were imported by people travelling from epidemic regions during the Réunion outbreak. Despite the *Ae. albopictus* host being prevalent in the Americas, there were no

reports of local transmission beyond the initial cases. On the other hand, Italy reported over 200 cases of CHIKV in 2007 (24). This was the first time that CHIKV had been spread autochthonously (human to mosquito to human) in Europe. In 2010 there were reports of autochthonous transmission of CHIKV in Southern France, and sequencing and phylogenetic analysis demonstrated that the strain detected was similar to the Indian strains in the ECSA lineage (25).

The Asian lineage of CHIKV was detected in the Pacific Islands in 2011, and then spread to 10 of the 22 countries/territories in this region (26). In 2013 the first case of locally transmitted CHIKV was reported in Saint Martin, in the Caribbean Islands (27, 28). Both the *Ae. albopictus* and *Ae. aegypti* vectors are commonly found in the region from South Florida to Brazil, allowing for the development of an urban cycle and *Ae. aegypti* remained the principal vector for CHIKV transmission across the Caribbean (29). By 2015 there had been over 1 million reported cases and autochthonous transmission was observed in 45 countries across the Caribbean and North, South and Central America (27, 30).

Cases of CHIKV have now been reported in over 100 countries, across nearly all continents as shown in figure 1.2. Importantly, autochthonous transmission has been reported in temperate climates where there is no sylvatic cycle. These regions have no immunity against CHIKV and are therefore susceptible to the increased morbidity and mortality that was observed in the 2004 epidemic.

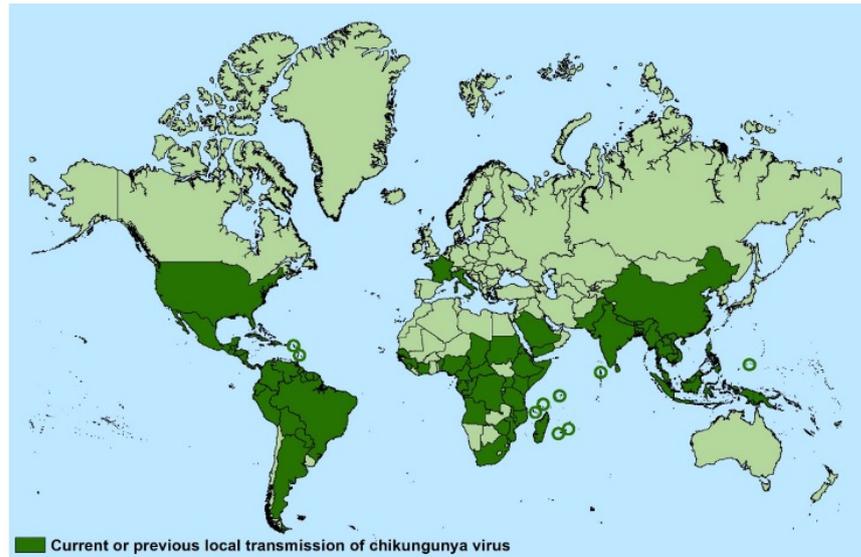


Figure 1.2: Geographical distribution of CHIKV

A map representing the countries and territories that have had reported cases of CHIKV (not including imported cases). The first case of CHIKV was reported in Tanzania in 1952 and since has spread to over 100 countries, across the Americas, Africa, Europe and Asia. Taken from CDC (31).

1.1.3 CHIKV Phylogeny

There have been three distinct lineages defined for CHIKV: the Asian, ECSA and West African lineage, with the virus responsible for the Indian Ocean outbreak forming a distinct cluster within the broader ECSA lineage (figure 1.3). These lineages roughly correspond to different outbreaks. It is of note that despite the close geographical location, the West African and ECSA lineages are genomically distinct (32).

CHIKV was isolated from a variety of different locations throughout the Indian Ocean outbreak and phylogenetic analysis of the strains identified that the A²²⁶V mutant in the E1 glycoprotein associated with increased infectivity in *Ae. albopictus* was acquired independently (23). This mutant, along with M²⁶⁹V and D²⁸⁴E in E1, were described as molecular signatures of the Indian Ocean outbreak (15). Virus sequenced from the Indian outbreak in 2006 lacked this mutant, but strains isolated from India in 2007 were observed to contain the A²²⁶V mutant. Furthermore initial CHIKV sequences isolated at the start of the La Réunion

outbreak did not contain the mutant, whereas >90% of sequences from 2005 onwards did (33). These observations demonstrate the ability of CHIKV to adapt over time. The acquisition and continued upkeep of the A²²⁶V mutant across various strains suggests that this is a concurrent pathway for the rapid infection rates of CHIKV in previously naïve locations. It also highlights the importance for continued genomic monitoring of CHIKV as a single substitution can lead to alterations in vector dynamics and increased spread of disease.



Figure 1.3: CHIKV phylogenetic tree

A phylogenetic tree of different CHIKV strains derived from the genomic sequence using Bayesian methods. Numbers indicate Bayesian posterior probability values. Strains highlighted magenta show the presence of the mutant in E1 (A²²⁶V) that allows for enhanced infection in *A. albopictus*. Image adapted from (32).

1.1.4 Transmission of CHIKV

CHIKV is transmitted through a bite from an infected female mosquito during a blood meal. There are two distinct transmission cycles: a sylvatic and urban cycle

(figure 1.4). These two systems are distinct from one another and the spillover from a sylvatic cycle poses a great threat to human health by acting as reservoir for initiating the urban cycle.

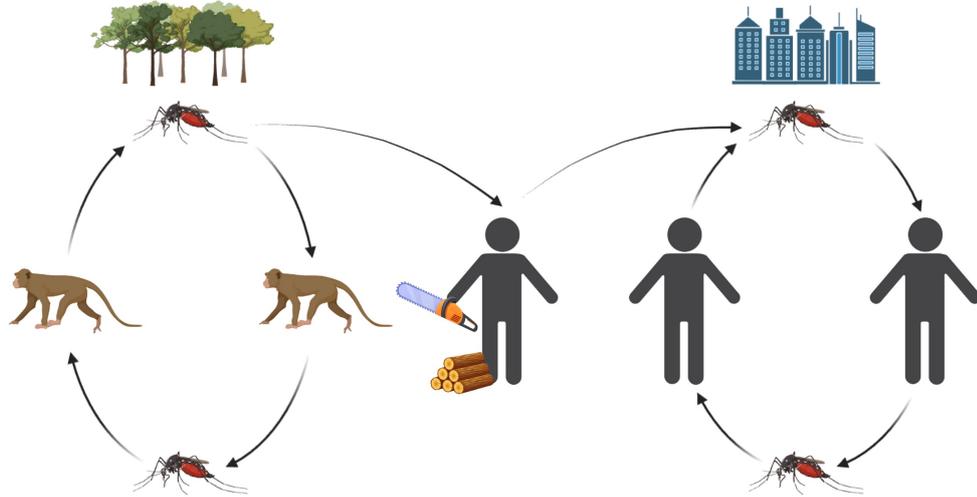


Figure 1.4: A schematic of the transmission cycle of CHIKV

CHIKV is transmitted in a sylvatic cycle between multiple *Aedes* mosquitoes in rural areas. The main reservoir host in the sylvatic cycle are non-human primates. The virus is transmitted between the mosquito vector and hosts during a bloodmeal. Spillover events occur where humans become infected with CHIKV and then travel to urban environments. CHIKV is transmitted between *Ae. albopictus* and *Ae. aegypti* mosquitoes and humans in the urban cycle.

1.1.4.1 Sylvatic cycle

The sylvatic cycle is the rural transmission of CHIKV from *Aedes spp.* mosquito vectors to non-human vertebrate hosts. CHIKV has been isolated from a variety of mosquitoes within the *Aedes* genus from a rural/forest environment, including *Ae. furcifer*, *Ae. taylori*, *Ae. africanus* and *Ae. luteocephalus*. There is an approximately 4-year amplification cycle of CHIKV in mosquitoes (34, 35), which occasionally results in spill over events into the human population (35). It was presumed that the gap between amplification events in mosquitoes was related to the time required for a sufficient amount of susceptible non-human primates (NHP) to be born. Conversely, a study reported a continuous seroprevalence of CHIKV in newly born (<1 year old) monkeys in the 3 years following an amplification event in mosquitoes. They concluded that NHP may be acting

as an amplification host for CHIKV rather than a reservoir hosts and proposed that there were other mammals acting as reservoir hosts (35). CHIKV has been successfully isolated from a variety of other mammals, such as bats and rodents (34, 36).

Whilst epidemic autochthonous CHIKV transmission occurs in the Americas and Asia, there is no evidence that sylvatic transmission cycles have been established in these areas. Moreira-Soto et al tested over 200 NHP in urban Brazil for CHIKV infection and determined that it was unlikely for CHIKV to have established a sylvatic cycle due to a low seroprevalence and lack of CHIKV detectable by RT-PCR (37). When trying to determine whether CHIKV had established a sylvatic transmission cycle in Asia, the virus was isolated from wild long-tailed macaques in Malaysia between 2007 - 2008. However, PCR and sequencing data determined that the macaques were infected with a different strain to the one that was circulating in humans within the region (38). A later study in the same location also determined the seroprevalence of CHIKV in macaques to be very low, despite an outbreak in humans occurring at the time (39). Further, a study in Thailand reported that the infection of captive macaques was independent of infection in humans (40).

1.1.4.2 Urban cycle

It is proposed that the arboreal *Ae. furcifer* vector is most likely involved in spillover events into humans as it is prevalent across a large area of land, and importantly is frequently observed within human settlements (41, 42). There are two mosquito vectors identified in the urban cycle of CHIKV: *Ae. aegypti* and *Ae. albopictus*. A person becomes infected with CHIKV when bitten by an infected mosquito (i.e *Ae. furcifer*), during the mosquito's blood meal. This person then travels to an urban environment where different species of *Aedes* mosquitoes are present. An uninfected mosquito (i.e. *Ae. albopictus*) then feeds on this (now infected) individual and the mosquito becomes infected and is able to transmit the virus onto other non-infected humans during future blood meals. A novel urban cycle can be established if an infected individual travels to a new location

and infects naïve mosquitoes when female mosquitoes have a blood meal.

It is important to continue to further investigate the spillover events from NHP in order to prevent regular resurgence of CHIKV in humans and to protect the health of individuals living in close proximity to sylvatic transmission of CHIKV. Further studies into other reservoir/amplification hosts should also be performed to predict whether novel sylvatic cycles can be established in the Americas, Asia and Europe. It is far easier to limit the spread of CHIKV when there is only an urban cycle within a region, as there is only the need to control the population of *Ae. albopictus* and *Ae. aegypti* mosquitoes, rather than attempting to limit the spread between multiple primatophilic mosquito vectors and hosts.

Furthermore, the current spillover events from established sylvatic cycles have only been from the ESCA lineage, with no cases of the West African lineage reported beyond this area. There is a potential that future spill over events could occur from the West African lineage, resulting in the generation of novel CHIKV strains and even a new lineage (similar to the evolution of the Indian Ocean lineage).

Previously CHIKV was limited to humid, tropical climates where the mosquito vectors were able to thrive. A rise in global temperatures has resulted in *Ae. aegypti* and *Ae. albopictus* mosquitoes being able to survive in areas that were previously temperate and inhospitable. *Ae. albopictus* mosquitoes are now able to survive over winter within these temperate countries, which has ultimately led to the establishment of CHIKV (and other arthropod-borne diseases) in previously naïve countries, and importantly this allows for CHIKV to be endemic in a region without the presence of a sylvatic cycle. Furthermore, *Ae. albopictus* mosquitoes have a large geographical range which in combination with the A²²⁶V mutant in E1 has increased the geographical spread of CHIKV. International trade and travel allows for vectors and infected hosts to be introduced into a new country with increased ease.

1.1.5 CHIKV infection in humans

As described above, CHIKV is transmitted to humans via infected female *Aedes* spp. mosquitoes when they take a blood meal. The typical incubation period in humans is from 1 to 12 days, with most people experiencing symptom onset between 4 - 7 days. There are three defined clinical stages following CHIKV infection; acute stage, post-acute stage and chronic stage (discussed in more detail in section 1.1.5.1).

Upon introduction to the body through the mosquito bite, CHIKV replicates in fibroblasts at the site of inoculation. Some evidence also suggests it is able to replicate in macrophages (43). CHIKV enters the bloodstream through the lymphatic system, where it is able to disseminate to multiple organs. There are high viral loads (10^9 - 10^{12} virions/mL) in the blood due to the replication of CHIKV in peripheral tissues (44, 45). This is unusual for alphaviruses compared to other arboviruses (such as dengue virus (DENV)) and allows for efficient transmission to mosquitoes during a blood meal (26, 46). The most common sites of CHIKV replication are the lymph nodes, spleen, skin, peripheral joints, muscle and tendons (where prominent symptoms are observed)(47). In more severe cases of CHIKV, the virus replicates in the brain and liver (48). It is proposed that chronic arthralgia is due to persistent replication of CHIKV in joints (49).

1.1.5.1 Symptoms of CHIKV infection

The acute phase is classed as the first 3 weeks of infection. The typical disease presentation of CHIKV is the sudden onset of fever, inflammatory arthralgia and arthritis. The extremities and peripheral joints (such as wrists and ankles) are usually swollen and painful. Other symptoms associated with CHIKV infection are headaches, myalgia, macular or maculopopular rash, edema on the face and extremities (27). It is not uncommon for individuals to suffer from asthenia and anorexia after initial symptoms have subsided. Other complications can occur during the acute phase, such as gastro-intestinal symptoms and neurological

symptoms, as well as more severe occurrence of rhabdomyolysis and bullous dermatosis which can both be fatal. There are reports of individuals suffering from respiratory and renal failures, pancreatitis and extensive epidermolysis, however, these are uncommon.

The majority of patients will recover after 1 - 2 weeks, however it is more than likely that they will relapse again. The post-acute phase is from the 21st day up to 3 months post infection. Patients who present symptoms during this stage suffer from continued severe arthritis and inflammation in joints. Patients may also present peripheral vascular disorders, neuropathy and neuropsychiatric disorders. Anti-inflammatory or pain relief drugs can be taken in order to manage the arthralgia. The percentage of patients that suffer from persistent symptoms beyond the third week varies from 50 - 90% (50). The persistence of infection within individuals is dependent on multiple factors such as genetic susceptibility, age, gender, prior musculoskeletal complications and differences in how cultures treat the pain (50, 51). It is also of note that during the post-acute stage there are records of non-specific symptoms that are not usually associated with CHIKV infection. Some of the more frequently observed manifestations include chronic fatigue, alopecia and changes in skin colour (50, 52).

The third stage of infection is the chronic stage, which is when symptoms persist for over 3 months. 80-93% of infected individuals within the La Réunion outbreak suffered chronic disease, with 47% of these suffering with symptoms for over 2 years following initial infection (27). The presentation of symptoms in the chronic stage differs between individuals, with some suffering bilateral arthralgia and others symmetrical; it can also be intermittent or constant pain (50). The effects of chronic CHIKV are severe and reduce the patients quality of life for several years following the initial infection with individuals claiming they are unable to lift heavy objects or walk 3 years after the La Réunion epidemic (53).

1.1.6 Alphavirus infection in *Aedes* spp. mosquitoes

CHIKV is known to infect a range of *Aedes* spp. mosquitoes across the sylvatic and urban transmission cycles. *Aedes* spp. mosquitoes are the biological vector for CHIKV and infection is typically asymptomatic and lifelong, in contrast to the acute pathogenic infection associated with humans. A female mosquito ingests CHIKV during a bloodmeal on an infected host and the midgut epithelial cells are the primary location for CHIKV replication (54). The virus passes through the basal lamina from the midgut into the haemolymph, where it is able to disseminate to other organs in the mosquito through circulation in the haemocoel. Replication of CHIKV in the salivary glands is essential for successful transmission to naïve hosts during a bloodmeal. It was generally thought that alphavirus infection in mosquitoes was benign. However, there have been studies performed that demonstrate the importance of the mosquito siRNA response in reducing viral infection and dissemination in mosquitoes (55, 56). This allows for both the virus and mosquito to survive for long enough for further viral transmission (57).

1.1.7 Therapeutics and preventative measures against CHIKV

Whilst there are currently no licensed vaccines or antiviral therapeutics against CHIKV there have been several candidates progressed through to Phase 3 clinical trials with promising results so far. There has been a vaccine developed by a team in the USA which is a CHIKV virus-like particle (VLP). The VLPs were generated in mammalian cells and contain the structural proteins but do not contain the CHIKV genome or replication machinery. This novel vaccine has been tested in Phase 1 and Phase 2 trials, where it has been successfully shown that vaccine is safe in humans and able to elicit an immune response against CHIKV that was still present 72 weeks after the vaccine administration (58).

In March 2022 the company Valneva announced the success of the Phase 3

trial of their CHIKV vaccine (59). The vaccine is a live-attenuated, single shot vaccine, which was generated through the deletion of a 60 aa region in the nsP3 hypervariable domain (Prof Andres Merits, personal communication). The trial involved over 4,000 participants, aged 18 years and above, across USA, with 96.3% of participants maintaining CHIKV neutralising antibodies 6 months post vaccination. Valneva are now working towards gaining full approval from the U.S Food and Drug Administration to allow for the vaccine to become fully licensed and available.

Whilst waiting for the Valneva vaccine to become fully licensed, the best preventive measure against CHIKV infection is to try and limit individuals being fed on by infected mosquitoes. This can be done through efforts to control mosquito populations, such as reducing the number of habitats that allow for breeding of mosquitoes (whether they are natural or artificial). This is dependent on communities working together to empty and clean water containers on a regular basis, but if sustained this is an effective way at reducing the mosquito vector population. Health authorities can also be involved in using insecticides to kill mosquitoes and their larvae if there is a known outbreak. Individuals can also apply mosquito repellents and sleep under mosquito nets to avoid mosquito bites. The rapid introduction of measures to control the *Ae. albopictus* population in Italy and France was successful in limiting the spread of CHIKV during the 2007 and 2010 outbreaks (24, 25).

1.2 CHIKV Molecular Biology

1.2.1 CHIKV Virion Structure

CHIKV is a small, enveloped, single-stranded, positive-sense RNA virus, approximately 60-70 nm in diameter. The alphavirus virion is composed of three, concentric layers; the genome is surrounded by 240 copies of the nucleocapsid core protein which are arranged as an icosahedral, a host-derived phospholipid membrane and the glycoprotein heterodimers which form 80 trimers (figure 1.5A).

The core proteins and glycoproteins form a $T = 4$ symmetrical lattice, which is unique to alphaviruses (60).

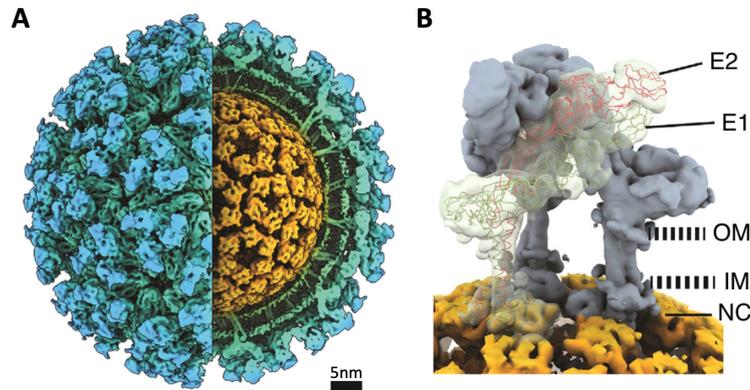


Figure 1.5: Alphavirus virion structure

A A cutaway isosurface representation of the alphavirus virion structure, determined by cryo-EM and 3D reconstruction. The nucleocapsid core is highlighted in yellow, the host derived lipid bilayer is highlighted in green, and the glycoprotein spikes are highlighted blue. **B** An isolated view of the trimeric spikes formed by heterodimers of E1 and E2. Two heterodimers are displayed in grey, and one is shown as an atomic trace (E1 in green and E2 in red). OM denotes outer membrane, IM denotes inner membrane and NC denotes nucleocapsid. Figure adapted from Button et al, 2020 (60).

The E2 glycoprotein contains a cytoplasmic domain that has been shown to be embedded in the lipid bilayer and interacts with the core protein (figure 1.5B). Early structures of the alphavirus virion (using SINV and SFV as models) suggested that there is one-to-one pairing between the 240 capsid and E2 proteins, helping to facilitate in the symmetry between the two layers (61, 62). It is of note however, that more recent publications have shown that the core protein is not essential for the assembly and egress of infectious CHIKV particles, creating a viable and potential live-attenuated vaccine (63).

1.2.2 CHIKV Genome Organisation

The CHIKV genome is approximately 11.8 kb in length, consisting of a 5' and 3' untranslated region (UTR) flanking 2 open-reading frames (ORF) that are separated by an intergenic UTR (figure 1.6). The importance of the UTRs are discussed in detail in sections 1.3.5.1 and 1.3.5.3. The genome has a 5' type-0

7-methylguanosine cap and is polyadenylated at the 3' end, similarly to host mRNA. The first ORF encodes the non-structural proteins 1-4 (nsP1-4) and is translated from the full-length genomic RNA strand (49S). ORF-2 encodes the structural proteins (capsid, E1-3 and 6K) and is translated from a sub-genomic mRNA (26S).

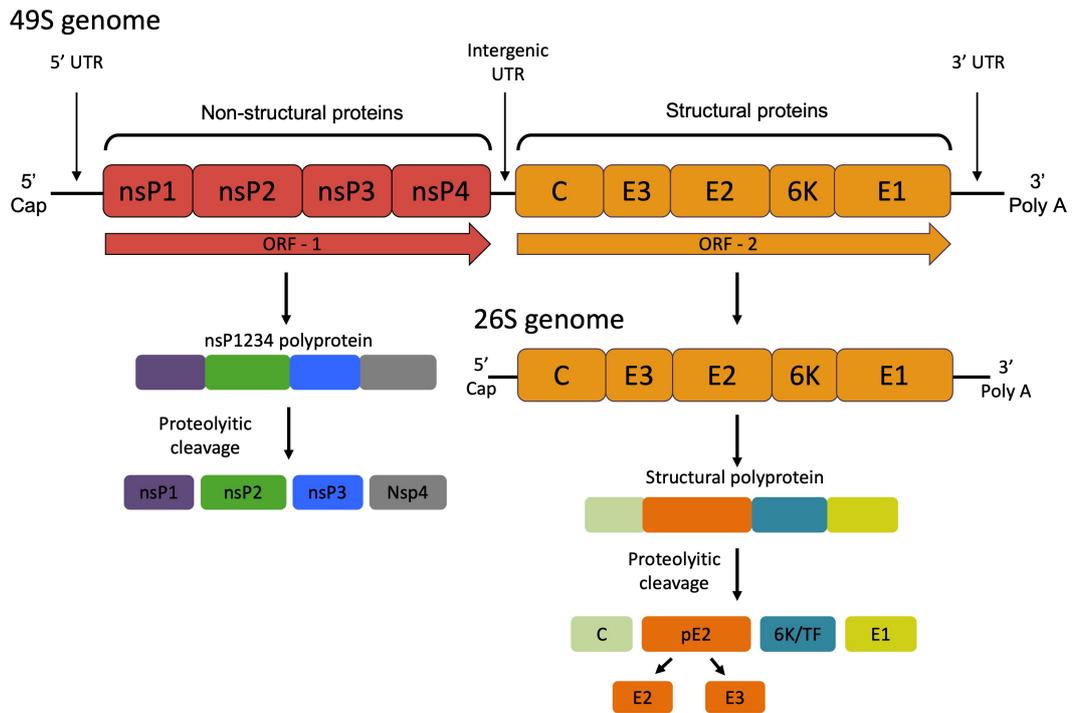


Figure 1.6: CHIKV genome schematic

The CHIKV genome is 11.8 kb. There are two ORF; ORF-1 encodes the non structural proteins. ORF-2 encodes the structural proteins. There are three untranslated region (at the 5' and 3' ends and between ORF-1 and 2).

1.2.3 Function of non-structural proteins

1.2.3.1 nsP1

The functions of CHIKV nsP1 are discussed in detail in section 3.1.1. Briefly, nsP1 is a 58 kDa protein whose primary function is the addition of the 5' cap onto nascent CHIKV genomic RNA (49S and 26S) (64, 65). This function is essential to ensure efficient and successful viral RNA translation. Importantly, the capping mechanism of alphavirus nsP1 is different to that of the host pathway. There is a

membrane binding domain within nsP1, that is conserved across the *alphavirus* genus, which is important in the anchoring of the replication complex to the plasma membrane, forming spherules where RNA replication occurs (66–69). nsP1 also contains a palmitoylation site that has been shown to be important in the formation of filopodial extensions in both mosquito and mammalian cells (70, 71). Finally, there have been studies performed that underline the importance of interactions between nsP1 and nsP4 for efficient negative strand synthesis (72). Following the submission of this thesis Tan et al., reported the molecular architecture of the replicase complex and the interactions between nsP1, nsP2 and nsP4 (73).

1.2.3.2 nsP2

nsP2 is the largest of the four non-structural proteins (≈ 90 kDa) and has multiple roles across the CHIKV replication cycle. There are 3 well characterised domains and 2 putative domains within nsP2: the two putative domains are located at the N-terminus, followed by the NTPase (nucleoside triphosphatase) domain, the protease domain and an MTase (methyltransferase) like domain at the C-terminus. It is believed that this MTase like domain does not exhibit enzymatic activity due to the lack of residues associated with traditional MTase enzymatic activity (74).

There are three catalytic activities performed by nsP2. The helicase activity of nsP2 was initially determined through bioinformatics analysis (75) and later confirmed through mutational and recombinant approaches (76). The helicase activity of nsP2 is dependent on the NTPase domain and nsP4. Rikonen et al., introduced mutations into the NTPase Walker A motif and noted a reduction in both NTPase and helicase activity (77). More recent studies have demonstrated that full length CHIKV nsP2 is essential for 5' - 3' helicase activity, NTPase activity and RNA rewinding, with reduction in activity observed when both the N- and C-terminal are truncated (78). During the addition of the 5' cap on nascent RNA, the NTPase domain of nsP2 is responsible for the removal of γ -phosphate from 5' RNA to produce diphosphate-RNA, which can then be used

as a substrate for the addition of a 5' m⁷-cap by nsP1 (figure 3.1).

The third catalytic function of nsP2 is the cleavage of the nsP1234 polyprotein by the C-terminal protease domain (79, 80). This process is essential for the production of individual non-structural proteins and replication of alphaviruses genome. Multiple studies of SINV and SFV nsP2 have demonstrated the ability of the protein to cleave the polyprotein both in *cis* and in *trans*, and it is proposed that the favouring between *cis* and *trans* cleavage provides temporal control to different stages of the virus life cycle (74). Generally, *cis* cleavage of the nsP1234 polyprotein occurs during early stages of replication when there is a low concentration of non-structural proteins. As the concentration of protease increases over time, *trans* cleavage is favoured (81). The preference of the protease to cleave the polyprotein alters between the different protein junctions in such a way to facilitate the preferential *cis* cleavage between nsP3 and nsP4 first, in order to create the nsP123+nsP4 early replication complex (82). For example, if the polyprotein contains nsP1 the protease activity between the nsP2/3 junction is poor (79). The increased concentration of protease that is noted during later stages of replication is associated with favoured *trans* cleavage of the nsP2/3 junction (83). Therefore, in the later stages of replication the polyprotein is observed to be cleaved into P12 and P34.

Furthermore, in addition to the catalytic activity of nsP2 there is evidence to suggest that the protein is important in the transcription of sub-genomic RNA, through binding to the sub-genomic promoter and acting as a transcription factor (74). There is a temperature sensitive mutant that is associated with reduced sub-genomic RNA synthesis activity, and further investigation of this mutant mapped it to nsP2, suggesting that this protein is involved in 26S RNA synthesis (84). It is of note that direct binding between nsP2 and the sub-genomic promoter has not been demonstrated, and it is known that the temperature sensitive mutant reduces nsP2 protease activity. It is therefore possible that the reduction in sub-genomic RNA synthesis was due to a knock-on effect of the reduced cleavage into individual non-structural proteins. However, reduction in sub-genomic RNA was still observed, even after the complete cleavage and assembly of the replicase complex (85).

1.2.3.3 nsP3

nsP3 is a ≈ 60 kDa protein with three distinct domains: the macrodomain (at the N terminal), the hypervariable domain (at the C terminal) and the alphavirus unique domain (AUD, in the center). nsP3 is critical for the efficient replication of alphaviruses within mammalian and mosquito cells and has been associated with a variety of activities/roles, but the specific function of the protein is yet to be fully characterised.

The hypervariable domain is hyperphosphorylated and involved in multiple host cell interactions. Fros et al., reported that CHIKV nsP3 interacts with the host G3BP protein (mammalian) and rasputin (mosquito) to prevent assembly of stress granules during viral infection (86, 87); suggesting a role for nsP3 in the regulation of the host cellular stress response following infection. Later studies determined that depletion of host G3BP proteins severely reduced negative strand synthesis of CHIKV (88). Interactions between nsP3 and Hsp90 have been shown to be important for CHIKV replication, and knockdown of Hsp90 resulted in reduced inflammation in mice (89). These various interactions underline the understanding that the hypervariable domain plays a part in the permissivity of cells to CHIKV.

The AUD is homologous across the alphaviruses and contains a highly conserved zinc binding region of 4 cysteines. Mutagenesis of this area has shown the importance of nsP3 in early stages of the virus lifecycle (83). Another study introduced mutations throughout the AUD and highlighted the importance of nsP3 in the transcription of sub-genomic RNA, with downstream negative effects in the translation of the structural proteins and virus assembly (90).

The macrodomain is conserved across the alphaviruses and is known to contain a binding pocket for ADP-ribose. ADP-ribosylation is a reversible post-translational modification that is important across a variety of cellular mechanisms/processes and modification of enzymes can inactivate them. Cellular proteins modify nsP2 by mono-ADP-ribosylation, which inhibits the protease activity of the protein, and therefore ablates virus replication through the lack of mature non-structural

proteins. Krieg et al. reported that the macrodomain of nsP3 is able to hydrolyse and remove the ADP-ribose on modified nsP2 to ensure that correct polyprotein processing occurs (91). This suggests that nsP3 may have an important role in evading the host immune response, however it is important for further studies to be performed to gain a deeper understanding about this relationship. Finally, correct folding and structure of the alphavirus nsP3 domain has been shown to be vital for specific cleavage of the 2/3 site in the non-structural polyprotein (92).

1.2.3.4 nsP4

nsP4 is the RNA-dependent-RNA-polymerase (RdRp) and is therefore responsible for the synthesis of nascent negative- and positive-sense viral RNA (both full-length and sub-genomic). It is the second biggest non-structural protein at ≈ 70 kDa and ≈ 600 aa in length. The first 100 aa are unique to alphaviruses, with suspected roles in polyadenylation (although this process is still not fully understood) (93) and interactions with the other non-structural proteins. The remaining 500 aa contain the typical fingers, palm and thumb domains seen in RdRps. The active site is characterised by a GDD motif within the palm domain (74) and replicase activity can be abolished by introducing mutations (such as GAA) at this site. The amino acid residues involved with binding of the sub-genomic promoter and genomic promoter have been identified (331 - 332 and 531 - 538 respectively) and mutations at these residues inhibit nsP4-promoter binding (94).

Purified recombinant nsP4 requires the presence of the other non-structural proteins, so whilst nsP4 is solely responsible for RNA synthesis, the whole replication complex is important in the synthesis of nascent RNA. The unique region at the N-terminal of nsP4 is predicted to be disordered and important in the interactions with the other three non-structural proteins. Mutations were introduced across this region in SINV which reduced overall levels of viral RNA synthesis. Following infection and passaging with the mutant nsP4, the genome was analysed for any compensatory mutations and all 3 non-structural proteins contained mutants that restored viral RNA synthesis (95, 96), strongly suggesting

that the N-terminal of nsP4 regulates protein interactions within the replication complex.

1.2.4 Alphavirus replication

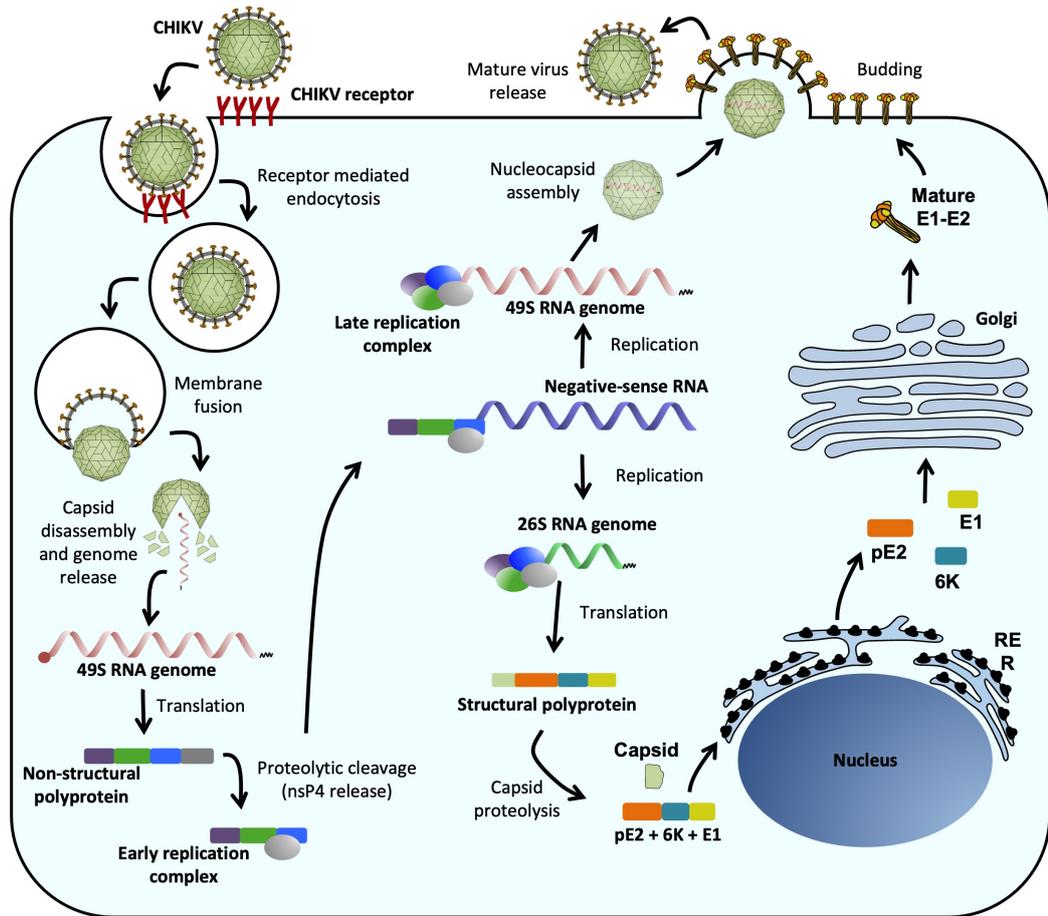


Figure 1.7: Alphavirus replication in mammalian cells

CHIKV binds to receptors on the host cell surface and the virus enters the cell through endocytosis. Within the endosome, the viral and host membranes fuse allowing the release of the nucleocapsid into the cytoplasm, where it disassembles and the genome is released. The non-structural proteins are translated and then a autoproteolytic cleavage event of nsP4 allows for the formation of the early replication complex (nsP4+nsP123). Negative-sense RNA is synthesised and is subsequently used as a template for replication of the positive-strand RNA and the sub-genomic (26S) RNA. The structural proteins are translated from the sub-genomic RNA into a polyprotein and autoproteolysis itself to release the capsid protein. The pE2+6K+E1 polyprotein enters the rough endoplasmic reticulum (RER) and ultimately the Golgi for post-translational processing and maturation. The capsid protein and newly synthesised RNA interact with each other to form the nucleocapsid, which is transported to the plasma membrane where it interacts with the mature glycoproteins. Budding of the host cell membrane occurs, allowing for release of mature virions (27).

1.2.4.1 Cell entry

There have been several different receptors and attachment factors associated with the alphaviruses over the years (97). Heparan sulphate, C-type lectins and phosphatidylserine have been suggested as alphavirus attachment factors; they are involved with the initial contact the virus makes with the cell prior to the virus interacting directly with the cell receptor to enter the host cell. An attachment factor is not classified as a cell receptor as it does not incite internalisation of the virus and an antibody can not be used to prevent the virus-receptor binding to inhibit infection.

The mammalian matrix remodelling associated protein (Mxra8) has been identified as a putative receptor for CHIKV through CRISPR-Cas9 screening experiments (98). This receptor was also associated with other arthritogenic alphaviruses such as ONNV, RRV and MAYV and is typically found on the surface of cells that are infected by these viruses (i.e. myeloid, epithelial and mesenchymal cells). Gene editing was carried out on mouse and human cells to deplete expression of Mxra8 on the cell surface and as a result a decrease in viral infection was observed, alternately, when the protein was ectopically expressed, viral infection increased. Further confirmation was achieved through structural imaging of Mxra8 in complex with the CHIKV virion (99) and the E1/E2 glycoproteins (100). It is of note that Zhang et al., still observed low levels of CHIKV infection in cells that lacked the Mxra8 protein, suggesting that there is a co-receptor that is associated with alphavirus fusion and cell entry (98).

Two other receptors have been proposed for CHIKV cell fusion and entry in mammalian cells; prohibitin-1 (PHB1) (101) and CD147 (102). Initial proteomics analysis identified PHB1 as a potential receptor for CHIKV. The CHIKV-PHB1 interaction was supported through co-localisation at the plasma membrane by immunofluorescence microscopy, reduction of viral infection following treatment of cells with anti-PHB1 antibody and decreased viral infection when the PHB1 gene was silenced by siRNA. Secondly, a recent study reported significantly reduced viral infection levels in knockdown cell lines not expressing CD147, which could be recovered through re-introduction of CD147 recovered infection levels.

Interestingly, CD147 shares a similar tertiary structure with Mxra8, despite no amino acid sequence homology between them. This proposes that CHIKV does not bind to a specific protein, but rather to a specific structure, to mediate virus fusion and cell entry. This supports the potential for a co-receptor as mentioned above.

The discussed receptors are not present on mosquito cells and instead the ATP synthase β subunit (ATPS β) has been characterised as a receptor for CHIKV entry into mosquito cells (103). Similarly to PHB1 and CD147, ATPS β was initially identified as a receptor for CHIKV through proteomics analysis, and confirmed through the observation of reduced viral infection following treatment of cells with anti-ATPS β antibody and siRNA knockdown of ATPS β in cells.

1.2.4.2 Uncoating and genome release

It is widely accepted that alphaviruses utilise clathrin-mediated endocytosis (CME) for cellular entry (104). Lee et al., were able to visualise CHIKV within invaginations on the plasma membrane of mosquito cells, that resembled clathrin-coated vesicles (105). They went on to report a significant reduction in virus entry following inhibition of CME by a variety of different inhibitors and treatment of C6/36 (*Ae. albopictus*) cells with siRNA targeting clathrin. A genome wide siRNA screen further demonstrated the importance of the clathrin-dependent pathway for internalisation of alphaviruses in human cells (106). In addition, live tracking of single CHIKV virions demonstrated that CHIKV enters the cells through CME (107). However, there are also studies suggesting that a clathrin-independent mechanism is used for alphavirus entry. There was no observed effect on CHIKV infection in HEK293T cells following treatment with siRNA against clathrin heavy chains (108), and instead the importance of Eps15 was reported. It is of note however, that Eps15 is associated with both clathrin-dependent and clathrin-independent endocytosis. The majority of research points towards a clathrin-dependent pathway for CHIKV entry, but the possibility of a clathrin-independent mechanism should not be ignored.

Regardless of the pathway used, the reduction of pH is essential for nucleocapsid

release into the cytoplasm. The acidification of the endosome causes the heterodimer of E1-E2 to dissociate and a conformational change within E1, exposing the fusion loop. This allows for fusion between the endosomal membrane and viral envelope to fuse together and the nucleocapsid is released into the cytoplasm. The pH required for fusion differs across the alphaviruses, for example it is understood that SFV fusion occurs in early endosomes (pH 6.2) whereas SINV requires a lower pH (pH 5.6) which is typical to that of the late endosome (104). There is also contradictory evidence for the importance of early and late endosomes for CHIKV fusion and nucleocapsid release through knockdown of the Rab5 (early) and Rab7 (late) endosomal proteins. Bernard et al., (2010) reported that Rab5, and not Rab7, was important for CHIKV entry in mammalian cells suggesting that fusion occurs in the early endosomes. Alternately, Lee et al., reported the importance for both Rab5 and Rab7 for CHIKV entry into mosquito cells, concluding that trafficking of CHIKV through early to late endosomes is required for efficient CHIKV entry (105).

1.2.4.3 Genome replication

Host ribosomes are known to interact with the capsid proteins to initiate nucleocapsid uncoating and release of the viral genome into the cytoplasm (109, 110). More recent studies have highlighted the importance of interactions between the viral RNA and capsid protein for efficient initial translation through direct interactions of the ribosome:RNA:capsid complex (111). The non-structural proteins are immediately translated as a polyprotein (nsP1234) from the released viral RNA. There is an opal stop codon in nsP3 in the majority of CHIKV strains and termination codon readthrough (TCR) of this alters the translation of nsP123 or full length nsP1234 (112). The translation of the full-length polyprotein is less frequent than nsP123, resulting in a reduced concentration of nsP4 in comparison to the other non-structural proteins.

The nsP1234 polyprotein undergoes auto-proteolytic cleavage by nsP2 to release nsP4, forming the early-replication complex nsP123 + nsP4. The early replication complex is anchored to the host membrane by nsP1, resulting in the formation of

spherules (113). Synthesis of the full-length negative sense RNA intermediate occurs within these spherules. Interactions between nsP4 and the other non-structural proteins is essential for negative strand synthesis (95). It has been demonstrated that the early replication complex is also capable of positive sense RNA synthesis in CHIKV (114).

The cleavage of nsP123 into individual proteins marks the change from synthesis of negative-sense RNA to positive-sense RNA (96, 115). All four non-structural proteins interact with one another to form the late replication complex responsible for synthesis of the genomic (49S) and sub-genomic (26S) positive-sense RNA. The minus strand molecule acts as the template for both full-length, and sub-genomic RNA synthesis and it is of note that the 26S RNA is synthesised in excess of the 49S, suggesting that synthesis of the structural proteins is favoured (74).

1.2.4.4 Structural protein synthesis

The structural polyprotein is translated from the sub-genomic RNA and then auto-proteolytically cleaved to release the capsid protein from the pE2-6K-E1 polyprotein (116). The capsid protein is immediately used for encapsidation of nascently transcribed 49S genomic RNA (117). A signal sequence in E3 results in the translocation of the polyprotein to the endoplasmic reticulum, where host signalases cleave at both the N- and C-terminus of 6K to release pE2, 6K and E1 (117). pE2 and E1 form heterotrimers within the Golgi complex, and formation of these is essential for correct folding (118). These heterotrimers form premature viral “spikes” which are further modified in the *trans*-Golgi system, whereby pE2 is cleaved by host-derived furin into the mature E2 and E3 glycoproteins (119).

1.2.4.5 Virus assembly and egress

The glycoproteins are transported to the plasma membrane via the secretory pathway, where the mature E2 and E1 heterodimers are inserted into the membrane. The capsid protein interacts with nascent genomic RNA to form the nucleocapsid, which is also trafficked to the plasma membrane. The capsid protein interacts with

the phosphorylated tail of E1 to form a complete virus particle. In mammalian cells, newly formed virus particles bud from the plasma membrane. Alternately, in mosquito cells budding also occurs in the lumen of cytopathic vacuoles (120). The E3 glycoprotein is associated with immature virions and is released following egress of the virion from the host cell and is hence found in the extracellular fluid but not in purified virus particles (121). Interestingly, alphaviruses have been shown to infect naïve cells via direct transport of virions from infected cells. This is believed to aid in the development of persistent infections due to the bypassing of extracellular medium during virus dissemination (120, 122).

1.3 RNA structures

1.3.1 Formation of RNA structures

Unlike DNA which is synthesised as two strands to form a double helix, RNA is synthesised as a single stranded molecule. This allows for the formation of base pairs between nucleotides to generate a multitude of possible structures. Similarly to protein, the folding of RNA is hierarchical and is determined by 1) secondary: base pairing between complementary sequences and 2) tertiary: interactions between long-range and distant sequences. Base pairing of RNA can occur with the typical Watson-Crick pairing (A-U, G-C), as well as non Watson-Crick pairing. The most common non Watson-Crick base pairing is G-U, often referred to as a wobble base pair, and is frequently observed in functional RNA across all three phylogenic domains demonstrating their importance across all aspects of life (123).

One of the most prevalent secondary structures within RNA is a stem loop (also referred to as a hairpin loop). A stem loop forms when the single-stranded RNA folds back on itself, creating a double-stranded stem which usually has a loop of varying size at its apex (figure 1.8) (124). Bulges occur when one or more nucleotides (on either one strand, or both) are un-paired resulting in a break within the double-stranded stem, which then reforms again after the bulge. A

multi-branch hairpin structure can occur when a stem loop starts from a bulge, resulting in the possibility of a wide variety of complex secondary structures. Long-range interactions can occur between complementary sequences at the 5' and 3' ends of RNA, resulting in circularisation of the RNA.

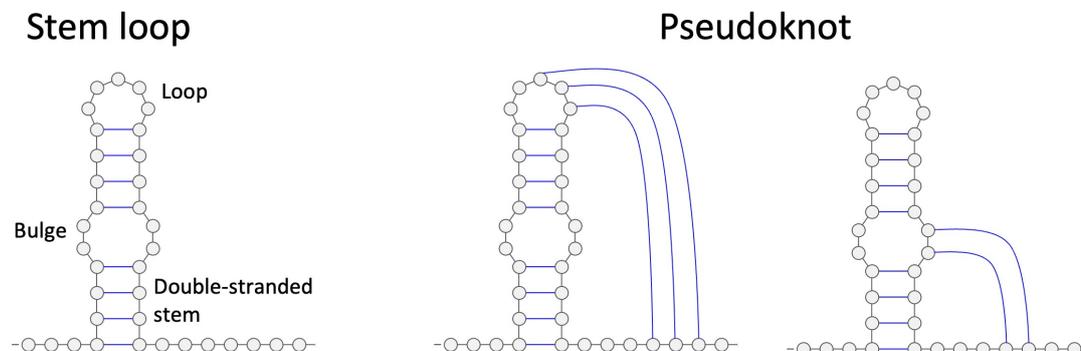


Figure 1.8: Schematic representation of RNA structures

Secondary structure shown as a stem loop, with single-stranded bulge and terminal loop. Tertiary structure shown as a pseudoknot with base pairing occurring from either the bulge, or terminal loop.

The presence of tertiary structures increase the complexity of RNA structures. Interactions can occur between two double-stranded helix regions, distant single-stranded sequences or a double-stranded helix and single-stranded sequence. Pseudoknots are formed when the exposed nucleotides in a stemloop interact with a complementary sequence elsewhere in the genome (as shown in figure 1.8), and these interactions can occur in either a 5' or 3' direction. Pseudoknots are known to be vital components of RNA structure and a key aspect in many cellular functions, such as regulation of gene expression (125). The combination of different secondary and tertiary structures allows a long RNA strand to form a small, compact molecule. Examples of these are discussed in more detail in section 1.3.4.

1.3.2 Predicting and visualising RNA structures

There are a wide variety of techniques that can be used to predict or visualise RNA structures. Initial analysis of RNA secondary structures can be performed through computational analysis, using the RNA sequence alone. The algorithms

are able to determine which nucleotide base pairings require the lowest amount of energy to occur (and therefore the most energy favourable) to generate minimum free energy models. These algorithms are limited as they require very little input information and is therefore unable to take into account external factors such as RNA:RNA interactions, and protein:RNA interactions, and cellular conditions (e.g. temperature, salt concentrations and pH levels). It is therefore important to use them alongside biochemical or structural mapping approaches.

Biochemical analysis is commonly used and provides the ability to closely examine interactions between individual nucleotides and model potential RNA secondary structures. This analysis can be performed on *in-vitro* transcribed RNA or *in-vivo* cellular RNA. Initial methods involved radiolabelling non base-paired nucleotides whereas more modern techniques utilise chemical probes or nuclease activity. Parallel analysis of RNA structure (PARS) utilises two nucleases; V1 RNase (preferentially cleaves the phosphate RNA backbone of double-stranded RNA) and S1 RNase (preferentially cleaves single stranded RNA) (126). The cleaved RNA is reverse transcribed to generate a cDNA library, which is then analysed by Next Generation Sequencing (NGS).

An alternative method is selective 2' hydroxyl acetylation analysed by primer extension (SHAPE), which involves the incorporation of a chemical probe. There are a multitude of probes that can be used, such as NMIA, DMS or 1M7, which all preferentially react with the 2'-hydroxyl groups of flexible (unpaired) nucleotides, forming a 2'-O-adduct. Similarly to PARS, the RNA is subjected to reverse transcription to generate a cDNA library which can then be analysed through NGS or capillary electrophoresis. A more in-depth description of this method is provided in section 4.2.1.

Techniques to determine RNA tertiary structures at an atomic resolution include crystallography, nuclear magnetic resonance (NMR), and cryo-EM. NMR is more easily performed on small sequences and as such, the majority of RNA structures uploaded onto the protein data bank (PDB) resolved through NMR are 20-50 nts (127). X-ray crystallography is more commonly used to visualise larger RNA structures but the flexibility of RNA structures mean that they can be difficult

to crystallise. Recent advances in image processing algorithms have led to a large increase in the number of structures resolved through cryo-EM (128). However, cryo-EM is mainly applied to the determine protein-RNA complexes, and there are only a handful of RNA only structures uploaded to the PDB. As with X-ray crystallography, the high flexibility of RNA structures is one of the main challenges when trying to obtain a high-resolution of RNA through cryo-EM. Furthermore, molecules smaller than 50 kDa are difficult to image through cryo-EM but recent breakthroughs in technology have led to the resolution of RNA structures below 40 kDa (\approx 119 nts) (129).

1.3.3 Importance of RNA structures

It was originally thought that RNA was merely an intermediary between genomic DNA and protein synthesis. It is now broadly understood that RNA is important in many aspects of cellular biology, such as translation regulation, splicing and localisation. The majority of RNA generated within a cell does not encode for protein, and is referred to as non-coding RNA (ncRNA). These ncRNAs are classified into groups roughly depending on their size, short ncRNAs are $<$ 200 nts, and long ncRNAs (lncRNA) are $>$ 200 nts. The dysregulation of lncRNAs has downstream effects on cellular functions, such as cell proliferation and evasion of apoptosis. Inevitably either the up- or down-regulation of lncRNAs have been associated with an array of cancers (prostate, lung, liver and breast cancer) (130).

Despite the association with multiple cellular functions and cancer progression, lncRNAs remain to be largely uncharacterised, and in particular the understanding of lncRNA structure is limited. There is a low cellular expression level of lncRNAs in comparison to other RNAs, making it difficult to obtain a sufficient quantity of material for structural analysis (131). The primary sequence of lncRNAs displays a large amount of diversity (132). Interestingly, despite diversity in the primary sequence the secondary and tertiary structures are often highly conserved. The conservation of structures regardless of sequence diversity is likely achieved through the integration of compensatory mutations elsewhere in the genome (132). This phenomenon is less likely to occur within protein-coding regions

as the primary nucleotide sequence is important for the ultimate amino acid sequence in the protein. A combination of chemical and enzymatic techniques were utilised to determine the structure for the full length steroid receptor RNA activator (SRA) lncRNA (870 nts) (133). Studies have demonstrated that SRA directly binds to nuclear receptor coactivator proteins, RNA helicases and two pseudouridylases (Pus1p and Pus3p) (134), which supports studies demonstrating SRA roles in nuclear co-activation of hormone-related pathways (including the oestrogen receptor) (135, 136). Mutations were introduced to disrupt the secondary structures and a reduction in SRA activity by 40% was reported, supporting the hypothesis that the structure of lncRNAs are important in their function and interactions with cellular proteins (133).

As well as ncRNAs, the presence of structures in the UTRs of mRNA (at both the 5' and 3') are crucial for the regulation of protein expression (137, 138). Translation is initiated at the 5' end of mRNA, and mRNAs with long UTRs require the ribosome to scan along the UTR until it finds the translation initiation site (139). The presence of structures within the 5' UTR can inhibit the entry of the 43S preinitiation complex. The iron response element (IRE) has been well studied and comprises a 30 nt stem loop in the 5' UTR of mRNAs involved with the transport/storage of iron. In low iron conditions, the IRE structure is bound to iron-regulator proteins (140), this interaction inhibits the binding of the 43S complex and therefore inhibits translation. When iron levels are increased, the proteins no longer bind to the RNA and translation is effectively initiated to allow for increased synthesis of proteins to transport/store iron (141).

Finally, structures within the 3' are understood to be important for RNA stability. β -Thalassemia is caused by a mutation of the stop codon in the β -globin gene, reducing haemoglobin expression. Due to this mutation, the ribosome continues into the 3' UTR and 'masks' the RNA secondary structures that would usually aid in mRNA stability. With these structures 'masked', the mRNA is no longer stable and is degraded, therefore decreasing the synthesis and levels of β -globin, causing the onset of disease symptoms (142).

1.3.4 Importance of viral RNA structures

As well as cellular mRNA UTRs being highly structured, it has long been understood that viral RNAs also possess UTRs that are structure rich. As with cellular mRNAs (143), the UTRs of viral RNAs can vary considerably in size, ranging from a few nucleotides to > 1kb in length (144, 145). One of the best studied viral UTR is that of human immunodeficiency virus-1 (HIV-1), which has a highly structured 5' UTR. One structure in particular is referred to as TAR and interacts with both viral and host proteins, the introduction of mutations in this region severely inhibit the translation and synthesis of HIV-1 proteins (146).

There are some viruses which do not have a 5' cap structure added post RNA transcription, which makes them more susceptible to 5' endonuclease degradation by host proteins. The 5' cap structure is also crucial for the initiation of translation. The host protein eIF4E recognises and binds to the 5' cap (147) and signals for the recruitment of the other eukaryotic translation machinery to the mRNA molecule to ultimately allow ribosome scanning and translation (148, 149). Uncapped viruses can clearly not initiate translation through this mechanism and therefore initiate translation in a cap-independent manner. There is a collection of highly structured RNA stem loops in viral 5' UTRs (figure 1.3.4A) which were first observed in the *Picornoviridae* family, known as internal ribosome entry sites (IRES), and interestingly they have recently been discovered in cellular mRNAs (150), conflicting the understanding that all cellular mRNAs are capped (151, 152). The IRES element in poliovirus was discovered to recruit the ribosomal 40S subunit with a reduced set of eIFs (153), and disregarded the requirement of eIF4E for cap recognition, and translation initiation. This discovery led to the understanding that many uncapped RNA viruses are able to gain a competitive advantage over the host cell and translation is favoured of viral proteins, rather than host proteins, and allows viruses evade host translational shut-off. It is of note that not every IRES is the same and they are known to differ in their sequence and secondary structure, with differing numbers of stem loops. It is not surprising that the differing structures also mean that different host cell interactions are observed between the IRES types and some interact directly

with the 40S ribosome subunit, whilst others require the involvement of eIFs for recruitment of the 40S ribosome subunit (148, 152).

RNA structures are also known to be integral in the packaging of viral genomes into virus particles, even more so in segmented viruses where it is imperative that the correct assortment of genomes are packaged. The packaging signal, Ψ , in HIV-1 has been well studied and consists of four stem loops. The 155 nt long region is known to act independently for genome packaging and is thereby able to direct the packaging of reporter genes (e.g. GFP-RNA) into VLPs (154). The structure of Ψ was determined by NMR and revealed that there is a tandem three-way junction between the stem loops in which long-range interactions result in the sequestering of the splice donor and translation initiation sequences (155) (figure 1.9B). Furthermore, the exposed guanosines in this region bind to the HIV-1 Gag protein with high affinity (155, 156), and when this interaction is inhibited there is a reduction in viral genome packaging (157). These studies demonstrate how RNA structures can be used as direct targets for anti-viral therapeutics.

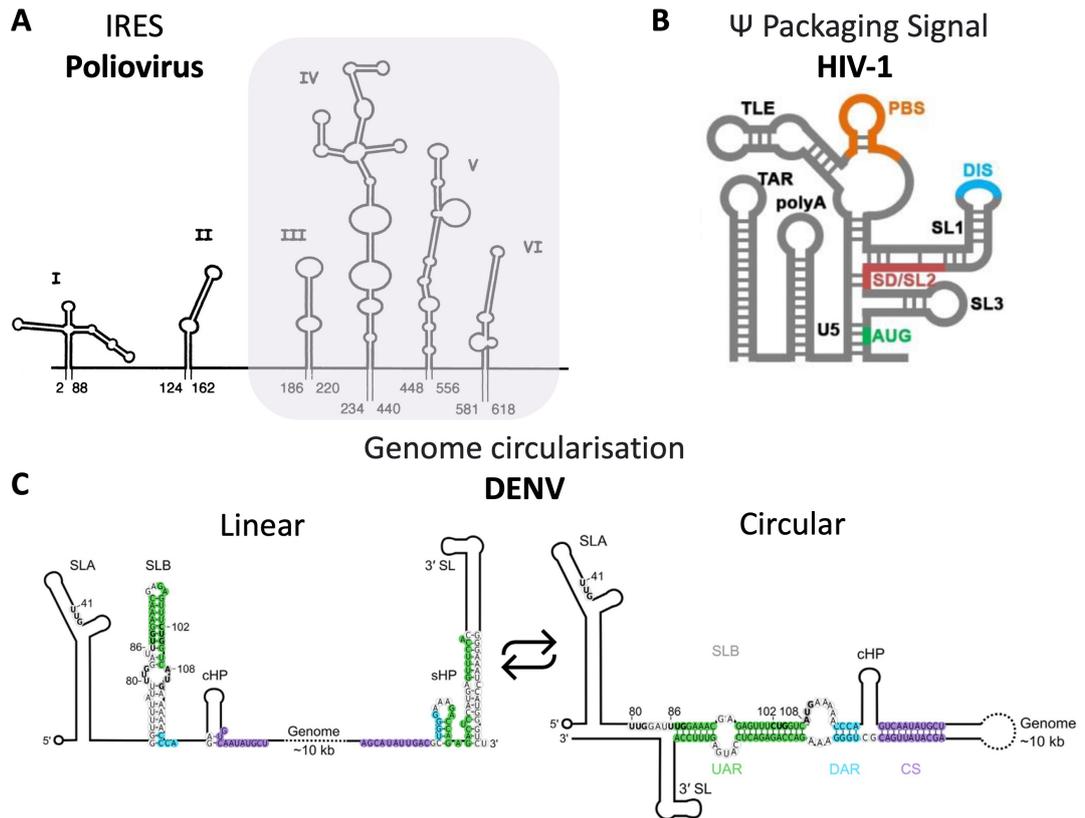


Figure 1.9: Examples of viral RNA structures important for replication

A Poliovirus IRES site highlight in purple (158). **B** HIV-1 Ψ structure involved with correct genome packaging (159). **C** Long-range interactions between the 5' and 3' regions for DENV genome circularisation (160).

Kissing loop interactions are another important structure within the viral lifecycle, and these interactions can direct circularisation of the genome through interactions at the 5' and 3' ends. These interactions are known to be important in the circularisation of DENV. The interactions between complementary sequences in the 5' and 3' termini stabilise the circularised RNA molecule and are essential for DENV replication. There are three distinguished sequences known to be important in DENV genome circularisation (upstream of AUG region (AUR), downstream of AUG region (DAR) and cyclisation sequence (CS)) (figure 1.9C). The DENV RdRp is known to bind to SLA in the 5' UTR and is translocated to the 3' end to initiate negative strand synthesis (161, 162). An ablation of genome replication was reported following the introduction of mutations that would prevent the circularisation of the DENV genome (163), suggesting that the circularised RNA is critical for RNA replication. However, mutants that actually

stabilised the circular genome, rather than linear RNA, resulted in a reduction of overall DENV replication, suggesting that the linear RNA form is also important in the DENV cellular lifecycle (164). A more recent study has confirmed the importance of both the circular and linear RNA forms in flavivirus replication, and that each form has its own defined role (RNA replication and translation, respectively) (160).

1.3.5 Alphavirus RNA structures

There are a variety of different RNA structures across the alphavirus genome that are associated with functions during the virus cellular lifecycle, including packaging (165), regulation of translation (166, 167), replication regulation (168–170) and evasion of the innate immune response (171, 172). The following sections will discuss the importance of RNA structures in context to the alphavirus cellular lifecycle. A large amount of research was performed on other viruses from the genus (e.g. SINV, VEEV, SFV), so much of what is discussed below is also assumed to correspond to CHIKV given the close nucleotide homologies and genome organisation.

Two individual studies determining the RNA secondary structure of the CHIKV 5' region (5' UTR + adjacent upstream region of ORF-1) were published since the project began in 2018 (173, 174) (figure 1.10). These are discussed in further detail in section 4.1. Briefly, Kendall et al. utilised biochemical SHAPE mapping to produce an RNA secondary structure map of the 5' \approx 330 nts and determined the importance of the novel stem loops in CHIKV genomic replication through a reverse genetics approach. Madden et al. used SHAPE and mutational profiling (SHAPE-MaP) to model RNA structures across the entire CHIKV genome and present 23 highly specific structured regions. They also introduced a range of mutations across the 5' CSE in order to determine the importance of stem loops in this region for efficient CHIKV genome replication. It is of note that this study did not include analysis of compensatory mutations to validate their predicted structures.

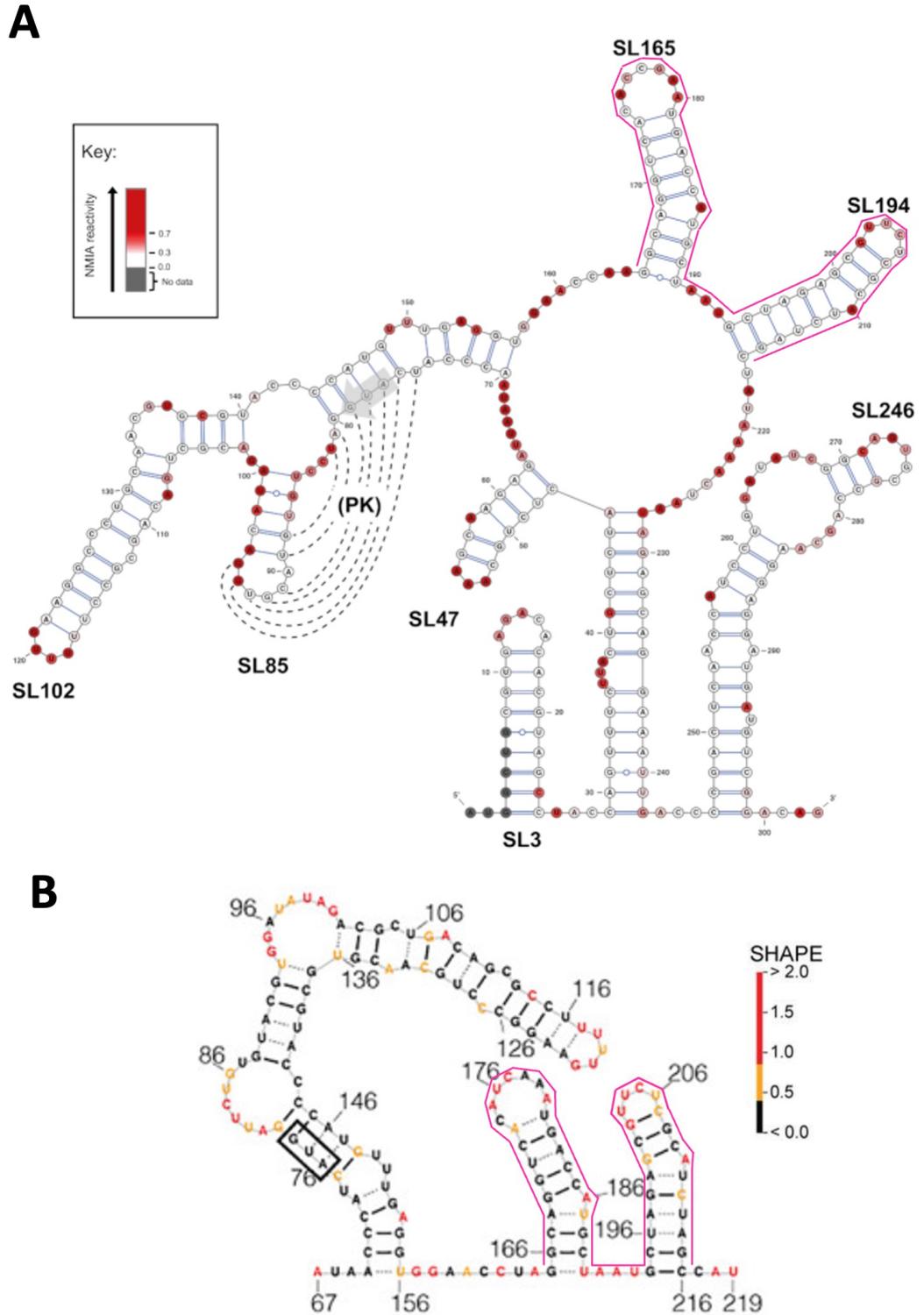


Figure 1.10: SHAPE-constrained thermodynamic folding prediction of the CHIKV 5' region

A The nsP1 start codon is highlighted by a grey arrow. SHAPE reactivities shown as a heat map. PK denotes a potential pseudoknot, with the dotted lines represent putative base-pairing. Pink line denotes the 5' CSE (nts 165 - 216). Adapted from (173). **B** SHAPE-MaP informed secondary structure model of the 5' region from nt 70 - 190. The nsP1 start codon is highlighted in a black box. Nucleotide colour corresponds to SHAPE reactivity scale. Pink line denotes the 5' CSE (nts 165 - 216). Adapted from (174).

1.3.5.1 Alphavirus 5' UTR

The length of the 5' UTR varies greatly across the *alphavirus* genus, with SAV having the shortest (27 nts) and SFV having the longest (85 nts). The CHIKV 5' UTR is 76 nts in length. All the alphavirus genomes have a AU dinucleotide at the 5' end, followed by a stem loop, referred to as SL3 for CHIKV (and throughout this thesis) but is known as SL1 in other alphaviruses. Despite the fact the presence of a stem loop is observed in this region of the 5' UTR, the sequence and structure is not conserved between species. The SINV SL3 has an exposed 11 nt bulge-loop on the 3' side of the double-stranded stem that is not present in other alphaviruses (175). This stem loop functions as part of the promoter during negative strand synthesis and the complement of this region in the 3' UTR acts as a promoter for positive strand genomic RNA synthesis (176–178).

The CHIKV genome includes a type-0 cap structure at the 5' end allowing for cap-dependent translation initiation. This cap is recognised by interferon-induced protein with tetratricopeptide repeats 1 (IFIT-1) due to a lack of methylation at the 2'-O which is seen in the eukaryotic type-1 caps (figure 1.11). Although the mechanism of this is not yet fully understood, IFIT-1 inhibits viral RNA translation (179). Interactions were observed between IFIT-1 and viral RNA when mutations were introduced in SL3, resulting in an overall a lack of virus recovered following infection of cells (172, 180). Compensatory mutants were discovered following passaging of the SL3 mutant virus which restored base-pairing in SL3, and recovered viral levels similar to wild-type. This study demonstrated the requirement of SL3 for the evasion of the host innate immune response and therefore the prevention of viral RNA degradation by host factors.

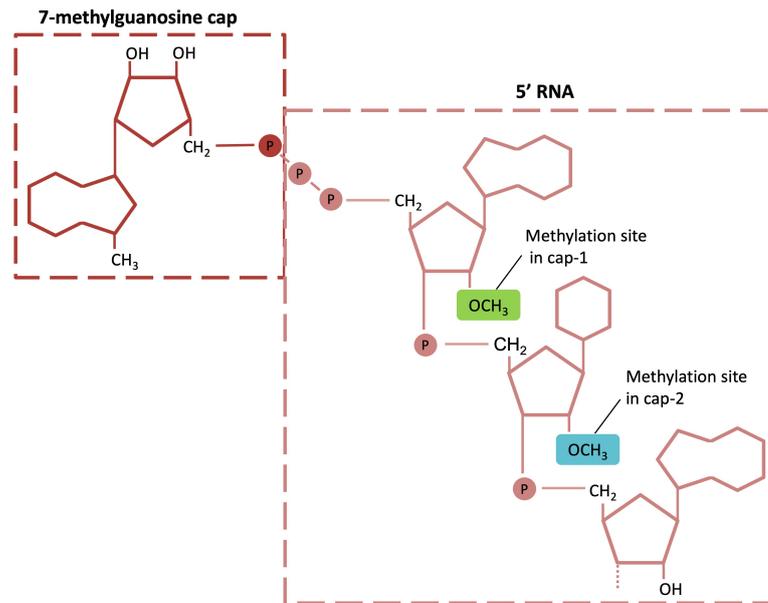


Figure 1.11: Schematic of the 5' 7-methylguanosine mRNA cap structure

The 7-methylguanosine is shown in dark red and the 5' end of RNA shown in pink. The 2'-O site of the first and second nucleotide can be methylated to produce a cap-1 and cap-2 structure respectively (shown in green and blue). CHIKV RNA has a cap-0 structure; neither nucleotide is methylated at the 2'-O site.

Circularisation of the genome is essential for efficient replication of many viruses (such as DENV, discussed in section 1.3.4), and has been suggested to occur during SINV and SFV negative-strand genome synthesis, despite no recognised complementary sequences in the 5' and 3' UTRs (181). It is possible that genome circularisation can occur through interactions of host cell proteins; cap-binding proteins can interact with poly-A binding proteins resulting in transient genome circularisation. Panhandle structures were observed between the 5' and 3' ends of the SINV genome through electron microscopy (181) and biophysical assays were utilised to demonstrate that extracted RNA from virally infected cells could cyclise. Linear and circular RNA can be purified and separated by a glucose gradient (182). When mutations were introduced to disrupt SL3 in SINV, the level of translation increased however there was a decrease in negative-sense genome synthesis in both mammalian and mosquito cells, strongly suggesting that this stem loop is important for this process irrespective of the host/vector (175). A series of chimeric SFV and SINV viruses further demonstrated the importance of SL3, and the terminal AU dinucleotide, for genome replication (176).

A series of mutational studies were performed on VEEV and underlined the importance of the 5' AU dinucleotide, as well as the sequence of the SL3 stem, for efficient promoter function (183). Substitutions of the 5' A or U were detrimental to the virus and compensatory mutations were observed whereby multiple AU's were inserted at the 5' terminal, demonstrating the importance for these nucleotides in genome replication. Mutations to the base of the stem (G-C > C-G) inhibited replication, and passaging of these mutations resulted in the virus gaining compensatory mutations of AU or AUG repeats at the 5' terminal (171). Other disruptions to this stem loop resulted in a gain of heterologous sequences that were capable of forming a similar structure. Further analysis of the VEEV genome further showed that mutations were gained in nsP4, the C-terminus of nsP1 and N-terminus of nsP2, indicating their involvement with genome replication. Similar experiments were performed in SINV, and the passaging of 5' AU deletion resulted in the addition of multiple AU nucleotides (181). This study also determined that variants with multiple AU dinucleotides were capable of replicating to higher levels in mammalian cells compared to mosquito cells.

1.3.5.2 Alphavirus 51 nt CSE

A 51 nt conserved sequence element (CSE) is observed across the alphavirus at the beginning of the coding region for nsP1 (nucleotides 165 to 216 in CHIKV). Due to the differing 5' UTR lengths between species, the CSE alters in nucleotide location but the stem loop structures associated with this region are highly conserved (174).

An early study demonstrated that the sequence of this region was important rather than the structures present; a double mutant was introduced to alter the sequence but maintain the stem loop and there was a significant reduction in virus replication in mosquito cells (177). It is of note that no reduction in replication was observed in mammalian cells for this double mutant, strongly suggesting this region is involved in host-specific interactions. Substitutions to the nucleotides in the terminal loops were also associated with a reduction in virus replication (in

avian and mosquito cells), further suggesting that sequence-specific interactions occur rather than structure-specific during replication (177).

A more recent study uncoupled the presence of secondary structures from the expression of nsP1 in VEEV through extensive synonymous mutations to completely remove all secondary structures within the nsP1 coding region, and then re-introduced the structures individually (184). There was almost no virus recovered when both structures in the CSE were disrupted. However, the presence of at least one stem loop was sufficient for virus generation, with the largest reduction of virus replication observed with the disruption of SL165. This reduction in virus replication was observed in both mammalian and mosquito cells. The stem loop knockout viruses were then passaged in BHK-21 cells twice, and further mutations were observed in the 5' upstream of the ORF-1 start codon and at the N-terminal of nsP2 and nsP3. To confirm whether these mutations were sufficient to rescue virus replication to wildtype levels, they were reincorporated into VEEV and it was reported that mutants in nsP2 and 5' UTR were not synergistic, suggesting that these regions are involved in different processes during replication (184). The importance of interactions between nsP2 and nsP3 with the CSE were confirmed when similar passaging experiments were performed in SINV and mutations arose again in these two proteins, and the 5' UTR, as well as E2 (169). The pseudo-revertant mutation in nsP2 was not sufficient to recover the CSE mutant to wildtype levels of replication and instead required the presence of the nsP3 pseudo-revertant mutant or the pseudo-revertant UTR mutant. Interestingly, the restoration of replication observed in mosquito cells was not observed within mammalian cells, suggesting that the CSE region is important in host-specific interactions.

1.3.5.3 Alphavirus 3' UTR

As with the 5' UTR, the length of the 3' UTR varies across the alphaviruses, with SAV having the shortest at 87 nts and CHIKV the longest at 723 nts. The length of the 3' UTR also differs between CHIKV species, from \approx 500 - 700 nts (185). The CHIKV 3' UTR is comprised of three direct repeat regions (DR1 -

3) which occur in a lineage specific pattern, with the Asian lineage having the longest UTR due to a second repeat of DR3 region and DR(1+2) region. There is a 19 nt CSE prior to the poly(A) tail, which functions as the promoter for negative-strand synthesis (185, 186). Multiple studies that introduced mutations into this region have suggested that the importance of the 3' UTR structural elements is host-specific (174, 187).

A recent study discovered that three variants co-existed during infection of CHIKV with varying 3' UTR variants (deemed canonical, duplication (partial duplication of the DR(1+2) repeat) and deletion (removal of the 3' end of the first DR(1+2) repeat)). SHAPE-MaP analysis on *in-vitro* transcribed RNA of the three 3' UTR variants was performed in order to determine structural differences between them (174). They reported that the start of the UTR had the greatest difference in RNA structure, while the 3' end remained conserved between variants. The duplication variant contained a repeat of a two-hairpin loop structure present in the canonical variant, and the deletion variant did not contain this structure. Interestingly, they also reported that compared to the canonical UTR the duplication variant enhanced replication levels in mosquito cells and that the deletion variant inhibited replication. There was no observed difference in replication rates across the three variants in mammalian cells and in a mouse model, strongly suggesting that the structures at the beginning of the 3' UTR are involved with mosquito-specific interactions.

The generation of mutants in the 3' CSE was associated with a greater reduction in virus replication in mosquito cells than avian cells (186), and more specifically inhibition of RNA replication (188), suggesting that this region functions as a promoter for negative-strand synthesis (189). The 19 nts CSE terminates in a cytosine and mutagenesis or deletion of this nucleotide in SINV prevents negative-strand synthesis, the insertion of 3 or more uridylates recovers RNA synthesis, further confirming the importance of this region to function as a promoter (168, 170). Early studies reported that SINV can tolerate the conserved cytosine being displaced by at least 7 nts from its usual place adjacent to the poly(A) tail (186), whereas a later study demonstrated that 11 or more A residues immediately following the CSE were required for efficient negative-strand synthesis (170).

The 3' UTR of alphaviruses is known to interact with RNA binding proteins; human-antibody R (HuR) and the La autoantigen homologue in mosquitoes (190, 191). The majority of alphaviruses, excluding CHIKV, ONNV and RRV, contain a U-rich sequence preceding the 3' CSE, that is understood to be the interaction site with HuR (190). In ONNV and RRV, this interaction occurs with the repeating elements further downstream the 3' UTR (192). The interaction between RNA and HuR aids in RNA stability in the cytoplasm and prevents deadenylation and viral RNA decay (192, 193). Due to the high affinity of this interaction, HuR is sequestered in the cytoplasm and cellular mRNAs which usually depend on HuR for their stability are ultimately degraded. There is a requirement for additional studies to further elucidate the functional implications of interactions between viral RNA and cellular La protein in mosquitoes (185).

1.3.5.4 Additional structures within protein coding regions

There are RNA structures present within the CHIKV genome that determine recoding during translation to allow for the translation of nsP4 and TF protein. There is an opal stop codon present in between nsP3 and nsP4, and TCR occurs in the presence of an elongated stem loop at a rate of 10% (166, 194). SHAPE analysis of this region revealed a terminally forked stem loop downstream of the opal stop codon (166). Mutations to interrupt this structure resulted in ablation of TCR and a reduction in translation (166). Furthermore, this forked stem loop structure is conserved across CHIKV lineages and alphaviruses through co-variation (173, 195). A -1 frameshift in translation of the structural 6K gene results in the translation of the TF protein (196). This frameshift requires the presence of a programmed ribosomal frameshifting (PRF) stem-loop, which was determined to be a short, bulged stem loop with homology to structures observed in EEV (166, 197). Mutations to abolish this stem loop were shown to decrease the expression of the TF protein (166).

During virus infection, host cells induce translational shut off to inhibit virus replication. This would mean that during late stages of replication, the translation of structural proteins from the sub-genomic replicon would be limited. There is

a predicted highly stable stem loop that forms downstream of the sub-genomic AUG which is critical in the initiation of translation without eIF2 (198). This structure is referred to as the downstream stable hairpin (DSH). The DSH is believed to stall the 43S ribosomal subunit at the start codon for ORF-2 and initiate translation; potentially through the direct binding with the P site (199, 200).

1.4 Aims of the project

Preliminary work from the group identified an in-frame AUG start codon that was conserved at amino acid residue 24 across CHIKV and ONNV (deemed M²⁴). They initially demonstrated that mutations introduced at M²⁴ were detrimental to virus replication and observed a potential truncated version of nsP1 by Western blot analysis, however, these results were inconsistent. They hypothesised that M²⁴ functioned as an alternative start codon to translate a truncated form of nsP1, and that this protein was essential for efficient CHIKV replication with functions that had not been previously characterised in CHIKV infection.

The overarching goal of this project was to further understand the importance of nsP1 within the CHIKV lifecycle and to determine whether it had novel, non-canonical functions. The project was broken down into 3 main aims:

1. Verify preliminary data to confirm whether M²⁴ functions as an alternative start codon to translate a truncated version of nsP1 through the use of an infectious clone to analyse virus replication as a whole, alongside luciferase based replicon systems to isolate genome replication from other aspects of the lifecycle.
2. Determine mechanistic functions/interactions of the M²⁴ codon, through SHAPE analysis to produce minimum free energy models of M²⁴ mutant RNA secondary structures and passaging assay of mutant CHIKV in mammalian and mosquito cells.
3. Express and purify recombinant full-length nsP1 for X-ray crystallography

analysis in order to gain insight into the protein structure and compare full-length nsP1 to N-terminally truncated nsP1.

Chapter 2

Materials and Methods

2.1 DNA/RNA

2.1.1 Transformation of bacterial cells

Chemically competent *Escherichia coli* (*E. coli*) (DH5 α) cells were transformed with ICRES CHIKV plasmid DNA. 0.1-100 ng/ μ L of DNA was incubated with DH5 α for 5 minutes on ice and then spread out on lysogeny broth (LB) agar plates supplemented with 0.1 μ g/mL ampicillin. Plates were incubated at 37°C overnight. Colonies were then picked for inoculation of LB broth (100 μ g/mL ampicillin) and incubated at 37°C overnight. Plasmid DNA were extracted from bacteria using either the ThermoScientific Genejet Plasmid Miniprep or Maxiprep kit where appropriate, according to the manufacturer's recommended protocol. Glycerol preparations were used for long term storage of plasmid DNA, 2 mL of bacteria overnight culture was pelleted and resuspended with 20 % glycerol:LB broth and stored at -80°C.

Chemically competent DH5 α cells were transformed with the *trans*-replicase plasmids. 10 ng of DNA was incubated with DH5 α for 5 minutes on ice, 250 μ L of SOC media (New England Biolabs) was added and the bacteria was allowed to recover at 37°C for 1 hour, shaking at 180 x g. The mixture was then spread out on LB plats that were supplemented with 50 μ g/mL kanamycin. Plates were incubated and DNA extracted as described above.

E. coli Rosetta 2 cells were transformed with pET28a-SUMO-CHIKV-nsP1 plasmid DNA. 100 ng/ μ L of DNA was incubated with *E. coli* Rosetta 2 cells for 30 minutes on ice and then heat shocked at 42°C for 30 secs and returned to ice for 2

minutes. Pre-warmed SOC media (New England Biolabs) was added to the cells and they were allowed to recover for 1 hour at 37°C in a shaking incubator (at 180 x g). Finally the cells were then plated out onto LB agar plates supplemented with 50 µg/mL kanamycin. Colonies were picked as described above (with the LB broth supplemented with 50 µg/mL kanamycin) and plasmid DNA extracted using the ThermoScientific Genejet Plasmid Maxiprep kit.

2.1.2 Endonuclease digestion

For digestion of DNA during cloning, DNA was digested with the appropriate restriction enzyme in custmart buffer and incubated at 37°C for a minimum of 1 hour.

For the linearisation of CHIKV ICRES plasmid DNA prior to *in vitro* transcription, 10 µg of DNA was digested with NotI and incubated at 37°C for 4 hours.

2.1.3 Agarose TAE gel electrophoresis

The DNA quality and size was analysed by agarose gel electrophoresis. Fragments > 500 bp were analysed on a 1 % gel and fragments < 500 bp were analysed on a 2 % agarose gel. Agarose was mixed with 1x TAE buffer (40 mM Tris, 20 mM Acetate and 1 mM EDTA) and SYBR® safe DNA gel stain (1:10,000, Invitrogen), the mixture was heated to ensure the agarose fully melted. DNA samples were electrophoresed at 90 V for 1 hour, alongside a 1 kb Plus DNA ladder (New England Biolabs).

2.1.4 Extraction of DNA

DNA was visualised by blue light (470 nm) and the appropriate bands were excised with a clean scalpel and placed in nuclease free microcentrifuge tube. DNA was extracted using the Monarch Gel Extraction kit (New England Biolabs) according to the manufacturer's recommended protocol and DNA was eluted in

10 μ L nuclease free water.

2.1.5 Ligation of DNA

Vector DNA that had been digested with the appropriate restriction enzyme was treated with 10 U Quick CIP (New England Biolabs) for 15 minutes at 37°C. Vector and insert DNA were ligated at a ratio of 1:3 (vector:insert), with a total concentration \leq 100 ng. The ligation reaction was assembled with T4 DNA Ligase (New England Biolabs), T4 ligase buffer (New England Biolabs), nuclease free water, vector DNA and insert DNA, and then incubated at room temperature for 30 minutes. The DNA was then transformed into chemically competent DH5 α as described in section 2.1.1).

2.1.6 Phenol-chloroform extraction and ethanol precipitation of DNA

Following linearisation, dH₂O was added to the digest mixture to a total volume of 200 μ L. 1 volume of UltraPureTM Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Invitrogen) was added and vortexed for 1 minute before being centrifuged for 5 minutes at 13,000 x g. The upper, aqueous phase was extracted and placed in a new microcentrifuge tube. 1 volume of chloroform (Sigma) was then added and the sample was vortexed and centrifuged as previously described and the upper phase was extracted as before. 2 volumes of 100 % ethanol and 0.1 volume of 3 M sodium acetate (pH 5.8) were added and the samples were incubated at -20°C for 3 hours. Following incubation, samples were centrifuged at 4°C for 20 minutes at 13,000 RCF. The supernatant was removed and the pellet washed with 500 μ L 70 % ethanol. The samples were centrifuged for a final time for 5 minutes at 4°C at 13,000 RCF and then the pellet was allowed to air dry before being re-suspended in 20 μ L dH₂O.

2.1.7 Oligonucleotide primers

DNA oligonucleotides were ordered from Integrated DNA Technologies and dissolved with nuclease free water at 100 μ M. All oligonucleotides were stored at -20°C.

2.1.8 DNA sequencing

In order to confirm the correct sequencing following PCR cloning steps, 100 ng of DNA was mixed with 25 μ M of appropriate primers and submitted to Eurofins Genomics for Sanger sequencing.

2.1.9 *In vitro* transcription

RNA for the generation of CHIKV and for the sub-genomic replicon experiments required a 5' cap. Capped RNA synthesis was achieved using AmpliCap™ SP6 High Yield Message Maker Kit (Cellscript) following the manufacturers guidelines. Briefly, 1 μ g of linearised template DNA (see 2.1.2) was mixed with SP6 enzyme solution, a cap/NTP mix, DTT, RNase inhibitor and transcription buffer and incubated at 37°C for 2 hours. Following this incubation, DNase I was added and further incubated for 15 minutes at 37°C.

The SHAPE protocol required uncapped *in vitro* transcribed RNA. The T7-Scribe™ Standard RNA IVT kit was used to generate uncapped RNA according to the manufacturer's recommended protocol. Briefly, 1 μ g of purified linear DNA was mixed with 100 mM ATP, GTP, UTP and CTP, 10x T7-Scribe™ reaction buffer, T7 enzyme mix, 100 mM DTT and 40 U/ μ L RNase inhibitor. The transcription reaction was incubated at 37°C for 2 hours, 1 μ L DNase I was added to the reaction and incubated at 37°C for a further 15 minutes. The RNA was purified as described in either section 2.1.11 or 2.1.12.

2.1.10 Lithium chloride purification of RNA

In vitro transcribed RNA (ICRES and monoluciferase sub-genomic replicon RNA) was purified using lithium chloride precipitation. 30 μ L diethyl pyrocarbonate (DEPC)-dH₂O and lithium chloride (7.5 M, Invitrogen) were added to the RNA mixture from 2.1.9 and incubated overnight at -20°C. Following incubation, the samples were centrifuged at 13,000 rpm for 20 minutes at 4°C. The supernatant was discarded and the RNA pellet was washed with ice cold 80 % ethanol. The washed pellet was centrifuged for at 13,000 x g for 10 minutes at 4°C and the supernatant discarded. The RNA pellet was resuspended in DEPC-dH₂O. Purified RNA was quantified on the Nanodrop by absorbance at A₂₆₀, followed by analysis on a 0.1 % 3-(N-morpholino)propanesulfonic acid (MOPS) gel to confirm purity. RNA was stored at -80°C.

2.1.11 Phenol-chloroform extraction of RNA

In vitro transcribed RNA for SHAPE analysis (section 4.2.1) was purified using phenol chloroform extraction. The RNA sample was made up to 200 μ L with nuclease free water. One volume of phenol:chloroform:isoamyl alcohol (pH 6.7; Ambion) was added to the RNA sample and mixed by vortexing for 1 minute, then centrifuged for 10 minutes at 16,000 x g at 4°C. The upper, aqueous phase was aspirated and transferred to a clean nuclease free microcentrifuge tube. 0.1 volume of sodium acetate (3 M, pH 5.2) and 2.5 volumes ice cold 95 % ethanol were added to the sample and mixed well by pipetting, then incubated at -80°C for \geq 1 hour. The RNA was pelleted by centrifugation at 16,000 x g for 10 minutes at 4°C. The supernatant was discarded and the pellet washed with 70 % ethanol. Following a final centrifugation at 16,000 x g for 15 minutes at 4°C, the supernatant was removed. Once the ethanol had fully evaporated and the pellet was dry, RNA was resuspended in 50 μ L of nuclease free water. RNA was stored at -80°C.

2.1.12 Column purification of RNA

From 2022, all RNA was purified using the PureLink RNA kit (ThermoFisher) according to the manufacturer's recommended protocol. RNA was eluted in 50 μ L nuclease free water.

In vitro transcribed RNA for SHAPE analysis (section 4.2.5) was purified using the PureLink™ RNA kit, according to the manufacturer's recommended protocol and the RNA was eluted in 50 μ L of nuclease free water.

2.1.13 RNA agarose gel electrophoresis

The integrity and size of RNA was analysed on an agarose formaldehyde MOPS gel. A 1 % or 2 % agarose gel was prepared by mixing nuclease free agarose in 1x MOPS buffer (40 mM MOPS, 10mM sodium acetate and 1 mM EDTA). Once the agarose was fully melted, 6.5 % formaldehyde and SYBR® safe DNA gel stain (1:10,000) were added in a fume hood and then the mixture was poured into a gel cast. RNA samples were mixed at a 1:1 ratio with RNA loading dye (Invitrogen) and electrophoresed at 90 V for 1 hour, alongside the Millenium™ RNA Marker (Invitrogen).

2.2 Selective 2' Hydroxyl Acetylation Analysed by Primer Extension (SHAPE)

2.2.1 Template preparation

CHIKV wild-type mono-luciferase DNA was used as a template for PCR amplification of the region 1 - 337 nts of the CHIKV genome. 10 ng of DNA was combined with 10 mM dNTPs, 10 μ M forward and reverse primer, DMSO (final concentration of 3 %), 5X Phusion GC buffer, nuclease free H₂O and Phusion polymerase (0.02 U/ μ L, New England Biolabs) in a total volume of 50 μ L. The

reaction was incubated at 98°C for 2 minutes, 35 cycles of 98°C for 15 seconds, 55°C for 30 seconds and 72°C for 30 seconds followed by 72°C for 10 minutes. PCR productions were purified using the Monarch [®] PCR clean up kit according to the manufacturer's recommended protocol, and DNA was eluted in 12 μ L nuclease free water. RNA was *in-vitro* transcribed using the T7 *In-Vitro* Transcription kit ((Cellscript) see section 2.1.9), and purified by either column purification or phenol chloroform extraction (see section 2.1.12 and 2.1.11).

2.2.2 RNA folding and NMIA treatment

25 pmol of RNA diluted in 0.5 x TE buffer (Ambion) was heated at 95°C for 2 minutes, and cooled by incubation for 2 minutes on ice. 45 μ L of 3.3X folding buffer (333 mM HEPES (pH 8.0), 20 mM MgCl₂, 333 mM NaCl) and 80 units RNase inhibitor (Promega) was added to the RNA and the volume made up to 150 μ L with nuclease free water, the RNA was incubated at 37°C for 30 minutes. Each sample was divided into two tubes and the folded RNA was incubated with either 5 mM NMIA or DMSO at 37°C for 50 minutes.

Once treated, RNA was ethanol precipitated by the addition of 5 % v/v NaOAc, 2 μ L 100 mM EDTA, 20 μ g glycogen and made up to a total concentration of 70 % ethanol with nuclease free water. Precipitations were incubated at -80°C for at least 1 hour prior to pelleting by centrifugation at 4 °C at 16,000 x g for 30 minutes. The pellet was washed with 70 % ethanol and resuspended in 10 μ L 0.5 x TE buffer.

2.2.3 Primer extension

The treated RNA was combined with 1 μ L of 10 μ M 6-FAM fluorescein labelled primers and made up to a total volume of 12 μ L nuclease free water. The reactions were incubated at 85°C for 1 minute, 60°C for 10 minutes and 35°C for 10 minutes. The reaction was reverse transcribed using SuperScript III (ThermoFisher Scientific). Briefly, the RNA/primer mix was combined with a

master mix of 5 x First Strand buffer, 12 mM DTT, 40 units RNase inhibitor, 1.3 mM dNTPs, 200 units SuperScript III and nuclease free water. The reaction was incubated at 52°C for 30 minutes. Upon reverse transcription, 1 µL 4M NaOH was added to the reaction and then incubated at 95°C for 2 minutes. The samples were immediately cooled on ice and 2 µL of 2M HCl was added. The cDNA was ethanol precipitated as described in 2.2.2, with the replacement of being resuspended in 40 µL highly deionised formamide. Once resuspended, the cDNA was heated at 65°C for 10 minutes and then vortexed briefly prior to being stored at -80°C.

2.2.4 Ladder preparation

The sequencing ladder was prepared in the same way as described in 2.2.3, however 10 pmol of untreated RNA was combined with Hexachloro-flourescein labelled primer and the addition of 1 µL 10mM ddATP to the reverse transcription master mix.

2.2.5 Capillary Sequencing of SHAPE reactions

The sequencing ladder was mixed with the NMIA or DMSO treated sample in a 1:4 ratio, and submitted to Dundee DNA Sequencing and Services for analysis by capillary electrophoresis.

2.2.6 Data analysis

The raw sequencing trace files were analysed in the programme QuSHAPE to generate SHAPE reactivity values. Negative values were set to 0 and "super-high" reactive nucleotides were set to 3. The Vienna RNAProbing server was used to generate RNA structure predictions, using SHAPE reactivity scores to guide base pairing. VARNA was used to visualise the RNA secondary structure from the dot bracket file generated by Vienna. The reactivity values were overlaid onto

a predicted RNA structure and a scale bar set of white (0 - 0.3), a gradient of white to red (0.3 - 0.7) and red (>0.7).

2.3 Cell culture methods

2.3.1 Continuous cell culture

All mammalian cell lines were maintained at 37°C with 5 % CO₂ in humidified incubators. All mammalian cell lines were grown in Dulbecco's Modified Eagle's Media (DMEM; Sigma) supplemented with 10 % foetal bovine serum (FBS; Invitrogen) and 100 IU penicillin/mL, and 100 µg streptomycin/mL, this is referred to as complete media. Cell lines were passaged using trypsin (Sigma) as the cells were reaching 100 % confluency.

The mosquito cell line (C6/36, *Ae. albopictus*) was maintained at 28°C. C6/36 cells were grown in Leibovitz media (Sigma) supplemented with 10 % FBS (Invitrogen), 10 % tryptose phosphate brother (Sigma), 100 IU penicillin/mL and 100 µg streptomycin/mL (this is referred to as complete Leibovitz media). When cells were reaching 100 % confluency they were passaged by mechanical scraping.

2.3.2 Transfection of cells

2.3.2.1 Sub-genomic replicon assay

Huh7 cells was seeded at 1.5×10^5 cells per well in a 24 well plate the day prior to transfection and allowed to reach 70 % confluency. 500 ng of RNA was mixed with 50 µL of Opti-MEM™ per well. 1 µL of Lipofectamine-2000 (ThermoFisher) was mixed with 50 µL Opti-MEM™ and incubated at room temperature for 5 minutes. The RNA mix was gently pipette mixed with the Lipofectamine-2000 mix and incubated for 20 minutes at room temperature. Cells were washed with 1x PBS and media was replaced with Opti-MEM™. The RNA/Lipofectamine-2000 mix was added dropwise to each well. Cells were incubated at 37°C for the appropriate

time. Supernatant was removed from cells and the monolayer was lysed with 1x passive lysis buffer (PLB, Biotium) and stored at -20°C prior to analysis.

Lysates were analysed for luciferase activity levels using a FLUOstar plate reader (BMG LABTECH). 10 µL lysate was transferred to a white bottom 96 well plate and the firefly luciferase activity of samples was measured by the addition of firefly working solution (Biotium).

2.3.2.2 *Trans*-replicase assay

Huh7 cells was seeded at 1.5×10^5 cells per well in a 24 well plate the day prior to transfection and allowed to reach 70 % confluency. 500 ng of plasmid reporter and replicase plasmid DNA was mixed together with 50 µL Opti-MEM™. 1 µL of Lipofectamine-2000 (ThermoFisher) was mixed with 50 µL Opti-MEM™ and incubated at room temperature for 5 minutes. The DNA mix was gently pipette mixed with the Lipofectamine-2000 mix and incubated for 20 minutes at room temperature. Cells were washed with 1x PBS and media was replaced with Opti-MEM™. The DNA/Lipofectamine-2000 mix was added dropwise to each well. Cells were incubated at 37°C for the appropriate time. Supernatant was removed from cells and the monolayer was lysed with 1x PLB and stored at -20°C prior to analysis.

Lysates were analysed for luciferase activity levels using a FLUOstar plate reader (BMG LABTECH). 10 µL lysate was transferred to a white bottom 96 well plate and the firefly luciferase activity of samples was measured by the addition of firefly working solution . An additional 10 µL of lysates was used to measure the *Gaussia* luciferase activity by addition of *Renilla* working solution (Biotium).

2.4 Virological methods

2.4.1 CHIKV generation

2.4.1.1 BHK-21 Electroporation

Virus was generated by transfecting *in vitro* transcribed RNA into BHK-21 cells. Cells were trypsinised and centrifuged at 1,000 x g for 1 minute, cell pellets were washed in ice cold 1x DEPC PBS. This was repeated twice and then cells were resuspended in ice cold 1x DEPC PBS at 3×10^6 cells/mL. 1 μ g of RNA was added to 1.2×10^6 BHK-21 cells in a pre-cooled 4 mm cuvette and electroporated using a square-wave setting: 260 V for 25 ms. Immediately after electroporation, cells were resuspended in 10 mL complete media. Cells were seeded directly into a T75 flask and supernatant (containing infectious virions) was collected after 24 hours incubation at 37°C.

2.4.1.2 C6/36 Lipofectamine-2000 Transfection

Virus was generated by transfecting *in vitro* transcribed RNA into C6/36 cells. The day prior to transfection, 1.4×10^6 C6/36 cells were seeded in a T75 and incubated overnight to reach 70-80 % confluency. 10 μ g of viral RNA was mixed with 50 μ L Opti-MEM™ (Gibco). 10 μ L of Lipofectamine-2000 (ThermoFisher) was mixed with 50 μ L Opti-MEM™ per T75 flask, and incubated for 5 minutes at room temperature. The RNA/Opti-MEM™ mix was added with the Lipofectamine-2000/Opti-MEM™ and incubated for a further 20 minutes at room temperature. Cells were washed with 1x PBS and the media was replaced with 4 mL of Opti-MEM™. 100 μ L of the Lipofectamine-2000/RNA mixture was added drop wise onto the cells. Following a 4 hour incubation, the Opti-MEM™ was replaced with complete Leibovitz media. Supernatant containing released infectious CHIKV was harvested 48 hpt and stored at -80°C.

2.4.2 Infectious centre assay

The day prior to electroporation, 1.5×10^5 BHK-21 cells were seeded into a 12 well plate, and incubated overnight to reach 90-95 % confluency. Cells were electroporated according to section 2.4.1.1. Electroporated cells were resuspended up to 1 mL in complete media, 100 μ L of resuspended electroporated cells were used to perform a dilution series of 10^{-1} to 10^{-6} . The media was removed from the monolayer of BHK-21 cells seeded the day prior and replaced with 500 μ L of diluted electroporated cells. Cells were incubated for 1-2 hours. Complete media was diluted 1:1 with 1.6 % (w/v) high viscosity carboxyl-methyl cellulose (CMC, sigma). Supernatant was removed and cells were washed with PBS, complete media:CMC was added to the cells as an overlay. The cells were incubated for a 48 hours. Cells were fixed using 4 % formaldehyde for 30 minutes and stained using a 20 % ethanol and 0.1 % (w/v) crystal violet solution for 1 hour. Cells were washed thoroughly with dH₂O and allowed to dry. The following equation was used to determine titre:

$$\text{Titre (PFU/mL)} = \frac{\text{Number of plaques}}{\text{dilution factor} \times \text{volume (mL)}}$$

2.4.3 Virus infection

Once cells had reached 70-80 % confluency complete media was replaced with fresh medium and infected with CHIKV at the stated multiplicity of infection (MOI). Cells were incubated for 1 hour with rocking, cells were then washed twice with PBS and complete media was added. Infected cells were incubated at 37°C for the desired length of time.

2.4.4 CHIKV titre determination by plaque assay

1.5×10^5 BHK-21 cells were seeded into a 12 well plate and incubated at 37°C overnight to reach 90-95 % confluent. A serial dilution of CHIKV was made, from

10^{-1} to 10^{-7} in SFM. Complete media was removed from cells. Virus dilutions were added carefully to the cells and incubated at 37°C for 1 hour on a rocker. Complete media was diluted 1:1 with 1.6 % (w/v) high viscosity CMC (Sigma). Virus was removed from the cells, and complete media:CMC was added to the cells as an overlay. CMC prevents virus migrating through the media, ensuring that the secreted infectious virion from an infected cell could only infect immediately neighbouring cells. Plaques were formed due to CHIKV infection resulting in cell lysis in very defined regions. The cells were incubated at 37°C for 48 hours. Cells were fixed using 4 % formaldehyde for 30 minutes and stained using a 20 % ethanol and 0.1 % (w/v) crystal violet solution. The chromophore containing component of crystal violet binds to the negatively charged cell membranes (it is positively charged at pH 7). Healthy cells appears purple in the plaque assay, and areas where the virus has caused lysis appear transparent as they are not stained. Cells were incubated for 1 hour in stain solution, stain was removed and cells were washed thoroughly with dH_2O , and allowed to dry. Virus titre was determined using the equation in section 2.4.2.

2.4.5 Passaging

For each passage, a T25 flask of the appropriate cell line was allowed to reach 70-80 % confluency prior to infection. Passaging began by infecting the passage 1 (P1) monolayer at MOI 1. The supernatant was harvested 48 hpi and 1 mL of supernatant was used to infect a naïve T25 flask of appropriate cells. This was repeated up to the harvesting of supernatant from passage 10 (P10). The supernatant from all passages was titred. If the titres were suitable, an infection was performed in huh7 cells at MOI 1 or 0.1 for 24 hours, and the supernatant titred.

2.4.6 RNA Extraction

The monolayer of cells following each passage were harvested with Tri-Reagent (Sigma Aldrich). RNA was extracted from this using PureLink™ RNA Mini Kit

(ThermoFisher) according to the manufacturer’s instructions and eluted in 30 μ L RNase free water. Extracted RNA was stored at -80°C .

2.4.7 Next Generation Sequencing

The Oxford Nanopore Technology, PCR tiling of SARS-CoV-2 virus with rapid barcoding (SQK-RBK-110), protocol was adapted for Next Generation Sequencing analysis of the CHIKV genome. Primers to generate eleven 1,200 bp amplicons were designed using Primal Scheme (201). Primers were combined to make a 100 μ M stock deemed primer pool 1 and 2 (see table 2.1).

Table 2.1: Next Generation Sequencing primers

Primers were designed using Primal Scheme. "Left" indicates the forward primer, "Right" indicates the reverse primer with the number referring to the amplicon.

| Primer Pool 1 | |
|----------------------|-----------------------------|
| Name | Sequence |
| CHIKV_ICRES_1_LEFT | AGACACACGTAGCCTACCAGTT |
| CHIKV_ICRES_1_RIGHT | TGTTTCGTATTCCGTTGCGTTCT |
| CHIKV_ICRES_3_LEFT | TGTAAGAAGGAAGAAGCCGCAG |
| CHIKV_ICRES_3_RIGHT | AATAGCCCGCTGTCTAGATCCA |
| CHIKV_ICRES_5_LEFT | CGCAAAAACAGTTATGTGCGGT |
| CHIKV_ICRES_5_RIGHT | TGTTTCTTGTACGACCGGACAC |
| CHIKV_ICRES_7_LEFT | ACCCACTTTGGACTCAGCAGTA |
| CHIKV_ICRES_7_RIGHT | ACCTATTTAGGACCGCCGTACA |
| CHIKV_ICRES_9_LEFT | ATCCCAGTTATGTGCCTGTTGG |
| CHIKV_ICRES_9_RIGHT | CGGCCAATACTTATACGGCTCG |
| CHIKV_ICRES_11_LEFT | TTCATTGTGGGGCCAATGTCTT |
| CHIKV_ICRES_11_RIGHT | TTGTTACTATTCAGGGGTTTTATAGCC |
| Primer Pool 2 | |
| CHIKV_ICRES_2_RIGHT | TCCGAAGACTCCTATGACTGCA |
| CHIKV_ICRES_2_LEFT | AATGTCATTCTCGGTGTGCACA |
| CHIKV_ICRES_4_RIGHT | CTTTGCGACTTCTCGATAGGCA |
| CHIKV_ICRES_4_LEFT | TTGTTGGGCTAAGAGCTTGGTC |
| CHIKV_ICRES_6_RIGHT | CTGCTGCTTTTGGCCCTTTTAG |
| CHIKV_ICRES_6_LEFT | GGTAATGAGCACCGTACCTGTC |
| CHIKV_ICRES_8_RIGHT | CGGTGGGGAGAACATGTTAAGG |
| CHIKV_ICRES_8_LEFT | ATGGCAACGAACAGGGCTAATT |
| CHIKV_ICRES_10_RIGHT | GCCAAATTGTCCTGGTCTTCCT |
| CHIKV_ICRES_10_LEFT | AACACTCCTGTCCTACCGGAAT |

2.4.7.1 Library Preparation

Two individual reverse transcription reactions were performed on each sample; 8 μL of extracted total RNA was mixed with 2 μL of the LunaScript™ RT SuperMix (New England Biolabs) and then incubated in a thermocycler at 25°C for 2 minutes, 55°C for 10 minutes and 95°C for 1 minute, and then held at 4°C. A mastermix was prepared for two separate PCR reactions of: RNase free H₂O, primer pool 1 or 2 (100 μM) and Q5® Hot Start 2x Master Mix (New England Biolabs). The primer pool 1 or 2 mastermix was added to clean RNase free PCR tubes and then mixed with 2.5 μL of the cDNA generated from the RT step. PCR reactions were performed with the following cycle: 98°C for 30 seconds, 35 cycles of 98°C for 15 seconds and 65° for 5 minutes, and a final hold of 4°C.

The entire contents of the primer pool 2 reaction was combined with the primer pool 1 reaction and pipette mixed. 7.5 μL of the combined PCR products was mixed with 2.5 μL of each "barcode" from the Rapid Barcoding Plate (Oxford Nanopore Technology). The sample was incubated at 30°C for 2 minutes and 80°C for 2 minutes. Following attachment of the "barcodes", all individual samples were pooled together. The combined samples were mixed at a 1:1 ratio with AMPure XP SPRI beads (Beckman Coulter) and incubated at room temperature for 5 minutes. A magnetic rack was used to pellet the beads and the supernatant was aspirated and disposed of. The pellet was washed twice with 1 mL 80 % ethanol and the beads were allowed to dry at room temperature for 5 minutes to ensure all ethanol had evaporated. Elution buffer (Oxford Nanopore Technology) was added and incubated with the beads for 10 minutes at room temperature. The beads were pelleted using a magnetic rack and the elution was aspirated and transferred to a fresh nuclease free tube.

The eluted DNA was mixed with 1 μL of Rap F (Oxford Nanopore Technology), incubated at room temperature for 5 minutes and then stored on ice. The flow cell priming mix was prepared by mixing the Flush Buffer (Oxford Nanopore Technology) with the Flush Tether (Oxford Nanopore Technology). Finally, the prepared library was mixed with Sequencing Buffer II and Loading Beads II.

2.4.7.2 Sequencing Run

The flow cell was initially primed with 800 μL of priming mix through the priming pore, with care taken to not introduce any air into the flow cell, and incubated at room temperature for 5 minutes. The sample pore was opened and a further 200 μL of priming mix was used to finish priming the flow cell. The prepared DNA library was loaded into the sample port in a dropwise manner.

The sequencing run was set up at the recommend voltage, for 72 hours and with basecalling and barcoding enabled, mid-read barcoding was also enabled. The minimum barcoding score was overrode to 60, and the minimum mid-read barcoding score set at 50.

2.4.7.3 NGS Analysis

Real time basecalling was performed using Guppy that is integrated within MinKNOW; the 'Fast' basecalling option was selected. Guppy was also used to demultiplex samples based on their "barcodes". The InterARTIC web application was utilised for post-run analysis which aligned the FastQ files generated by Guppy to a reference genome, where single nucleotide variations and indels can be detected relative to the reference genome (202). The Medaka workflow was chosen for variant detection, and consensus genome sequences are ultimately generated. The consensus fasta files (.bed files) were used for downstream analysis on IGV 2.13.0 to determine population percentages of bases at specific nucleotides.

2.5 Protein production

2.5.1 Protein expression

Small scale starter cultures were grown from stabs of glycerol stocks of *E. coli* Rosetta-2 cells transformed with pET28a-SUMO-CHIKV-nsP1 40 mL of LB medium supplemented with 1 $\mu\text{g}/\text{mL}$ kanomycin at 37°C for 16 hours. 10 mL of

these starter cultures was used to inoculate 1L LB medium and left to grow at 37°C until OD₆₀₀ reached 0.4 - 0.6. Expression of 6xHis-SUMO-CHIKV-nsP1 was induced with isopropyl β-D-1 thiogalactopyranoside (IPTG) at a final concentration of 0.2 mM. Cultures were then incubated at 30°C for 16 hours.

2.5.2 Protein purification

Bacterial cells were harvested by centrifugation at 5,000 x *g* at 4°C for 20 minutes using a SLC-6000 rotor. Pellets were either resuspended in lysis buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.8), 20 mM imidazole, 0.1 mM TCEP, 5 mM MgCl₂, 2 μg/mL DNase I, 2 μg/mL RNase A, 1 mg/mL lysozyme, 3 tablets (/80 mL EDTA-free protease inhibitor) or stored at -20°C. 20 mL of lysis buffer was used per 1 L of bacteria culture. Cell lysis was completed by using a cell disrupter. Lysed cells were clarified by centrifugation at 40,000 x *g* at 4°C for 20 minutes to separate the soluble and insoluble fractions. The supernatant (soluble fraction) was then filtered with a 0.45 μM filter.

2.5.3 Nickel affinity chromatography

The HisTrap HP 5 mL column was first washed with 20 % ethanol and then dH₂O to remove residual ethanol. The HisTrap column was equilibrated with binding buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.8), 20 mM imidazole, 0.1 mM TCEP) immediately prior to applying the supernatant of bacterial lysate. The supernatant was re-applied to the column twice more, at a flow rate of 1 mL/min and the final flow-through was collected. The HisTrap column was then washed with binding buffer with 20 mM imidazole to elute any non-specific binding. The HisTrap column was washed with 2 M NaCl/20 mM Tris-HCl (pH 8.8)/20 mM imidazole to remove any RNA. The column was washed once more with binding buffer with 50 mM imidazole before 6xHis-SUMO-CHIKV-nsP1 was eluted with elution buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.8), 300 mM imidazole, 0.1 mM TCEP). A final wash of elution buffer with 1 M imidazole was used to elute any protein that was still bound to the HisTrap column.

Following dialysis and cleavage of the eluted fractions (section 2.5.4) of Ni-affinity chromatography was repeated to remove the 6xHis-SUMO tag from pure CHIKV-nsP1. As before, the HisTrap column was first washed with dH₂O to remove residual ethanol and equilibrated with dialysis buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.8)). The mixture of cleaved and dialysed samples were applied onto the column at a flow rate of 1 mL/min and the flow-through collected. The column was washed with binding buffer containing 50 mM imidazole, 500 mM imidazole and finally 1 M imidazole to remove residual 6-His-SUMO tag from the column.

2.5.4 Dialysis and cleavage of fusion tag

The elution fractions were dialysed overnight in order to remove any excess imidazole to aid in SUMO-protease efficiency. The total protein concentration was estimated with Bradford reagent and absorbance at 595 nm. 0.5 mg SUMO-protease was added for every 10 mg of eluted protein. The mixture was decanted into a 10k MWCO dialysis tubing (Thermo Scientific) and placed in 5 L of dialysis buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.8), 0.1 mM TCEP) for 16 hours overnight at 4°C.

2.5.5 Size exclusion chromatography

The flow-through from the secondary HisTrap column was concentrated to 10 mL using centrifugal filters (Pierce™ Protein Concentrator PES, 10K MWCO) and injected into a equilibrated (ÄKTA buffer, 200 mM NaCl, 20 mM Tris-HCl (pH 8.8), 1 mM DTT, 5 % (w/v) glycerol) HiLoad 75 Superdex pg column (GE Healthcare) using an ÄKTA prime plus at a flow rate of 0.5 mL/min. Size exclusion chromatography was performed at 4°C and 2.5 mL fractions were collected after void volume (80 mL). The purity of CHIKV-nsP1 in eluted fractions was analysed by SDS-PAGE followed by Coomassie staining. Finally, pure CHIKV-nsP1 was concentrated to 10 mg/mL and stored at -80°C.

2.5.6 Circular dichroism

Circular dichroism was carried out to investigate secondary structure information of purified CHIKV nsP1. Purified nsP1 was diluted to a final concentration of 0.6 mg/mL in 20 mM Na₃PO₄ (pH 7.4), 100 mM NaCl, 0.5 mM DTT, 5% (w/v) glycerol and 5 mM MgCl₂. Circular dichroism was measured between 250 and 180 nm in a 1 mm path length cuvette using a Chirascan Plus CD spectrometer (Applied Photophysics) and was performed by Nasir Khan (University of Leeds).

2.5.7 Sparse matrix screening - X-ray crystallography trials

Sparse matrix screening of CHIKV nsP1 was carried out using the JCSG core suite screens I-IV (Qiagen). Sitting drops were set up in 96-well MRC-3 drop plates (Molecular Dimensions) using a Mosquito crystal Nanolitre protein crystallisation robot (TPP Labtech) with 30 µL reservoir solution. Unbound protein was concentrated to 9 mg/mL and dispensed into 0.2-3 µL droplets containing protein:reservoir ratios at 1:1, 2:1 and 1:2. Plates were sealed with Viewseal pressure adhesive clear seals (Grenier Bio-One) and incubated at 20°C.

2.6 Protein analysis

2.6.1 SDS Polyacrylamide Gel Electrophoresis (SDS PAGE)

SDS PAGE gels were made with a 12 % resolving gel (4 mL 30 % bis-acrylamide, 2.5 mL 1.5 M Tris-HCl pH 8.8, 3.3 mL ddH₂O, 100 µL 10 % SDS, 100 µL 10 % ammonium persulphate (APS), 10 µL TEMED) and 5 % stacking gel (0.83 mL 30 % bis-acrylamide, 0.63 mL 1 M Tris-HCl pH 6.8, 3.4 mL ddH₂O, 50 µL 10 % SDS, 50 µL 10 % APS, 5 µL TEMED). Protein samples were mixed 4 x LDS (lithium dodecyl sulphate) sample buffer (1 x LDS; 106 mM Tris HCl, 141 mM

Tris Base, 2 % LDS, 10 % glycerol, 0.51 mM EDTA, 0.22 mM G250 Coomassie Blue, 0.157 mM phenol red; pH 8.5), supplemented with 50 mM DTT reducing agent. Samples were heated at 95°C for 5 minutes to denature the proteins and allowed to cool before loading onto the gel. Samples were loaded alongside colour prestained, protein standard, broad range ladder (New England Biolabs). Electrophoresis was performed in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS) at 180 V for 50-80 minutes (until the dye front had run off the gel).

2.6.2 Protein Visualisation by Coomassie

Proteins that had been resolved by SDS PAGE were visualised by incubating gels in Coomassie stain (0.25 % (w/v) Coomassie R-250, 50 % (v/v) methanol, 10 % (v/v) glacial acetic acid) at room temperature for 1 hour. The gels were destained for 1 hour at room temperature using destain solution (40 % (v/v) methanol, 10 % (v/v) glacial acetic acid) and rehydrated in ddH₂O prior to imaging.

2.6.3 Western Blot Analysis

Proteins were transferred from the SDS PAGE gels to a fluorescence compatible polyvinylidene fluoride (PVDF) membranes (Immobolin-FL transfer membrane; Millipore) using a Trans-Blot semi-dry cell (Bio-rad) in Towbin buffer (25 mM Tris, 192 mM glycine, 20 % methanol) for 1 hour at 15 V. The membranes were blocked for 1 hour at room temperature in 50 % Odyssey blocking buffer (LiCor) in PBS. The membranes were then incubated with primary antibody diluted to the appropriate concentration in 50 % Odyssey blocking buffer in PBS, overnight at 4°C. The membranes were washed 3 x in PBS before incubating with the appropriate fluorescent secondary antibody diluted in 50 % Odyssey blocking buffer in PBS, for 1 hour at room temperature. The membranes were washed 3 x in PBS and allowed to dry before using a LiCor Odyssey Sa Infrared imaging system (LiCor) to directly visualise the membranes.

Chapter 3

Investigating the importance of M²⁴ across the CHIKV lifecycle

3.1 Introduction

3.1.1 The function of nsP1

There have been two primary functions characterised for nsP1: it is essential in the addition of the 5' cap on nascent viral RNA and for anchoring the replication complex to host membranes. Computational analysis of homologous structural and functional predictions of CHIKV nsP1 and closely related alphaviruses, was utilised originally to determine three domains in nsP1 (203). The first domain is at the N-terminal (NT domain) and is the domain that is involved in the methyltransferase (MTase) and guanylttransferase (GTase) activities during 5' capping of the CHIKV genome. The 5' cap is essential for recognition by host translational machinery, RNA stability and the evasion of the innate immune response (204). The middle domain (MB domain) is involved in anchoring nsP1, and in turn the replication complex, to host cellular membranes. The C-terminal domain (CT domain) has yet to be definitively associated with a function, however there is some evidence to suggest it is involved in an essential interaction with nsP3 (203). There is a palmitoylation site within the CT domain that has been suggested to aid in stabilising nsP1 binding to the membrane. Throughout this study the molecular structure of CHIKV nsP1 was resolved and published (113, 205). Both studies reported that 12 monomers of nsP1 interact with one another to act as a nuclear pore complex to control viral RNA replication. It was clearly demonstrated that nsP1 interacts with the host plasma membrane and that this

is important for the correct formation of the replication complex. The findings of these studies are discussed in more detail in section 5.3.

The capping mechanism (see figure 3.1) and cap type of alphavirus nsP1 differs to that most commonly utilised in eukaryotic cells, making this pathway and protein an attractive target for potential therapeutics. The first reaction within the alphavirus capping reaction is the methylation of GTP to m^7 GTP which is catalysed by nsP1; S-adenosylmethionine (AdoMet) is the methyl donor (which is demethylated to S-adenosyl-L-homocysteine (AdoHcy)). Next, nsP1 is guanylated by m^7 GTP, releasing a pyrophosphate (PPi). The RTPase domain of nsP2 dephosphorylates RNA from 5'-triphosphate RNA to 5'-diphosphate RNA. The final reaction is the transfer of m^7 GMP from nsP1 to 5'-diphosphate RNA, producing the final capped product: m^7 Gppp-RNA (204, 206). It is of note that this reaction has only been studied in SINV, SFV and VEEV, and it is therefore important to continue this research within CHIKV. In eukaryotes, the capping activity of RNA is carried out in a different order. Likewise, a diphosphate RNA intermediate is produced through the removal of the 5' γ -phosphate, however, GMP is transferred onto the 5' end of RNA prior to the methylation of the guanosine. Capping of alphavirus RNA occurs in the cytoplasm whereas capping of host RNA occurs in the nucleus.

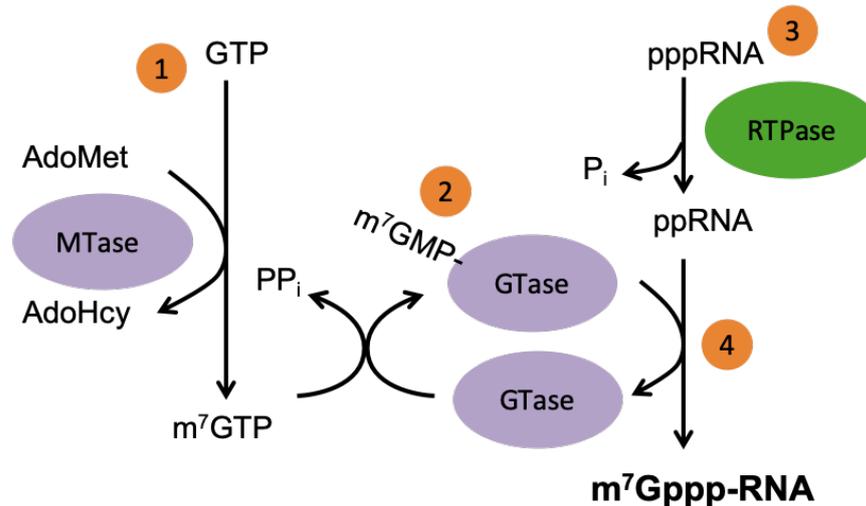


Figure 3.1: Capping mechanism of nsP1.

There are 4 main reactions involved in CHIKV RNA capping. **1.** The nsP1 MTase domain catalyses the methylation of GTP to m⁷GTP (S-adenosylmethionine = AdoMet, the methyl donor. S-adenosyl-L-homocysteine = AdoHcy). **2.** Guanylation of nsP1 by m⁷GMP, releasing pyrophosphate (PP_i). **3.** 5'-triphosphate RNA (pppRNA) is dephosphorylated by RTPase (nsP2) to 5'-diphosphate RNA (ppRNA). **4.** m⁷GMP is transferred onto 5'-diphosphate RNA to produce the final capped product (m⁷Gppp-RNA).

The formation of spherules at host plasma membranes and vacuoles is essential for productive alphavirus infection. This was first reported in 2000, whereby studies were performed on a synthetic peptide corresponding to the membrane binding domain of SFV nsP1 and the presence of anionic phospholipids resulted in a structural transition from a random coil to an amphipathic α -helix (68). Recent studies have reported a cryo-EM structure of CHIKV nsP1, showing that a dodecameric ring is formed of 12 nsP1 monomers (113, 205). This structure was found to interact with the host plasma membrane and revealed how oligomerisation, membrane binding and catalytic capping activity were coupled. It is proposed that nsP1 functions as a viral nuclear pore complex to control the transit of molecules in and out of spherules and is essential for the formation of the membrane-associated replication complexes.

In addition to the well characterised functions above, a palmitoylation site in nsP1 has been associated with the induction of short projections from the plasma membrane (70, 71). Real-time live imaging of fluorescently labelled-SINV

infection demonstrated the formation of filopodia which were associated with nsP1, and were observed in both mammalian and mosquito cells (120). Furthermore, the expression of nsP1 alone was sufficient to induce the formation of these filopodia, and the removal of the palmitoylation site in SINV and SFV inhibits filopodial extensions forming (207). The function of these projections remains unknown, but inhibition of nsP1 palmitoylation is associated with decreased pathogenesis and neurovirulence in mice (208).

Finally, the nsP1 palmitoylation site has been suggested as the interaction site between nsP1 and nsP4. There was an observed attenuation in negative-strand synthesis following the removal of the 3 aa palmitoylation site in nsP1, suggesting that there are interactions between nsP1 and the RdRp during genome replication (72). Replication and nsP1-nsP4 binding, but not palmitoylation, were rescued through compensatory mutations, suggesting that it is the amino acids and not palmitoylation required for efficient replication. Previous genetic analysis of SINV and SFV found that interactions between nsP1 and nsP4 are related to negative strand synthesis (209–211).

3.1.2 Project background and preliminary results

As discussed in section 1.3.5, our group has previously identified novel stem-loops within the 5' UTR and 5' coding region of ORF-1 that are required for efficient replication of the virus genome (173). Alongside this, our group also noted an in frame start codon (AUG) at amino acid residue 24 within ORF-1 (referred to as M²⁴ hereafter) that is unique to CHIKV and ONNV, and is not conserved within other closely related alphaviruses (figure 3.2).

was an alternative start codon that initiated the production of an N terminally truncated nsP1. In order to test this hypothesis they introduced a panel of mutations at M²⁴ (table 3.1). Three residues (M²⁴A, M²⁴Q and M²⁴S) were introduced to reflect residues seen at this position in other alphaviruses. A further two leucine mutations were introduced: M²⁴LI (CUG), as CUG is a known non-AUG start codon and M²⁴LII (CUC) is an alternate leucine that does not function as a start codon.

Table 3.1: Mutations introduced at M²⁴ in CHIKV ORF-1.

A summary of mutations introduced at M²⁴ to test the hypothesis that this amino acid functioned as an alternative start codon for a truncated version of nsP1.

| Mutation | Codon | Amino acid | Reason for choice |
|---------------------|-------|------------|---|
| Wild-type | AUG | Methionine | Wild-type CHIKV |
| M ²⁴ A | GCG | Alanine | Conserved in RRV |
| M ²⁴ Q | CAG | Glutamine | Conserved in VEEV, WEEV and SINV |
| M ²⁴ S | AGC | Serine | Conserved in SFV |
| M ²⁴ LI | CUG | Leucine | Alternative, non-canonical start codon |
| M ²⁴ LII | CUC | Leucine | Does not act as a non-canonical start codon |

Preliminary data utilised a sub-genomic replicon system and determined that introducing the panel of mutations at residue M²⁴ had no effect on genome replication in mammalian cells 24 hpt. However, when these mutations were introduced into the infectious CHIKV clone a reduction in virus replication by ≥ 2 logs relative to wild-type was observed following transfection of BHK-21 cells with ICRES (infectious clone of the La Reunion ECSA strain) RNA. This data suggested that M²⁴ was important in the virus lifecycle outside of genome replication as the sub-genomic replicon isolates genome replication from later stages of the virus lifecycle (such as virion assembly and egress). It was inconclusive whether these results were statistically significant due to a low number of repeats. Western blot analysis following infection of wild-type and mutant CHIKV demonstrated that a truncated form of nsP1 was produced during wild-type CHIKV infection but not following infection with the M²⁴A

mutant; supporting the hypothesis that M²⁴ acts as an alternative start codon to produce a truncated form of nsP1, and that mutations introduced at M²⁴ abolished its expression. This preliminary data suggested that nsP1 and M²⁴ have non-canonical functions beyond the currently known role in the capping of 5' genomic and sub-genomic RNA, such as virus encapsidation.

nsP1 is an attractive target for the development of novel antiviral therapeutics, as it is a multi-functional protein with essential roles, including 5' capping of the virus genome, and anchoring the CHIKV replication complex to the membrane (113, 205). Furthermore, nsP1 functions through a different mechanism to that of host methyltransferases so therapeutics against nsP1 are less likely to result in off target effects. The potential of novel functions within nsP1 provides a novel approach for the development of antivirals.

3.1.3 Aims

The aim of the work described in this chapter was to determine the importance of the M²⁴ residue in CHIKV nsP1 and whether it functions as an alternative start-codon to produce a truncated nsP1. A reverse genetics approach was utilised to introduce a panel of mutants at the site of interest. The mutants were then analysed to observe the impact on replication, in the context of infectious virus and replicon systems designed to isolate and measure different stages of the CHIKV replication cycle.

3.2 Results

3.2.1 Phenotype of M²⁴ mutations in infectious virus

As discussed previously, preliminary data from the group demonstrated that introducing mutations at M²⁴ resulted in a reduction in the production of infectious virus. Initial experiments were performed to confirm this data and to observe the phenotype of individual mutations across different cell types to determine whether

observed phenotypes were host specific. A reverse genetics approach was utilised in order to study the panel of 5 mutations at residue M²⁴ (see table 3.1) within the infectious CHIKV ICRES clone. The BHK-21 cell line is commonly used during CHIKV virus generation/experiments (212), as they are highly permissive to infection and allow the production of high wild-type viral titres. However, this cell line is not physiologically relevant to CHIKV infection as BHK-21 cells are not human or mosquito derived. The experiment was also performed in the human hepatoma derived cell line, Huh7, as the liver is a known replication site during human CHIKV infection (48) and this cell line has been previously shown to support CHIKV replication (212).

Full length wild-type and mutant ICRES RNA was *in vitro* transcribed and transfected into BHK-21 cells and Huh7 cells by electroporation. The supernatant of infected cells containing released CHIKV was harvested 24 hpt and stored at -80 °C. The viral titre of harvested supernatant was determined by plaque assay in BHK-21 cells, in order to confirm whether there was a reduced replication phenotype across the M²⁴ mutant panel (figure 3.3).

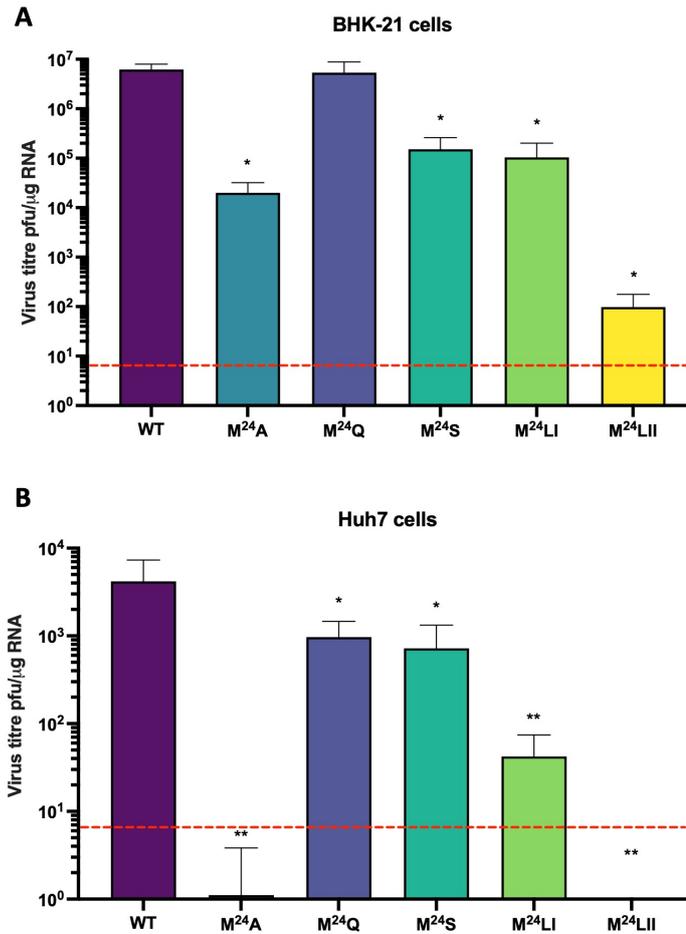


Figure 3.3: Wild-type and M²⁴ mutation viral titres following transfection of CHIKV RNA into BHK-21 and Huh7 cells.

Viral stocks of wild-type CHIKV and the panel of mutants were produced by transfecting BHK-21 cells with 1 μg RNA and the supernatant was collected 24 hpt. Virus titre in the supernatant was measured by plaque assay in BHK-21 cells. The red line denotes the limit of detection. **A** CHIKV titres generated in BHK-21 cells. **B** CHIKV titres generated in Huh7 cells. Error bars represent SEM and significance is shown when compared to the WT control. * = P<0.05, ** = P<0.01. n=3.

As mentioned, the BHK-21 cell line is known to be highly permissive to CHIKV infection and this is evident in figure 3.3, where wild-type virus generated in BHK-21 cells had viral titre that was ≈ 3 logs greater than the titre of wild-type virus generated in Huh7 cells. In BHK-21 cells, four of the five mutations showed a statistically significant reduction in infectious virus production, with M²⁴Q having a similar titre to that of wild-type CHIKV (figure 3.3A). In Huh7 cells, all 5 mutants showed a statistically significant reduction in infectious virus production (figure 3.3B). The reduction in viral titres across the mutant panel in both BHK-21 and Huh7 cells confirms that the M²⁴ residue is critical for

efficient CHIKV replication and that a reduced virus replication phenotype was observed regardless of the amino acid introduced at this residue, although the M²⁴Q mutation had no observed phenotype in BHK-21 cells.

In order to confirm that the observed phenotypes in figure 3.3 were not affected by any reversion mutations, infectious centre assays (ICA) were performed. This assay involves seeding transfected cells directly onto a naïve monolayer of BHK-21 cells and incubated under a layer of viscous media, similar to a plaque assay. The benefit of an ICA is that it reduces the opportunity for reversions/compensatory mutations which can be gained prior to harvesting virus to be selected across the whole culture. Plaques will only be formed by infectious virus that is initially released from successfully transfected cells. Full length wild-type and mutant CHIKV RNA was transfected into BHK-21 cells via electroporation, the cells were then serially diluted and plated onto naïve BHK-21 cells. Media was replaced after 2 hours with a 1:1 ratio of CMC:complete media, similarly to plaque assays, to ensure that secreted infectious virions could only infect neighbouring cells.

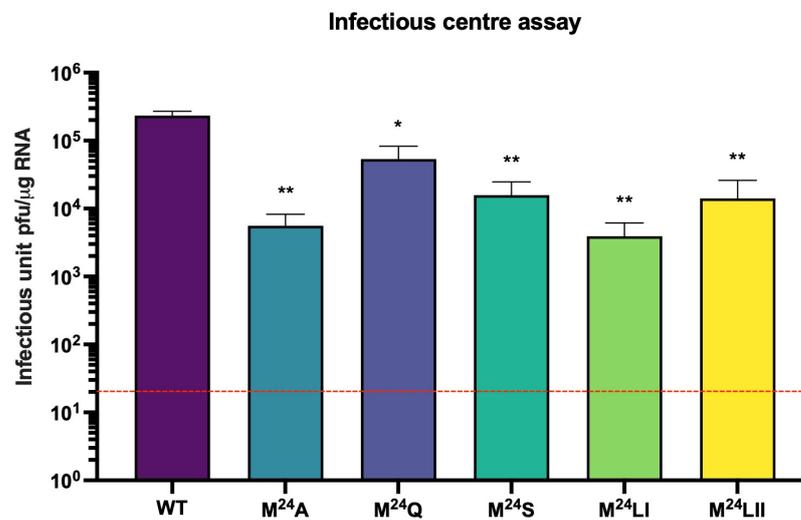


Figure 3.4: Infectious centre assay of M²⁴ mutations.

BHK-21 cells were transfected with 1 μ g of wild-type or mutant CHIKV RNA, serially diluted and plated onto a monolayer of naïve BHK-21 cells. Following a 72 hour incubation, cells were fixed and stained in order to visualise plaques. The red line denotes the limit of detection. Error bars represent SEM and significance is shown when compared to the WT control * = P<0.05, ** = P<0.01. n=3.

Differences between wild-type and mutant infectious units calculated from the ICA (figure 3.4) showed less variation than the differences in viral titres (figure

3.3). However, all 5 mutants were shown to have statistically significant reduced infectious units in comparison to wild-type levels. The M²⁴Q mutation was observed to have a reduced phenotype in the ICA, whereas the viral titre of M²⁴Q following virus production was similar to the wild-type titre, this suggested that there was selection of reversion/compensatory mutations during the 24 hour incubation following the transfection of BHK-21 cells. The phenotypes observed across the ICA aligned with those observed in the production of infectious virus, further confirming the importance of M²⁴ for efficient CHIKV replication. There was a significant reduction in infectious M²⁴Q virus produced in the ICA (figure 2.4.2) but not following standard transfection and harvesting (figure 3.3A), suggesting that reversion/compensatory mutations may have occurred. Sequencing of extracted cellular viral RNA would be required to confirm whether the M²⁴Q mutant was still present.

Following the observation of a reduced viral titre phenotype across the mutant panel, analysis of viral growth 24 hpi at a given MOI was performed to study the phenotype of the mutations during infection of Huh7 and RD cells (muscle derived). RD cells were chosen as CHIKV naturally replicates within muscle cells during CHIKV infection of humans (44, 45). Due to the low titres obtained from the generation of virus in Huh7 cells, viral stocks produced from BHK-21 cells were used for the following experiments. However, the mutant viral stocks were still low and consequently an MOI = 0.001 was the highest possible MOI that could be used in order to include M²⁴A, with the titre of M²⁴LII virus too low to be used at all. The cell lines were infected at an MOI = 0.001 with wild-type and mutant virus, the supernatant was harvested 24 hpi and productive virus infection determined by plaque assay in BHK-21 cells (figure 3.5 A and B). The infected cell monolayers were lysed 24 hpi and assayed by western blotting, probing against CHIKV nsP1 and actin (figure 3.5C).

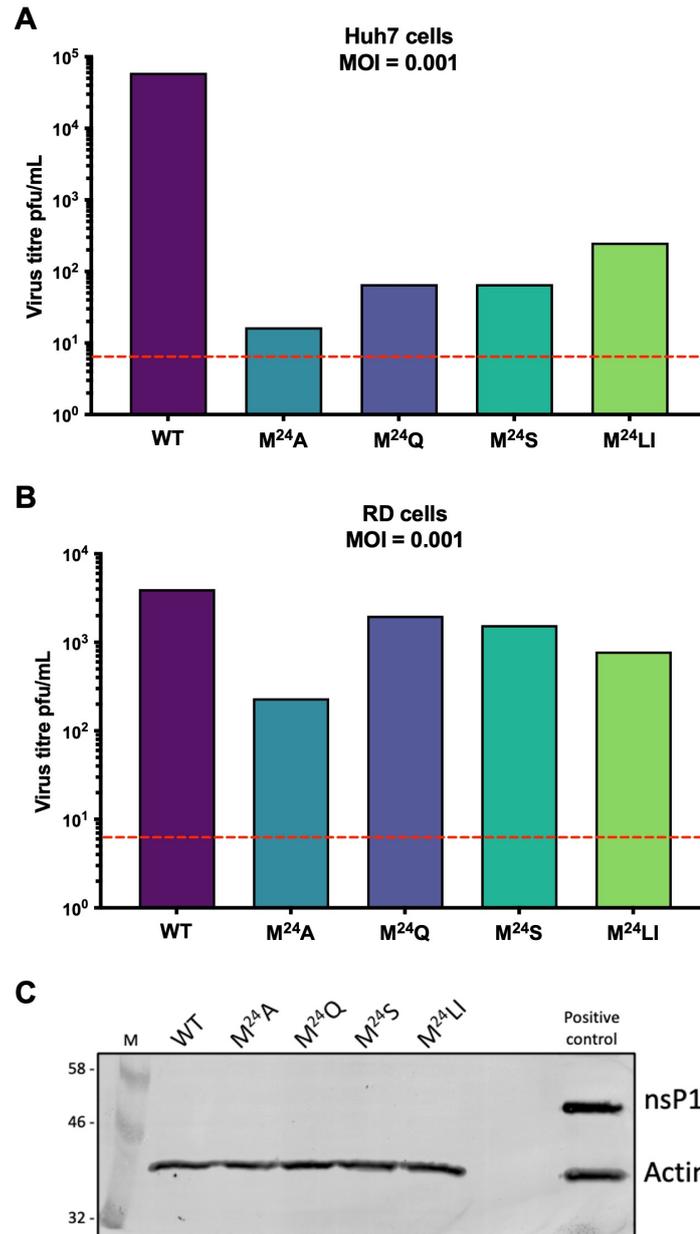


Figure 3.5: Viral titres of wild-type and M²⁴ mutations following infection at MOI = 0.001

RD and Huh7 cells were infected at MOI = 0.001 and supernatants collected 24 hpi, virus titre in the supernatant was measured by plaque assay in BHK-21 cells. The red line denotes limit of detection. **A** CHIKV titres following infection of Huh7 cells. n=1. **B** CHIKV titres following infection of RD cells. n=1. **C** Infected RD cells were lysed 24 hpi and analysed by western blotting, probing against CHIKV nsP1 and actin as a loading control. A positive control of wild-type CHIKV infected BHK-21 cells at an MOI = 1 was used. M denotes the size marker. n=1.

The viral titres of all 5 M²⁴ mutations following infection of both Huh7 and RD cell lines (figure 3.5) were lower than the wild-type titre. The difference between wild-type and mutant levels was greater in Huh7 cells than in RD

cells. The phenotype of a reduced viral titre following infection aligns with the phenotype observed when producing virus, further confirming that the M²⁴ residue is important for the CHIKV lifecycle. Overall, the viral titres following an MOI = 0.001 infection in Huh7 and RD cells were very low and so no nsP1 was detectable by western blot (figure 3.5 C). Due to this limitation and greater variability likely to be seen when calculating viral titres between 10 - 200 pfu/mL, an attempt to generate higher titre stocks of virus was performed, to allow for a higher MOI infection to be performed during equal MOI infections to compare wild-type and mutant viral growth.

To overcome the limitation of low mutant viral titres and the inability to detect nsP1 via western blot following an MOI = 0.001 infection, virus was generated in the mosquito cell line, C6/36 (*Ae. albopictus* derived). The *Ae. albopictus* species is the main vector involved with the spread of CHIKV in recent outbreaks meaning C6/36 cells are highly physiologically relevant. In addition, C6/36 cells have previously been shown to permit high levels of CHIKV replication (212). Full length wild-type and mutant CHIKV RNA was transfected into C6/36 cells using Lipofectamine 2000; the supernatant of infected cells containing released CHIKV was harvested 48 hpt and stored at -80 °C. Viral titres were determined by plaque assay, using the same method described previously following the transfection of BHK-21 and Huh7 cells.

The trend observed across the titres of wild-type and mutant virus produced in C6/36 cells (figure 3.6) was similar to the trend seen in BHK-21 and Huh7 cells (figure 3.3), yet the difference between wild-type and mutant virus titres was dramatically smaller. The M²⁴Q mutant produced virus with a similar titre to wild-type CHIKV, this was also observed in BHK-21 virus production. The titre of M²⁴S virus was slightly lower than the wild-type titre, however, it was not statistically significantly lower. The remaining three mutants (M²⁴A, M²⁴LI and M²⁴LII) all had statistically significantly lower titres than wild-type. The reduced viral titre phenotype of 3 of the 5 mutants further confirms the importance of the M²⁴ residue in CHIKV replication, and demonstrates that this was not cell type/host dependent. The overall levels of virus generated in C6/36 cells were far greater than previously observed in mammalian derived cells, allowing a higher

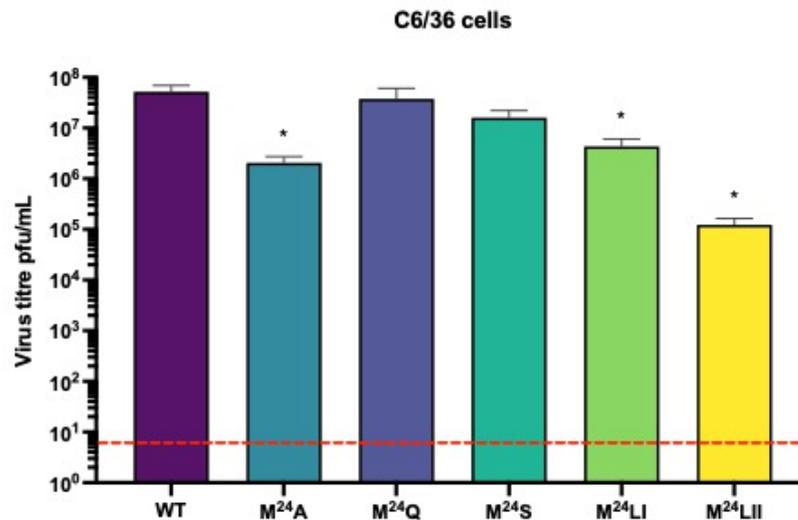


Figure 3.6: Wild-type and M^{24} mutation viral titres following transfection of CHIKV RNA in C6/36 cells.

Viral stocks of wild-type CHIKV and the panel of mutants were produced by transfecting C6/36 cells with 1 μ g RNA and the supernatant was collected 24 hpt. Virus titre in the supernatant was measured by plaque assay in BHK-21 cells. The red line denotes the limit of detection. Error bars represent SEM and significance is shown when compared to WT control. * $P < 0.05$, $n = 3$

MOI for infections going forward.

Equal MOI infections were performed in the same way as previously described, in Huh7, RD and C6/36 cell lines, at an MOI = 0.1, and viral titres analysed to observe the phenotype of M^{24} mutations throughout viral infection (figure 3.7). A similar trend was observed across all three cell lines, with Huh7 cell lines showing the most drastic reduction in mutant titres in comparison to the wild-type titre (figure 3.7 A). There was a significant reduction in all 5 mutant virus titres in comparison to wild-type titres in both Huh7 and C6/36 cell line infections. Whilst infection of RD cells with mutant virus (figure 3.7) showed a decrease in titre in relation to the wild-type titre, none of the decreased mutant titres were statistically significant.

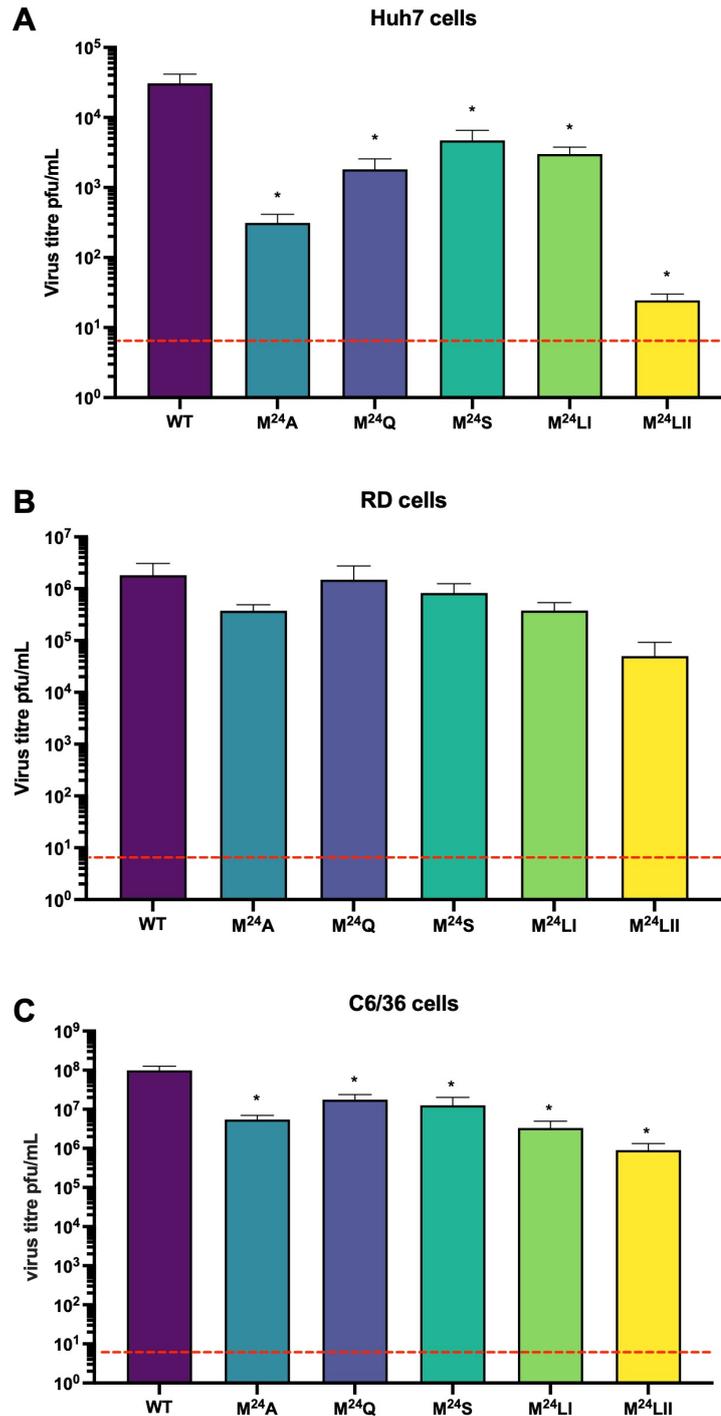


Figure 3.7: Viral titres of wild-type and M²⁴ mutations following infection at MOI = 0.1

Huh7, RD and C6/36 cells were infected at MOI = 0.1 and supernatants collected 24 hpi, virus titre in the supernatant was measured by plaque assay in BHK-21 cells. The red line denotes the limit of detection. Error bars represent SEM and significance is shown when compared to WT control. **A** CHIKV virus titres following infection of Huh7 cells. **B** CHIKV virus titres following infection of RD cells. **C** CHIKV virus titres following infection of C6/36 cells. * P<0.05, n=3

The observed reduced viral titre phenotype following viral growth analysis at equal MOI infection strongly suggests that residue M²⁴ is important for the efficient replication of CHIKV. In addition, this residue is not important in a host dependent manner as the phenotype is present in Huh7 and C6/36 cells. Whilst there was not a statistically significant reduction in productive virus replication observed in RD cells, the viral titre following an equal MOI infection was still reduced. Although this set of experiments was able to demonstrate that M²⁴ is required for CHIKV replication, the data did not allow us to determine what stage of the life cycle it was important for, i.e whether the introduction of M²⁴ mutations effected early stages (genomic/sub-genomic replication) or late stages (virion assembly, budding and egress) of the virus life cycle.

3.2.2 No truncated nsP1 expression observed during CHIKV replication in BHK-21 or Huh7 cells

Following infections at a higher MOI, western blot analysis of cell lysates was performed to confirm whether truncated nsP1 (consistent with translation initiation from M²⁴) could be observed in wild-type lysates, but not mutant virus lysates (other than M²⁴LI, where the leucine can act as a non-AUG start codon). Huh7 cells were infected at MOI = 0.1 with wild-type and all 5 mutant CHIKV virus, following incubation for 24 hours the cells were lysed and analysed by western blot for nsP1. Despite infecting at an MOI 100 fold greater than previously, there was no nsP1 band detected at 60 kDa, even in wild-type CHIKV infected Huh7 cells (figure 3.8A). Furthermore, the anti-nsP1 antibody is very non-specific and there were a multitude of bands present in both virally infected and mock cells. The supernatant from the infected cells was collected and the titre was determined for each virus, confirming that the infection was successful. This suggests that an MOI = 0.1 is still too low for detection of nsP1 in Huh7 cells.

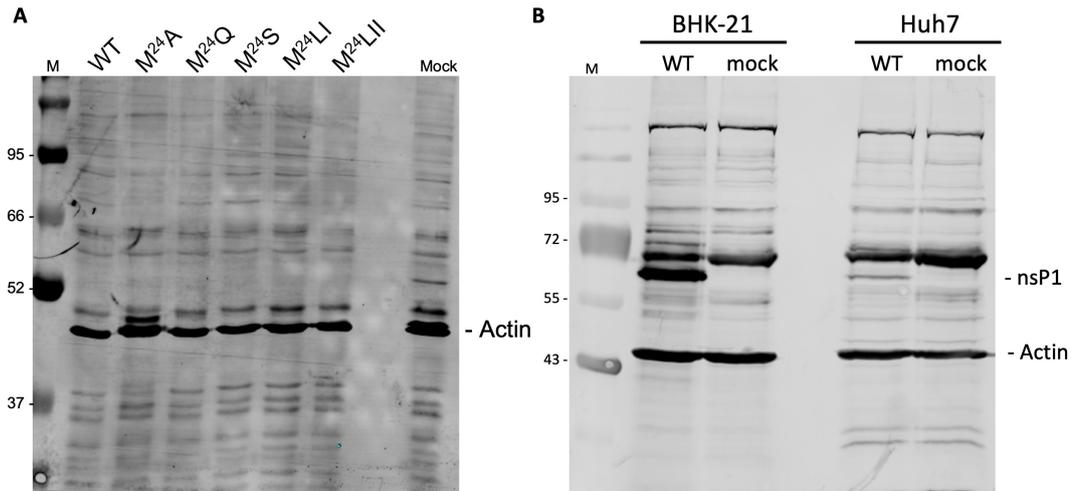


Figure 3.8: Western blot analysis of CHIKV infected cell lysates

Cells infected with wild-type and mutant CHIKV were analysed by western blot to detect a truncated nsP1. Anti-actin was used as a loading control, anti-nsP1 was used to probe against nsP1. **A** Huh7 cells were infected at MOI = 0.1 with wild-type and mutant CHIKV. Cells were harvested 24 hpi. **B** BHK-21 and Huh7 cells were infected at an MOI = 1 and the monolayer was harvested 24 hpi.

A further experiment was performed comparing the level of nsP1 expressed during a CHIKV infection in Huh7 cells and BHK-21 cells and to further confirm whether a truncated version of nsP1 was detectable. The full-length nsP1 is 60 kDa and the truncated version was expected to be 57 kDa. Huh7 and BHK-21 cells were infected with wild-type CHIKV at MOI = 1, and cell lysates were harvested 24 hpi. A band corresponding to nsP1 (at 60 kDa) was observed in the infected lane in both BHK and Huh7 cells, and absent in the mock infected lanes (figure 3.8). There was much lower expression of nsP1 in Huh7 cells compared to BHK-21 cells. As discussed, the anti-nsP1 antibody bound to a variety of non-specific bands in the both viral and mock infected lanes. There was a non-specific band, just below the 72 kDa marker, present in all infected and mock-infected lysates that had high affinity for the anti-nsP1 antibody. There was no observed truncated nsP1 in either BHK-21 or Huh7 wild-type CHIKV infected cells, suggesting that M²⁴ did not function as a start codon for the synthesis of a truncated version of nsP1.

3.2.3 Introduction of mutants at M²⁴ does not impact CHIKV RNA levels during infection

To determine whether the M²⁴ residue is important for CHIKV genomic replication, strand specific qPCR was performed, to compare the levels of positive and negative sense RNA between wild-type and the M²⁴ mutant panel. Huh7, RD and C6/36 cells were infected with wild-type and mutant virus at MOI = 0.1, the monolayer was harvested with TRI-reagent 24 hpi and total RNA was extracted. RNA was reverse transcribed using primers specific to the positive and negative strands and the cDNA products were used as templates for qPCR. Strand-specific qPCR was chosen over standard qPCR so that the difference in positive- and negative-strand RNA levels could be assessed.

Despite the M²⁴ mutations having a reduced viral titre phenotype, there was no difference observed when comparing wild-type genome copies to the genome copies of all five M²⁴ mutants (see appendix figure 7.1). The ratio of positive:negative strand RNA was analysed from the qPCR data to assess whether introducing mutations at M²⁴ affected the production of the negative-sense RNA. Whilst there was fluctuation between the wild-type ratio and the ratios across the mutant panel (figure 3.9), none showed statistical significance in comparison to the wild-type positive:negative strand ratio.

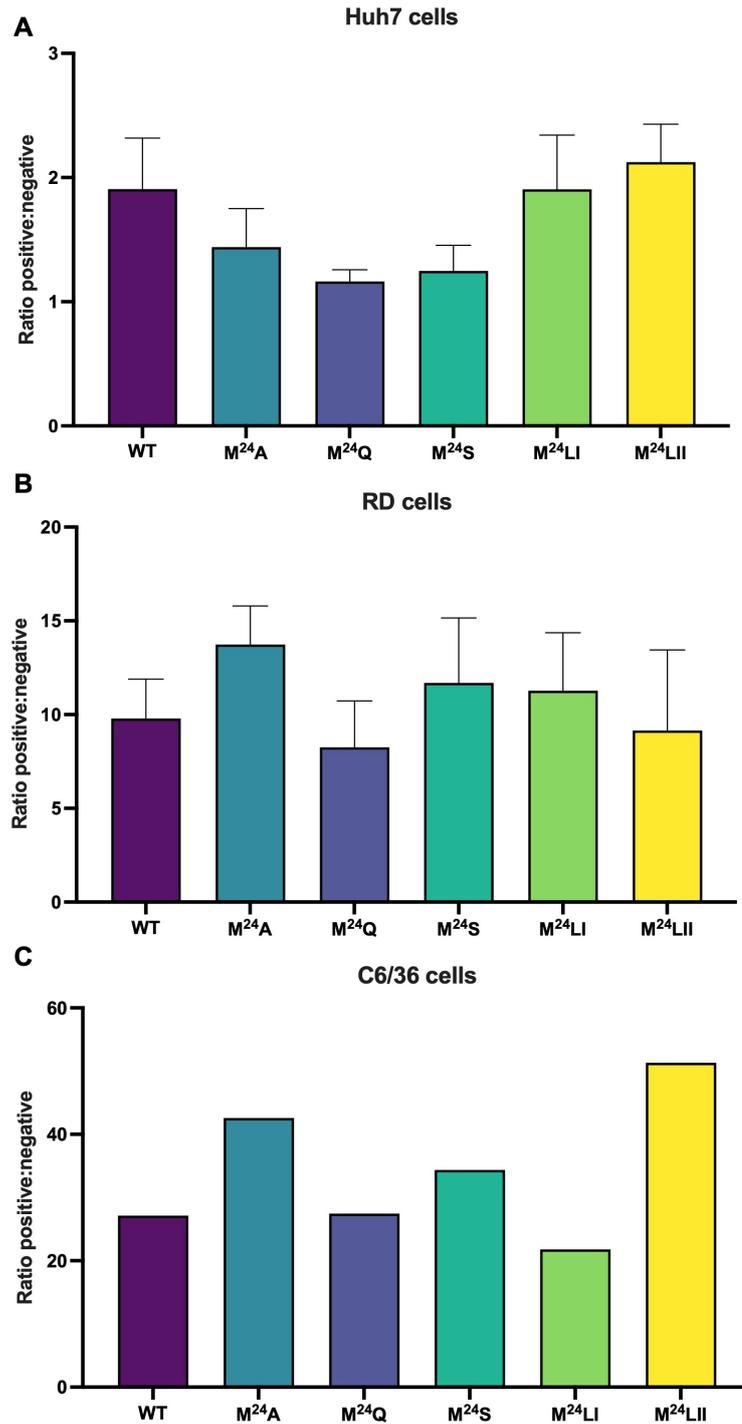


Figure 3.9: Ratio of positive to negative strand RNA levels in wild-type and M²⁴ mutant viral infections

Huh7, RD and C6/36 cells were infected with wild-type CHIKV and the M²⁴ mutant panel, at an equal MOI = 0.1. RNA was extracted from the monolayers 24 hpi with TRI-reagent. The RNA was reverse transcribed using primers specific to positive and negative sense CHIKV RNA. Strand-specific cDNA was used as a template for qPCR and a 94 bp region of nsP1 was amplified with strand specific primers (213). The ratio of positive to negative strand RNA was calculated. Error bars represent SEM. **A** Data from Huh7 cells, n=3 **B** Data from RD cells, n=3. **C** Data from C6/36 cells, n=2.

3.2.4 Mutations at M^{24} do not inhibit CHIKV genome replication

In order to investigate the role of M^{24} further a sub-genomic replicon system was used. This system allows us to isolate CHIKV genome replication and translation from other stages of the CHIKV lifecycle. In the sub-genomic replicon system the structural protein encoding region in ORF-2 was replaced with a firefly luciferase reporter gene (figure 3.10). A replication deficient dual-luciferase sub-genomic replicon, GAA, was used as a negative control. This construct contains two aspartic acid to alanine substitutions within the active site of the CHIKV RdRp, causing the polymerase to be inactive and therefore unable to replicate. The dual-luciferase construct also contains the renilla luciferase reporter gene in ORF-1 to measure translation, but only firefly luciferase was measured throughout this thesis. The panel of sub-genomic replicon mutants were transfected into cells, along with wild-type and GAA sub-genomic replicon RNA, the monolayer of cells were lysed at given time points and firefly luciferase expression measured. Firefly luciferase data was normalised to the 0 hpt time point, to ensure that the observed effects were due to true differences between wild-type and mutant levels, as opposed to any immediate translation of RNA following transfection.

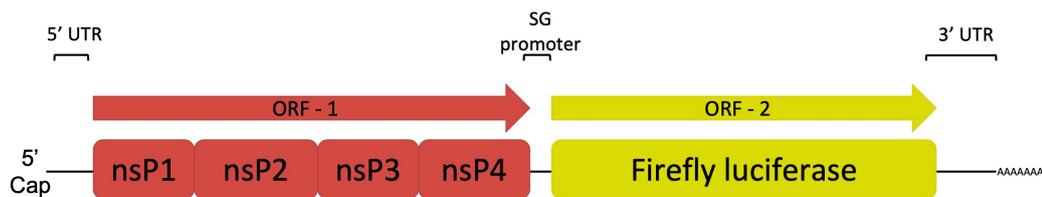


Figure 3.10: Firefly luciferase sub-genomic replicon

A schematic of the genome organisation of the firefly sub-genomic replicon. The structural genes in ORF-2 are replaced with a reporter gene, firefly luciferase.

Mutant sub-genomic replicon replication was assessed in Huh7 and BHK-21 cells, and replication was measured at 4 and 24 hpt. There was a slight variation in signal between wild-type and the five mutants at 24 hpt in Huh7 cells (figure 3.11A); with all mutants, except M^{24} LII, having a slightly higher luciferase signal than wild-type. However, these differences were not significant. The same trend

was observed at 4 hpt, with no significant difference in the luciferase expression between wild-type and the mutants present. In BHK-21 cells there was a slight reduction in luciferase signal observed between wild-type and mutant constructs at 24 hpt (figure 3.11B), however this reduction was not significant. At 4 hpt, $M^{24}A$ and $M^{24}LI$ showed a ≈ 1 log reduction in firefly luciferase expression, and $M^{24}LII$ displayed a ≈ 2 log reduction in luciferase; again, none of these were statistically significant. These results suggested that the presence of M^{24} is not a requirement for enhanced genome replication.

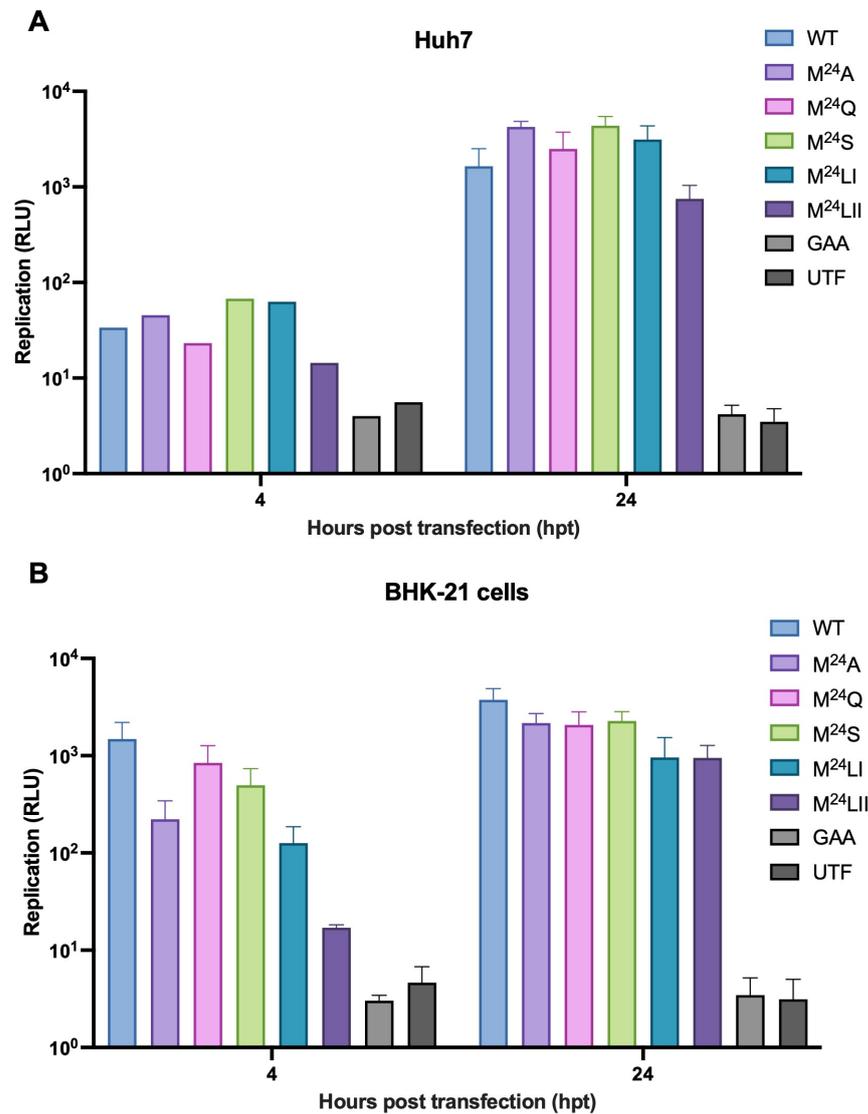


Figure 3.11: Phenotype of M^{24} mutants in CHIKV sub-genomic replicon

Cells were transfected with wild-type, mutant and GAA sub-genomic replicon RNA. Firefly luciferase signal was measured at 0, 4 and 24 hpt, data was normalised to 0 hpt. GAA denotes the inactive RdRp mutant. UTF denotes the mock transfected control. Error bars represent SEM. All data $n=3$, apart from 4 hpt Huh7 data $n=2$.

3.2.5 Substitutions at M²⁴ do not inhibit the function of nsP1

In order to further investigate which stage of the replication cycle was being affected, a *trans*-replicase system was used to interrogate the impact of introducing mutations at residue 24 specifically on the replication of the CHIKV genome. The system relies on two plasmids (figure 3.12A); the first is HSPoll-CHIKV-FG, which will be referred to as the CHIKV reporter throughout the thesis, and the second is CMV-P1234, which will be referred to as the CHIKV replicase. The CHIKV reporter contains two luciferase genes, firefly and *Gaussia* luciferase (Fluc and Gluc respectively), flanked by the CHIKV 5' UTR and the first 77 amino acids of nsP1 (N77), and the CHIKV 3' UTR, in addition, the CHIKV sub-genomic promoter lies between the firefly and *Gaussia* genes. N77 is included in order to include the replication elements that have been demonstrated to be essential for CHIKV genome replication (173, 174). It is of note that the CHIKV replicase construct was codon optimised in order to disrupt RNA secondary structures, but maintain the correct amino acid sequence. Fluc expression measures viral genomic replication, and Gluc expression measures viral sub-genomic transcription levels. The CHIKV replicase plasmid encodes the 4 CHIKV non-structural proteins downstream of a CMV promoter, allowing for the expression of CHIKV P1234 and subsequent cleavage events for the formation of the CHIKV replication complex. Similar to the sub-genomic replicon system, there is a CHIKV replicase construct that contains two aspartic acid to alanine substitutions in the RdRp active site (referred to as GAA replicase), resulting in a replication defunct replicase complex that is used as a negative control.

Briefly, upon transfection into cells, the replicase complex (nsP1-4) is translated from the CHIKV replicase construct by host machinery due to the CMV promoter (figure 3.12B). There is a baseline level of transcription of the CHIKV reporter RNA by RNA polymerase I. This RNA is uncapped and therefore there is minimal translation of either reporter gene (214, 215). The replicase complex recognises the 5' region of the reporter RNA and replicates the RNA, also generating sub-genomic RNA in the process. Firefly and *Gaussia* luciferase are expressed and

measured by luciferase expression following cell lysis.

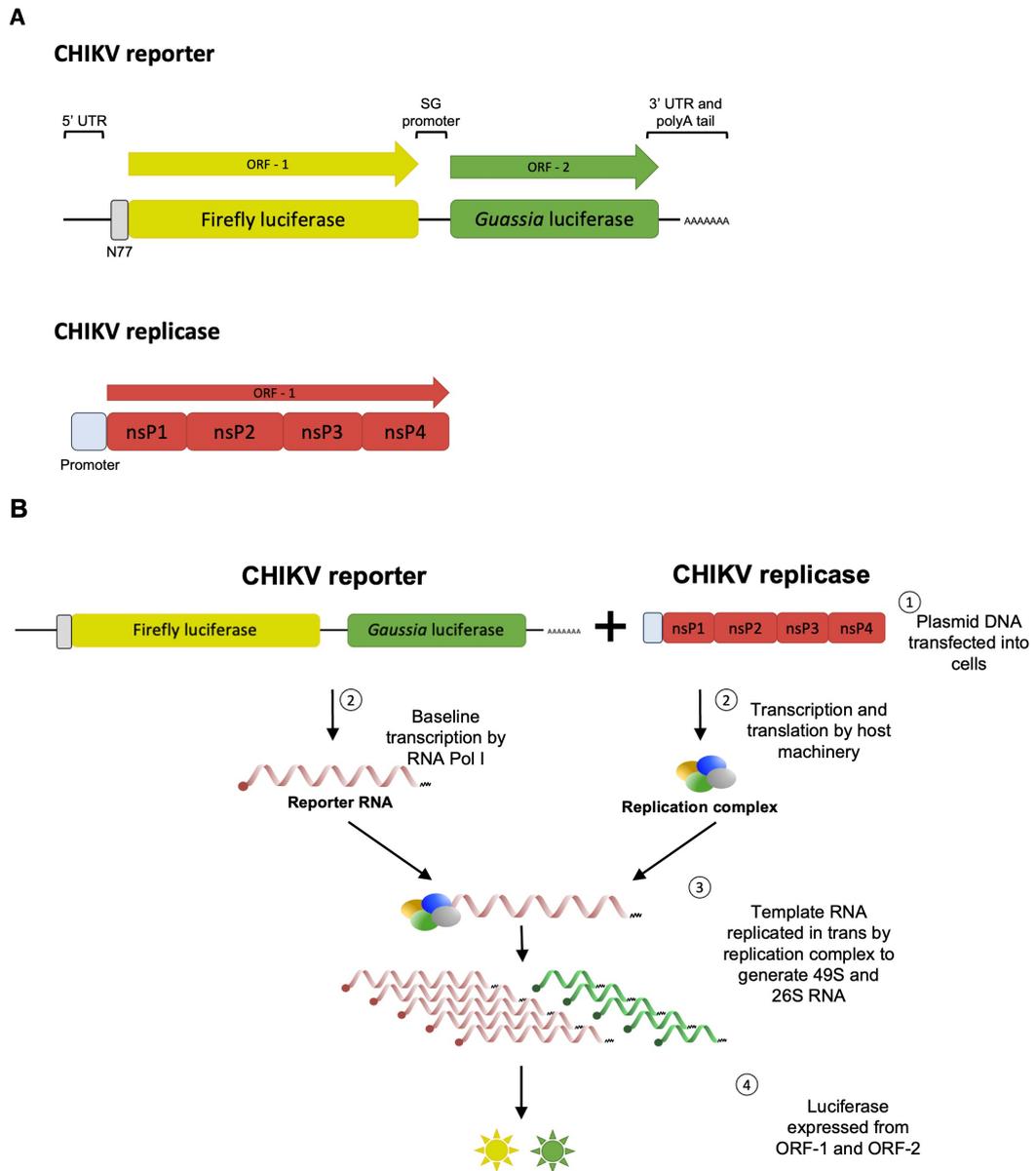


Figure 3.12: *Trans*-replicase assay schematic

A The *trans*-replicase system is comprised of two plasmids; the CHIKV replicase plasmid encodes the CHIKV non-structural polyprotein (P1234) downstream of a CMV promoter. The CHIKV reporter plasmid encodes the CHIKV 5' UTR upstream of a firefly luciferase reporter, followed by the CHIKV sub-genomic promoter upstream of a *Gaussia* luciferase reporter and the CHIKV 3' UTR. **B** **1.** Plasmid DNA is transfected into cells. **2.** The host machinery transcribes and translates the non-structural polyprotein from the CMV promoter present in the CHIKV replicase plasmid, ultimately forming the CHIKV replication complex. A baseline of transcription of reporter RNA is performed by RNA Pol I. **3.** The replication complex recognises the CHIKV 5' UTR and transcribes and replicates the CHIKV reporter RNA. **4.** The two luciferase reporters are expressed from the CHIKV reporter RNA and luciferase signal can be measured.

This system was chosen in addition to the sub-genomic replicon because mutations could be introduced at the residue of interest in both the CHIKV replicase plasmid, and the CHIKV reporter plasmid, due to the inclusion of the first 77 nucleotides of nsP1 in the 5' region of the CHIKV reporter. This allows us to determine whether the introduction of mutants at M²⁴ has a detrimental effect on the function of nsP1 within the replicase complex, or whether the nucleotide sequence is important in the RNA template. Importantly, this assay was also highly sensitive for measuring different stages of genome replication isolated from translation, which was not possible in the sub-genomic replicon system used in section 3.2.4.

Initial experiments were performed to confirm that the *trans*-replicase assay was a reliable system. Both the wild-type CHIKV reporter and replicase plasmids were transfected into Huh7 cells, using Lipofectamine-2000. Huh7 cells were chosen as this was the cell line that showed the greatest reduction in M²⁴ mutant viral titres in comparison with wild-type CHIKV. The GAA replicase construct was used as a negative control for luciferase expression. The wild-type reporter was transfected in stand-alone to confirm that the presence of the replicase was required in *trans* for increased luciferase expression. Cells were lysed at 0, 24 and 48 hpt and luciferase expression was measured. As shown in figure 3.13, the wild-type replicase and wild-type reporter produced robust replication at 48 hpt with Fluc and Gluc levels reaching an average of 8×10^4 and 6×10^5 RLU, respectively. The three negative controls (GAA replicase + wild-type reporter, wild-type reporter only, and UTF) all showed a significant reduction in Fluc and Gluc levels when compared to wild-type levels at 48 hpt. Of note, the inclusion of wild-type reporter alone confirmed that the wild-type replicase construct was required in *trans* to produce increased levels of Fluc and Gluc. Peak luciferase expression was observed at 48 hpt, thus time points in following experiments were taken to 48 hpt.

Following confirmation that the *trans*-replicase assay was reliable, the M²⁴A mutation was introduced into both plasmids, referred to as M²⁴A replicase and M²⁴A reporter. If the presence of a methionine at amino acid residue 24 was important for efficient nsP1 function, we expected to see a decrease in luciferase expression when transfecting the M²⁴A replicase in conjunction with wild-type

reporter. If the nucleotide sequence itself was important within the context of the RNA template, we expected to see a decrease in luciferase expression when transfecting the M²⁴A reporter alongside wild-type replicase.

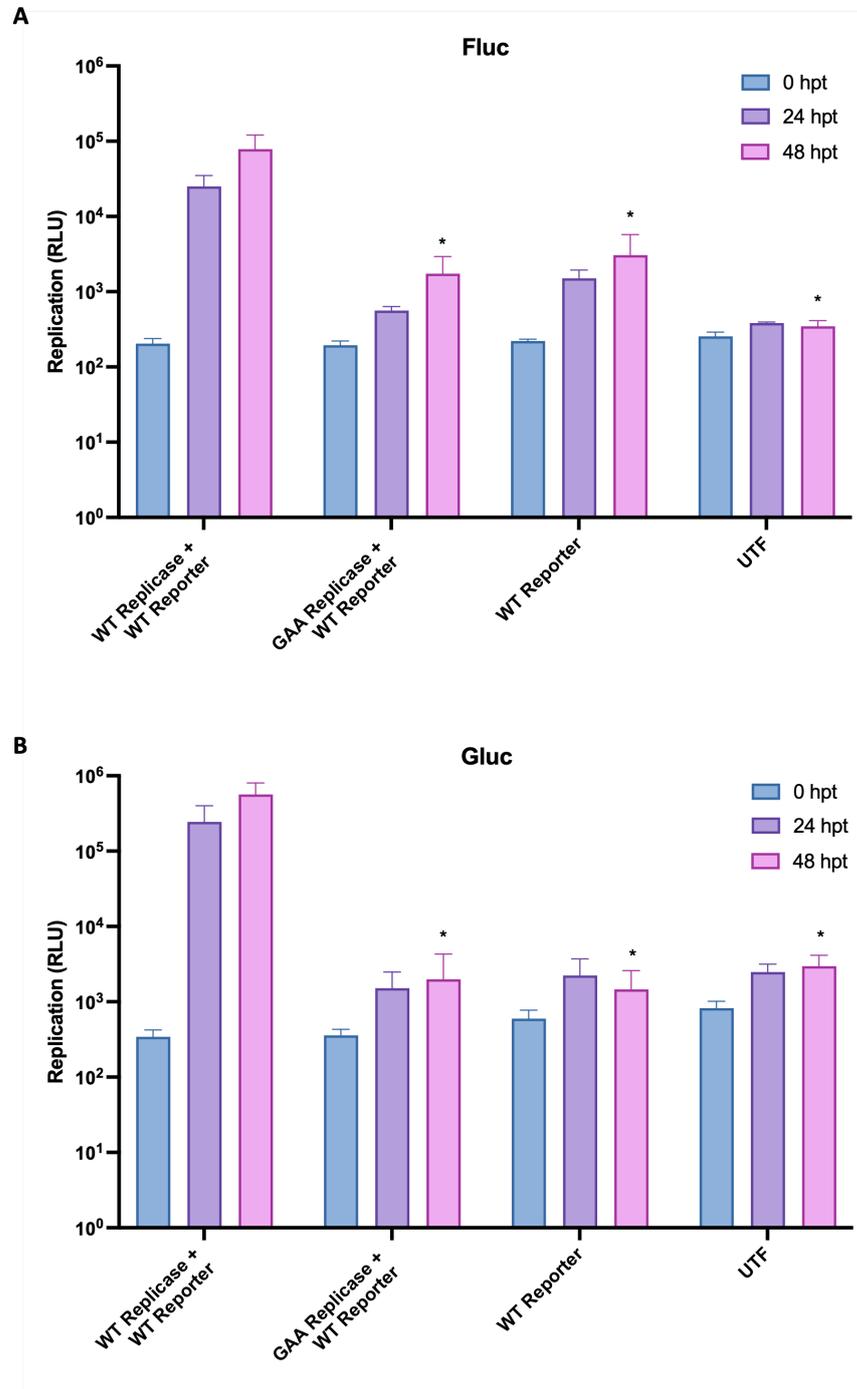


Figure 3.13: *Trans*-replicase assay development

Initial results to confirm that the *trans*-replicase system is reliable in producing increased luciferase expression from wild-type plasmids in comparison to GAA replicase, reporter only, and untransfected cells. **A** Fluc levels increase from 0 - 48 hpt. **B** Gluc levels increase from 0 - 48 hpt. Error bars represent SEM and significance is shown when compared to WT control. * P<0.05, n=3.

Fluc and Gluc expression levels were normalised to the negative control, GAA replicase (which has an inactive nsP4 (RdRp)) and the fold change is presented in figure 3.14. The raw luciferase data is shown in appendix figure 7.2. The introduction of M²⁴A in the replicase construct did not have a negative impact on either genomic or sub-genomic replication (figure 3.14). There was a slight increase in fold change with both luciferase activities when compared to wild-type levels, however, this increase was not significant. Interestingly, this trend was not observed when the M²⁴A mutation was introduced into the reporter construct. Conversely, there was a significant reduction in Fluc and Gluc levels in comparison to wild-type (> 50 fold and > 500 fold respectively). This data suggested that mutating M²⁴ to an alternative amino acid is not detrimental to the function of nsP1, but the nucleotide change is detrimental to the function of the RNA as a template for replication.

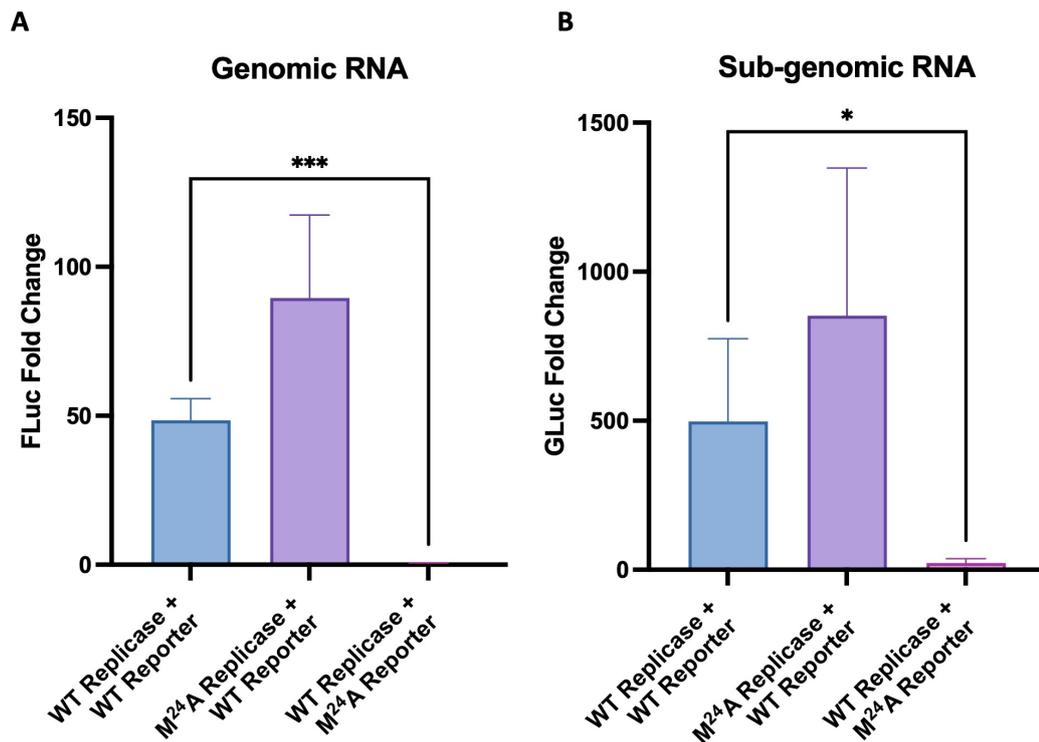


Figure 3.14: The phenotype of M²⁴A in the *trans*-replicase assay

The impact of introducing an alanine at residue 24 was analysed in the context of CHIKV nsP1's function, and the importance of nucleotide sequence in the RNA template. Fluc and Gluc activity levels were normalised to GAA replicase levels. **A** Fluc levels, corresponding to viral genomic RNA replication. **B** Gluc levels, corresponding to sub-genomic transcription levels. Error bars represent SEM and significance is shown when compared to WT control. * P<0.05, *** P<0.001 n=3

3.3 Discussion

Preliminary data from our group had previously demonstrated that introducing mutations at residue M²⁴ did not have an effect on CHIKV replication, however, an adverse effect was observed in infectious virus titres. This data suggested that nsP1 also has functions outside of its established roles during CHIKV genome replication and translation. The project first aimed to validate this data and continue to investigate the potential non-canonical functions of nsP1 through understanding the importance of residue M²⁴ at different stages of the virus life cycle.

This chapter has demonstrated that the introduction of five different amino acids at residue M²⁴ results in a significant reduction of infectious virus generated following either transfection of cells with RNA or mutant virus infection. This phenotype is not host specific, as the trend was observed in both mammalian and mosquito host cells. There was a slight reduction in viral titres across the mutant panel in RD cells, but the decrease was not as large or significant as observed in Huh7 cells. M²⁴LII displayed the greatest reduction in virus titre in RD and Huh7 cells, with a ≈ 1 log and ≈ 3 log decrease observed respectively. It has been previously shown that the differentiation of Huh7 cells results in reduced CHIKV replication (212), this may be why Huh7 cells displayed a generally reduced viral titre when compared to C6/36, RD and BHK-21 cells.

C6/36 cells were shown to be more permissive to CHIKV infection and were therefore used for the generation of higher titre viral stocks, allowing for a more suitable MOI to be used during subsequent infections. CHIKV is an arbovirus, and therefore has a lifecycle in mosquitoes, as well as humans. There is little to no viral-induced cytopathic effect (CPE) throughout infection with CHIKV, and no apparent cell death. This allows for the continuous production/replication of CHIKV with no adverse effects on cell health, and therefore the generation of high virus titres. This reflects the fact that in nature mosquitoes are the reservoir host and infection is systemic and life-long, and hence not lethal to mosquitoes.

In order to test the hypothesis that residue M²⁴ acts as an alternative start codon

to produce an N terminally truncated nsP1, a leucine (M²⁴LI) with the nucleotide sequence CUG, was introduced. This codon is known to have the potential to act as a non-AUG start codon (216) and hence should compensate for the function of M²⁴ as an alternative start codon for the synthesis of a truncated nsP1. It is therefore of note that M²⁴LI had a reduced viral titre phenotype following virus generation in BHK-21 and C6/36 cells, and following infection of Huh7, RD and C6/36 cells. This data suggests that the methionine is not functioning as an alternative start codon. On the other hand, CUG has a reduced efficiency in initiating translation or may simply not be functioning as a non-AUG start codon in this context, so it is possible that the presence of a leucine at residue 24 is not able to rescue the functionality of M²⁴ as an alternative start codon.

To confirm that any effects seen in the M²⁴LI mutant were not due to the introduction of a leucine at this residue, a second leucine was also introduced (M²⁴LII, AUG > CUC) that is not capable of initiating translation. Despite both the M²⁴LI and M²⁴LII mutants containing a leucine at residue 24, the M²⁴LII mutant displayed a greater reduction in virus titre following infection of Huh7 and C6/36 cells, and this mutant had the lowest virus yield when generating virus in BHK-21, Huh7 and C6/36 cells. It may be that the codon for M²⁴LI (CUG) does partially compensate for the function of the methionine to initiate translation of a truncated nsP1. Alternatively, the nucleotide change may be impacting the RNA secondary structures in the 5' region and in turn effecting genome replication, rather than the amino acid substitution debilitating the enzymatic activity of nsP1.

Western blot analysis was used to assay for a truncated nsP1 produced by wild-type CHIKV. Huh7 cells are not as efficient at replicating CHIKV as BHK-21 and C6/36 cells, and this resulted in difficulty detecting nsP1 following an MOI = 0.001 and 0.1 infection. It was possible for full length nsP1 to be detected when performing an MOI = 1 infection in Huh7 cells, however the band was much fainter when compared to BHK-21 cells that were treated the same (figure 3.8). Infections could be not carried out with mutant virus at this MOI due to the titre of viral stocks. There was no truncated nsP1 band observed in wild-type CHIKV infected BHK-21 or Huh7 cells, further supporting that M²⁴ does not

function as an alternative start codon to generate a truncated nsP1. For further confirmation of the results observed by western blot analysis, cell lysates could have been analysed by mass spectrometry.

Whilst the data is consistent with no truncated nsP1 produced during wild-type CHIKV infection, it was clear that introducing mutations at the M²⁴ residue was detrimental to the efficient replication of CHIKV. Next, the aim was to determine where in the lifecycle the residue was required; during genome replication, sub-genomic transcription or virion assembly and egress.

It was surprising to discover that despite viral titres differing between wild-type CHIKV and the mutant panel, there was no reduction in positive or negative sense genome copies observed by qPCR analysis. This initially suggested that nsP1 had non-canonical functions beyond its recognised roles in genome replication and translation. However, it may be that the impact on genome replication caused by introducing mutations at M²⁴ had recovered by 24 hpt, and further experiments with earlier time points should be conducted to discern whether the levels of genomic/sub-genomic RNA are reduced within mutant virus.

An alternative approach to analyse the impact of genome replication in M²⁴ mutant CHIKV is through by radiolabelling nascently transcribed RNA molecules. Radioactive uridine (³H uridine) has been used successfully to efficiently and specifically visualise the production of new RNA strands in CHIKV infected cells (90). First cells are pretreated with actinomycin D, stopping cellular RNA synthesis via inhibition of host cell RNA polymerases. This means the only RNA being produced in the cell should be newly synthesised CHIKV genomic and sub-genomic RNA. By chasing actinomycin D treatment with ³H uridine, we can see the newly generated CHIKV RNA as the ³H uridine is incorporated. This protocol was used to high effect by Gao *et al.* to show how mutation to the CHIKV nsP3 alphavirus unique domain effected sub-genomic RNA production but not genomic RNA (90). The same approach could be undertaken to further investigate the impact of genomic and sub-genomic RNA in the M²⁴ mutants.

Two different replicon systems were used to further interrogate the role of M²⁴ in genome replication; the sub-genomic replicon and the *trans*-replicase assay.

The sub-genomic replicon system is a measure of the combined effects of genome replication and translation, whereas the *trans*-replicase assay further isolates translation from genome replication. Similar to the qPCR, the phenotype observed within live virus was not reflected in the sub-genomic replicon system. The sub-genomic replicon data showed slight variation between wild-type and mutant constructs at 24 hpt in both Huh7 and BHK-21. Greater fluctuation in luciferase levels were observed at 4 hpt in BHK-21 cells but were not statistically significant at n=3. The firefly luciferase levels were normalised to the 0 hpt time point to allow for comparison between experiments performed on separate dates, despite this there was an outlier within the wild-type 4 hpt data set that was a log lower than the other 2 repeats. When this value was excluded from the analysis, significance was observed between wild-type and the three mutants M²⁴A, M²⁴LI and M²⁴LII (* P<0.05) however because these values were then only n=2 this significance cannot be confidently reported. Further experiments should be performed to confirm whether the reduction in mutant replication levels at early time points relative to wild-type levels is significant. An additional 12 hpt timepoint could also be investigated to observe how luciferase expression alters over time; previous work from the group introduced mutations into the 5' region of the CHIKV genome and observed a significant reduction in luciferase expression at 6 and 12 hpt, when compared to wild-type, but by 24 hpt levels were similar to wild-type (173).

The *trans*-replicase assay showed that introducing an alternate amino acid at residue 24 did not impact the activity of nsP1, as luciferase levels from constructs with the M²⁴A mutation introduced into the replicase remained similar to wild-type levels. On the other hand, introducing this mutation into the RNA template resulted in a significant reduction in both genomic replication and sub-genomic transcription. This data, coupled with no detection of a truncated nsP1 following wild-type CHIKV infection, suggests that the M²⁴ is not acting as an alternative start codon to produce a truncation form of nsP1 that is essential for CHIKV replication. The *trans*-replicase system allows a more in-depth insight into any effects on genome replication as the replicase cannot bind in *cis* to the RNA molecule that was used for the translation of the replication complex, which

increases sensitivity to small changes in the RNA template. The increased sensitivity may explain why this system showed that introducing a different nucleotide sequence at nucleotides 146-148 inhibited genome replication, where the qPCR and sub-genomic replicon data did not.

Future experiments should be performed utilising the *trans*-replicase system to interrogate the M²⁴ panel of mutants in insect cells. Work should be carried out in both C6/36 (*Ae. albopictus*) and Aag2 cell lines (*Ae. aegypti*) to determine if the effects of the mutations are the same across the two different mosquito vectors. It would be interesting to analyse the effects of the 5 different mutants across mammalian and insect cell lines and whether there are any amino acids introduced that are more tolerated by CHIKV than others.

The findings in this chapter lead to a new working hypothesis: the M²⁴ residue is important for efficient genome replication of CHIKV and mutations introduced at this location affect the function of the RNA molecule (perhaps through alterations to the RNA secondary structures), but not the efficiency of nsP1's function within the replicase complex.

Chapter 4

SHAPE-ing M²⁴ mutations in the CHIKV genome

4.1 Introduction

The previous chapter demonstrated that introducing mutations at M²⁴ decreased the level of infectious virus released following infection of mammalian and mosquito cells compared to wild-type CHIKV (figure 3.7) and that mutations at this amino acid affected the function of the RNA molecule rather than the function of nsP1 within the replicase complex (figure 3.14). However, the exact mechanism through which these mutations were inhibiting replication remained unknown. The results led to a new hypothesis that the nucleotides at M²⁴ were important in the RNA molecule for efficient genome replication of CHIKV and therefore further work was carried out to determine a functional mechanistic interactions, such as RNA:protein or RNA:RNA interactions.

4.1.1 CHIKV replication elements

It is well understood that the presence of secondary structures in positive-sense viral genomes are important for efficient virus replication and roles in genome replication (170), translation (153), genome packaging (155, 165) and evasion of the host-immune response (172) have been well characterised (discussed in detail in 1.3.5.1). There have been numerous studies performed to characterise the RNA structures across alphaviruses, however these were mostly on VEEV, SINV and SFV, and this knowledge was used to make assumptions about CHIKV.

The length of the 5' UTR varies between across the genus (20 - 85 nts) and even between different strains of a species (174, 185). There is a stem loop at the beginning of the 5' UTR (SL3) that is conserved across the alphaviruses, although the sequence and size varies; for example, the SINV stem loop contains an exposed bulge on the 3' side of the stem that is not present in CHIKV (173, 177). Despite the variation in UTR length and sequence there is a conserved dinucleotide (AU) at the beginning of the 5' UTR which is essential for genome replication and is believed to be involved with the binding of nsP4 to the 3' end of negative-sense genomes for positive-strand synthesis (171, 181).

There is a 51 nt CSE present across the alphaviruses within the 5' coding region of nsP1, that is known to be structured (figure 4.1). Both the sequence and structure of this region are highly important for efficient CHIKV replication. Silent mutations that disrupted the sequence but maintained base-pairing of one stem loop resulted in reduced growth rates in comparison to the wild-type virus (177) and disruption of both stem loops was non-viable whereas virus was recovered when at least one of the two stem loops was present (184). It was suggested that this region was involved with host-specific interactions as a greater reduction in replication was observed in mosquito cells compared to mammalian cells when this region was disrupted (177, 184). The stem of SL165 is comprised of 9 nts in CHIKV, SINV and VEEV, and 7 nts exposed in the terminal loop. Alternatively, the stem of SL194 in SINV is shorter than that of SL194 in CHIKV and VEEV (8 nts rather than 9 nts respectively) and the loop larger (6 nts rather than 4 nts respectively) (figure 4.1) (217).

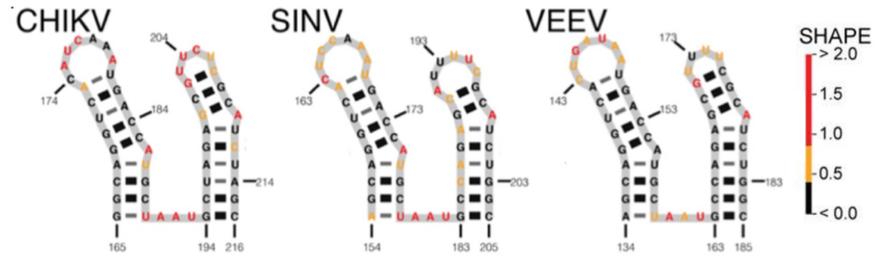


Figure 4.1: SHAPE-constrained thermodynamic folding prediction of the 5' 51 nt CSE in alphaviruses

SHAPE-constrained model shown for CHIKV, VEEV and SINV. Stem loops correlate to SL165 (left) and SL194 (right) according to the labelling scheme from Kendall et al.. Nucleotide colour corresponds to SHAPE reactivity scale. High reactivity (> 1 , red) denotes unpaired nucleotides, low reactivity (< 0.5 , black) denotes base-paired nucleotides. Adapted from (217).

There have been two recent publications reporting the importance of RNA elements specifically in CHIKV (173, 174). Kendall et al. used SHAPE to map the RNA structures in the 5' 300 nts of the CHIKV genome, and through structure-led reverse genetics identified six RNA replication elements essential for efficient CHIKV replication (figure 4.2).

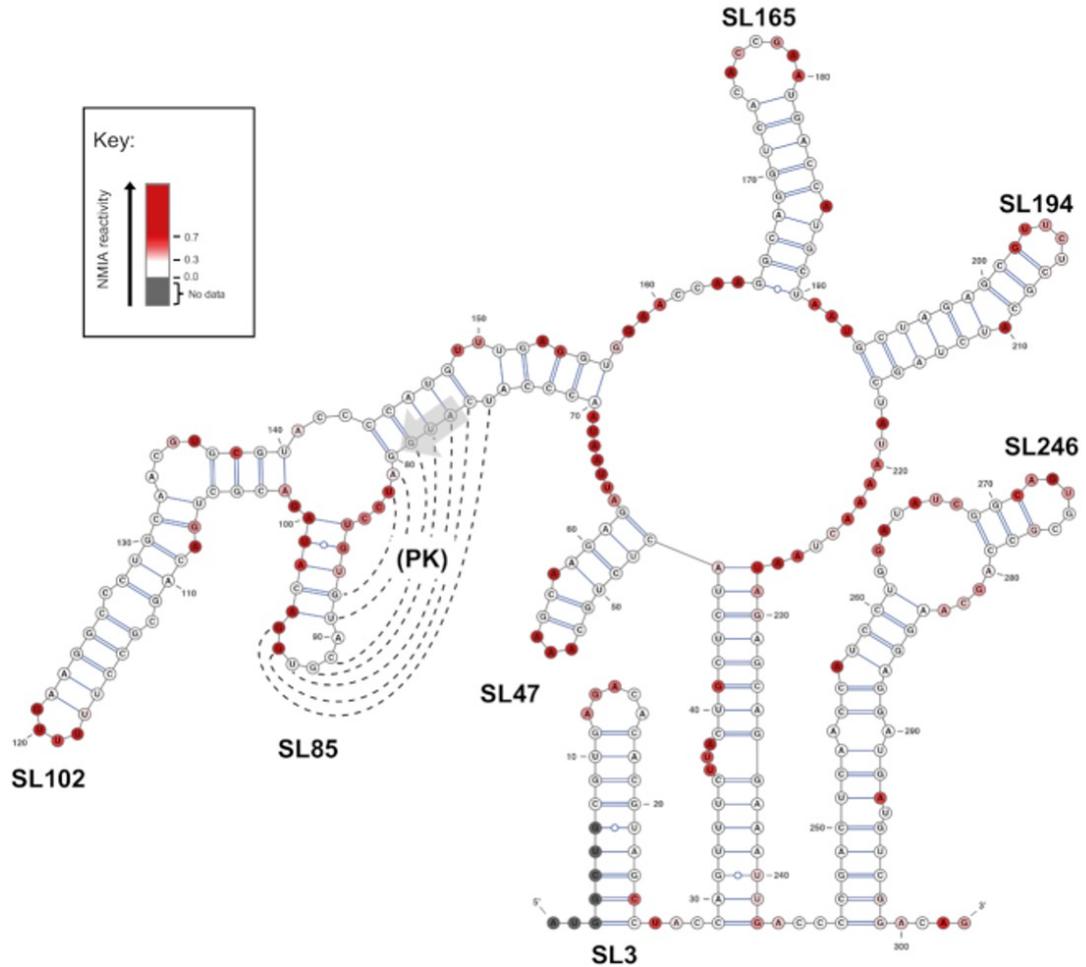


Figure 4.2: SHAPE-constrained thermodynamic folding prediction of CHIKV 5' 300 nts.

Minimum free energy model generated using SHAPE-directed constrains, with SHAPE reactivities overlaid. Nucleotide colour corresponds to SHAPE reactivity scale. Predicted stem loops labelled. The AUG start codon is denoted by a grey arrow. Adapted from (173).

As discussed in 1.3.5.2, the 5' CSE structures in VEEV and SINV are known to enhance replication in both mosquito and mammalian cells. Conversely, Kendall and colleagues reported that disruption of SL165 and SL194 in CHIKV decreased replication in human cells but not mosquito cells, suggesting that this region was involved in host-dependent interactions. In further experiments, they went on to demonstrate that the structure and sequence of SL194 was important and that the exposed nucleotides on the terminal loop are a CHIKV specific motif, whereas only the structure of SL165 was important.

In addition, they reported a novel stem loop in the 5' UTR (SL47), upstream of

the ORF-1 AUG start codon and that this structure was essential for CHIKV replication in both mammalian and mosquito cell lines. This structure is highly conserved across CHIKV strains, ONNV and Mayaro virus, with homologous structures observed in SFV and RRV. More distant viruses, such as SINV, VEEV and EILV, do not contain a sequence that would form a structure similar to that of SL47, suggesting that this structure is important for CHIKV specific RNA synthesis. A second novel stem loop was discovered within the nsP1 coding region, deemed SL246, that was shown to be conserved across the genus. Disruption of SL246 reduced replication in mosquito cells, but not human cells, suggesting that this region is important for host-specific interactions. The introduction of structure led compensatory mutations to restore base-pairing recovered virus replication to wild-type levels, demonstrating the structural importance of SL47 and SL246, rather than sequence, for efficient genome replication.

Finally, the structure-led reverse genetics analysis determined that the region downstream of the ORF-1 AUG in CHIKV contained two stem loops (SL85 and SL102), whereas a long arm has previously been observed in SINV (169). Kendall et al. reported that whilst the SHAPE mapping of the SL85 region was inconsistent with the *in silico* predicted model, reverse-genetic analysis to disrupt and then restore base pairing in this structure confirmed the presence and importance of SL85. They propose that this region is highly dynamic with further/alternative interactions involved in tertiary structures (such as a pseudoknots) potentially forming. Both SL85 and SL102 were shown to be structurally important for efficient CHIKV replication in mammalian cells, and not mosquito cells, suggesting they are involved in host-dependent interactions.

Another report was published whereby SHAPE-MaP was utilised to inform modelling of RNA structures across the entire CHIKV genome, and 23 regions with high structural stability were identified. One of these regions included the 5' CSE and the 70 nts upstream of this. Only a single stem loop was observed in the region upstream of the CSE (SL102) which contained the ORF-1 AUG start codon and was homologous to structures observed in SINV (169) (figure 4.3). The two conserved stem loops of the CSE (SL165 and SL194) were also identified. It was reported that SL102 was important for efficient replication in

both mammalian and mosquito cells, but that the magnitude of effect is greater in mammalian cells. Whereas disruption of base-pairing within the CSE resulted in reduced replication rates in both mammalian and mosquito cells. Furthermore, when SL102 and the CSE stem loops were disrupted in conjunction there was no virus recovered, demonstrating the importance of both RNA elements for CHIKV replication. It was also demonstrated that the disruption of SL102 inhibited RNA synthesis, but not protein synthesis. Further mutations were introduced that would disrupt the sequence of SL102 and the 5' CSE but allow base-pairing to maintain the predicted structures and these replicated to wild-type levels, indicating that it is the structure of these regions and not the sequence that is necessary for efficient genome replication.

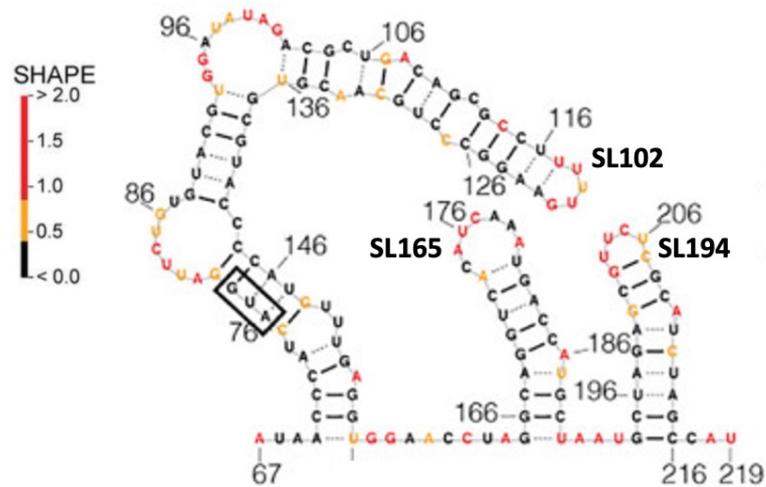


Figure 4.3: SHAPE-constrained thermodynamic folding prediction of CHIKV SL102 and 5' CSE

SHAPE-constrained model shown for CHIKV region 70 - 219. Stem loops are labelled according to the nomenclature from Kendall et al. Nucleotide colour corresponds to the SHAPE reactivity scale. High reactivity (> 1 , red) denotes unpaired nucleotides, low reactivity (< 0.5 , black) denotes base-paired nucleotides. The AUG start codon is denoted by a black box. Adapted from (174).

Despite the difference between predicted RNA secondary structure models from Kendall et al. and Madden et al., both groups demonstrated the importance of SL102, SL165 and SL194 during CHIKV replication. Furthermore, both models predict that M^{24} and the ORF-1 AUG have two nucleotides base-paired and therefore it was important to determine whether the mutations introduced in the M^{24} mutant panel (table 3.1) altered the RNA secondary structure, as this could

potentially impact virus genome replication.

4.1.2 Aims

The overall aim of this chapter was to establish a greater understanding of the importance of the M²⁴ codon (nucleotides 146 - 148 (AUG)) within the context of the RNA molecule throughout the CHIKV lifecycle and elucidate potential mechanistic interactions of this region in genome replication.

1. SHAPE analysis to produce SHAPE constrained minimum free energy models to confirm whether the RNA secondary structure was altered following mutation of M²⁴.
2. CHIKV passaging assay to generate selection of pseudo-escape mutations were gained to help determine potential mechanistic interactions for nucleotides 146 - 148. These mutations could occur one of three ways:
 - (a) Direct reversion (e.g. 146^{GCG} > 146^{AUG}).
 - (b) Substitutions to restore base-pairing and secondary structures (e.g. 146^{GCG} > 146^{GUG}).
 - (c) Compensatory mutations elsewhere in the genome to restore interactions with this region: (e.g. in nsP4).
3. To confirm whether mutations observed in passaging experiments were:
 - (a) Sufficient to restore wild-type levels of virus replication.
 - (b) Cell type specific.

4.2 Results

4.2.1 SHAPE analysis of the M²⁴ mutant panel

As explained earlier in sections 1.3.5 and 4.1 the CHIKV RNA replication elements of the 5' UTR and 5' coding region of nsP1 have been characterised and it has been

demonstrated that disruption of these structures reduces viral genome replication (173, 174). Here, we utilised biochemical methods to determine whether the reduction in CHIKV replication observed in chapter 3 across the M²⁴ mutant panel was due to alterations in RNA secondary structures at the 5' region of the CHIKV genome. Prior to SHAPE analysis, *in-silico* predictions were performed to discern that the M²⁴ mutants disrupted the RNA structures (appendix figures 7.4 - 7.9).

A biochemical SHAPE mapping approach was utilised to provide an RNA structure map of the CHIKV 5' \approx 300 nts region for WT and M²⁴ mutant CHIKV. The full protocol is described in Materials and Methods 2.2. Briefly; folded RNA was incubated with NMIA, which chemically modified non base-paired nucleotides to form a 2'-O-adduct. A negative control reaction was performed using DMSO in place of NMIA to control for early termination products produced during reverse transcription. An RNA sequencing ladder was also generated that did not undergo folding or NMIA/DMSO treatment, to align the sample sequence to. The chemically modified and negative control/sequencing ladder RNA was used as a template for a reverse transcription reaction, with fluorescent primers (FAM and HEX for the sample and sequencing ladder respectively). When the reverse transcriptase reached the NMIA adduct, extension would be terminated, resulting in a truncated cDNA copy. The pool of cDNA fragments was purified, combined with a sequencing ladder and analysed by fluorescently labelled size fragment capillary electrophoresis at the Dundee DNA Sequencing facility. The use of different fluorescent primers, and different capillaries (for NMIA and DMSO), resulted in different retention times which can be aligned during computational analysis to align the signal between the sample (FAM) and RNA sequence ladder (HEX). The tool, QuSHAPE, was used to perform the computational analysis on the sequencing data and create an output file of relative SHAPE reactivities at individual residues (218). This data was entered into the Vienna RNAprobing server, where it used the SHAPE reactivity data as pseudo energy constraints to predict a minimum free energy model of RNA secondary structure. In-depth detail of the analysis pipeline is shown in appendix figure 7.3.

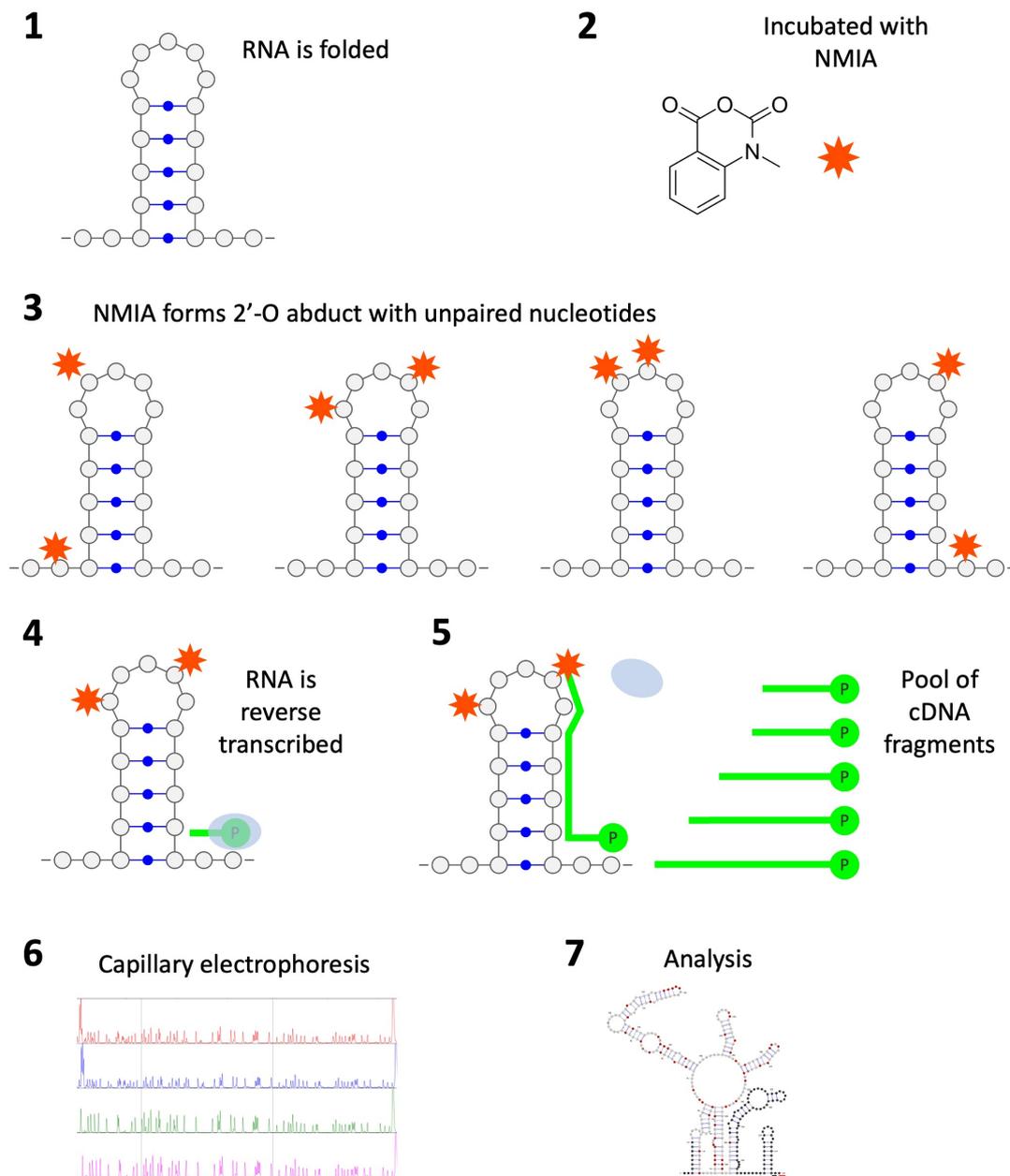


Figure 4.4: Schematic of the SHAPE protocol

1 *In-vitro* transcribed RNA is folded in folding buffer at 37 °C. **2** Folded RNA is incubated with NMIA (orange star). **3** NMIA binds with unpaired nucleotides to form a 2'-O-adduct. **4** Treated RNA is reverse transcribed using fluorescently labelled primers (HEX and FAM). **5** cDNA transcription is terminated when the polymerase (blue circle) reaches NMIA labelled nucleotides, generating a pool of cDNA fragments. **6** cDNA is analysed by capillary electrophoresis, alongside DMSO controls and sequencing ladders. **7** The raw sequencing files are analysed using QuSHAPE to determine SHAPE reactivity scores. Vienna RNAprobing server is used to predict base pairings. VARNA is used to visualise RNA structures.

Analysis was performed on wild-type CHIKV, alongside all five mutations analysed in chapter 3. The most notable differences between wild-type and mutant

structures were observed to be between nucleotides 67 - 219. Therefore, for simplicity only this region is shown in the SHAPE maps presented in this chapter; SHAPE maps for the entire region (nucleotides 1 - 337) are shown in the appendix (figure 7.10).

The wild-type structure that was predicted following SHAPE analysis closely matched the published structure from Madden *et al* (174), with an elongated arm culminating in SL102 and did not contain SL85, instead, a bulge of non-paired nucleotides was seen with nucleotides 83 - 87 and 143 - 147 (figure 4.5B). The presence of SL165 and SL194 were also observed in the wild-type structure. It is of note that this wild-type structure did not display base pairing between the first two nucleotides in M^1 and M^{24} that was observed in both the Kendall and Madden models.

The introduction of an alanine at amino acid residue 24 (in nsP1 (AUG > GCG)) resulted in a change of RNA structure and the number of stem loops observed in the region between nucleotide 70 - 219. Figure 4.6B shows that whilst the structure was disrupted, both SL102 and SL194 were conserved. SL165 was no longer present and instead there was a short stem loop from residue 137 to 146, and another longer stem loop from nucleotides 154 to 177. SL85, that was described in the Kendall model was present in the structure prediction for M^{24} . The base pairing observed between the AU nucleotides of the M^1 and M^{24} codon in published models had been disrupted within the M^{24} mutant and nucleotides 146-148 were no longer in the vicinity of the initial start codon.

The RNA structure was also disrupted in the $M^{24}Q$ (AUG > CAG) mutant, with the number of stem loops observed between nucleotides 70 - 219 reduced compared to wild-type (figure 4.7B). Once again, SL102 and SL194 were formed in the predicted structure, whilst SL165 was no longer present. As with the $M^{24}A$ model, there was a stem loop observed from nucleotides 154 - 177 rather than the SL165 that was observed in the published models and the wild-type structure shown in figure 4.5.

There was no significant observed change of structure in the $M^{24}S$ mutant (figure 4.8B) compared to the predicted structure for wild-type CHIKV. All three stem

loops (SL102, SL165 and SL194) were formed. The large single stranded bulge that was observed in the wild-type structure from nucleotide 82 is not present in the $M^{24}S$ structure and instead there were two smaller bulges of unpaired bases (bases 82 & 83 unpaired with bases 148 & 149, and base 87 unpaired with 143 & 144).

The structures predicted for $M^{24}LI$ and $M^{24}LII$ were very similar (figures 4.9B and 4.10B respectively). The structures SL102, SL165 and SL194 were observed in both $M^{24}LI$ and $M^{24}LII$. There was an additional stem loop observed between nucleotide 145 - 155 in the two structures; this stem loop contained the nucleotide that was mutated for $M^{24}LI$ and $M^{24}LII$ (nucleotide 146 A > C or G respectively). Furthermore, this loop differed in size between the two leucine mutants; $M^{24}LI$ had 5 bases in the terminal loop whereas $M^{24}LII$ had 3 bases. The arm that contains SL102 started at residue 78 in the two leucine mutants, whereas it started at nucleotide 72 in the wild-type structure (figure 4.5B).

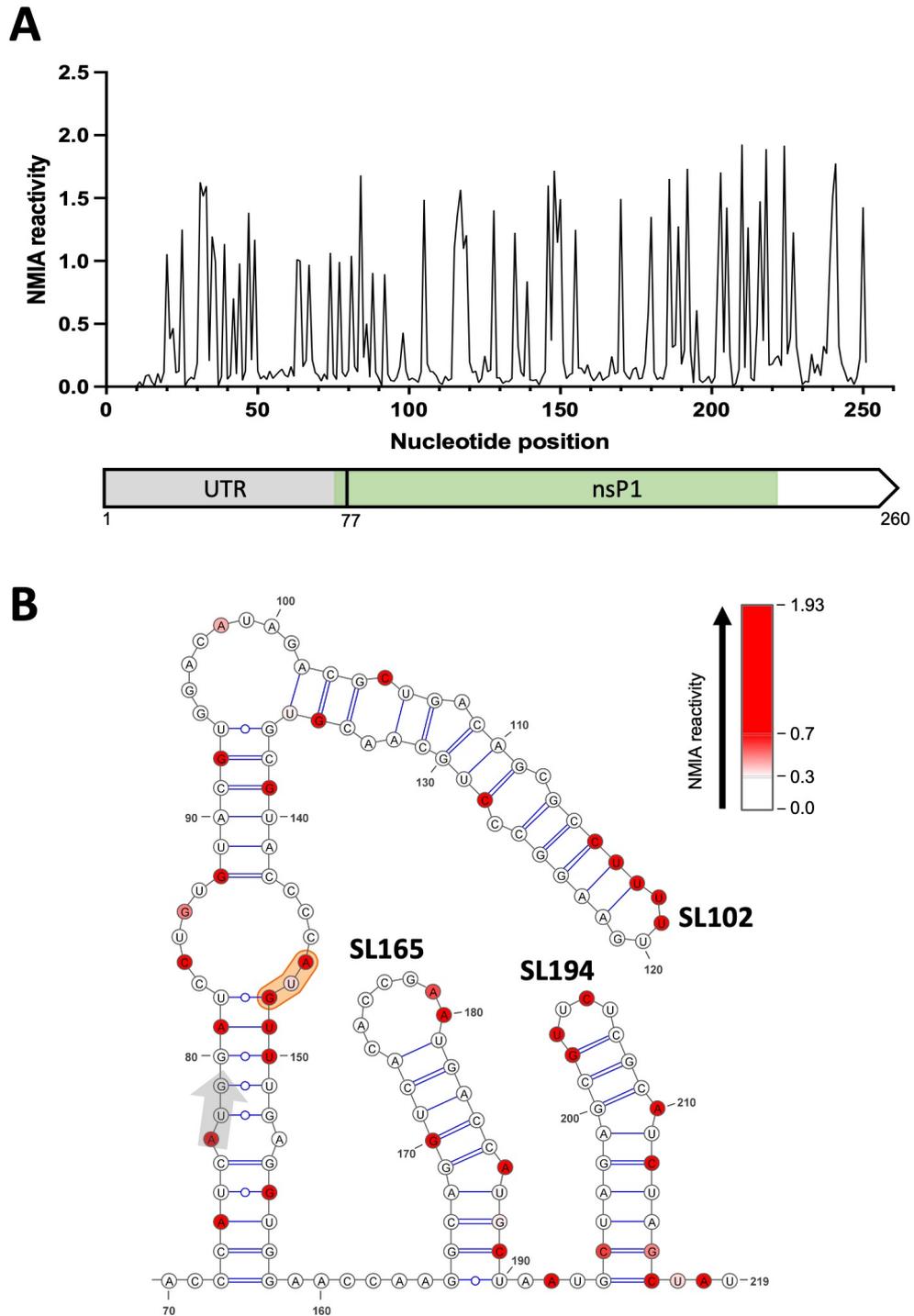


Figure 4.5: SHAPE constrained thermodynamic model of wild-type CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

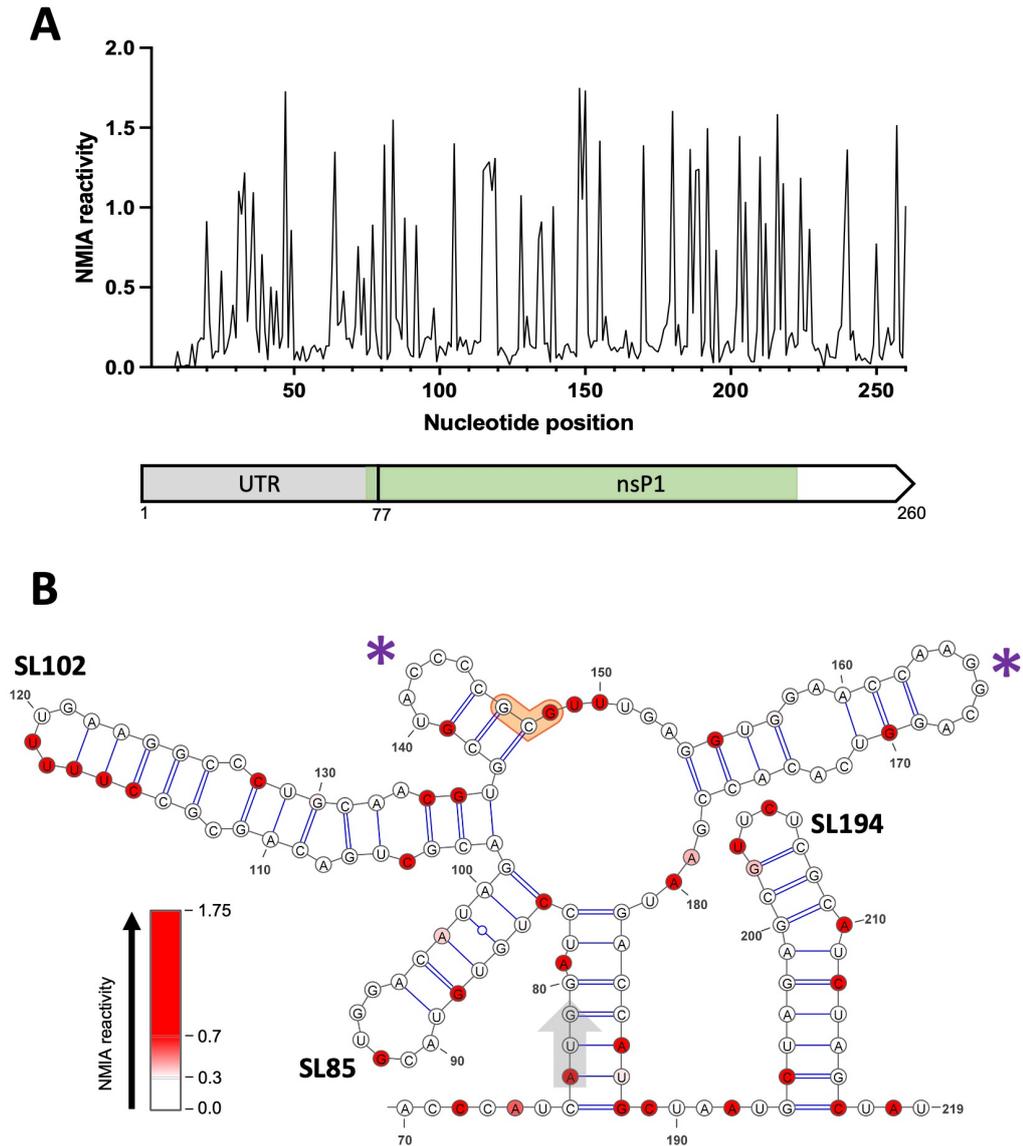


Figure 4.6: SHAPE constrained thermodynamic model of $M^{24}A$ CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

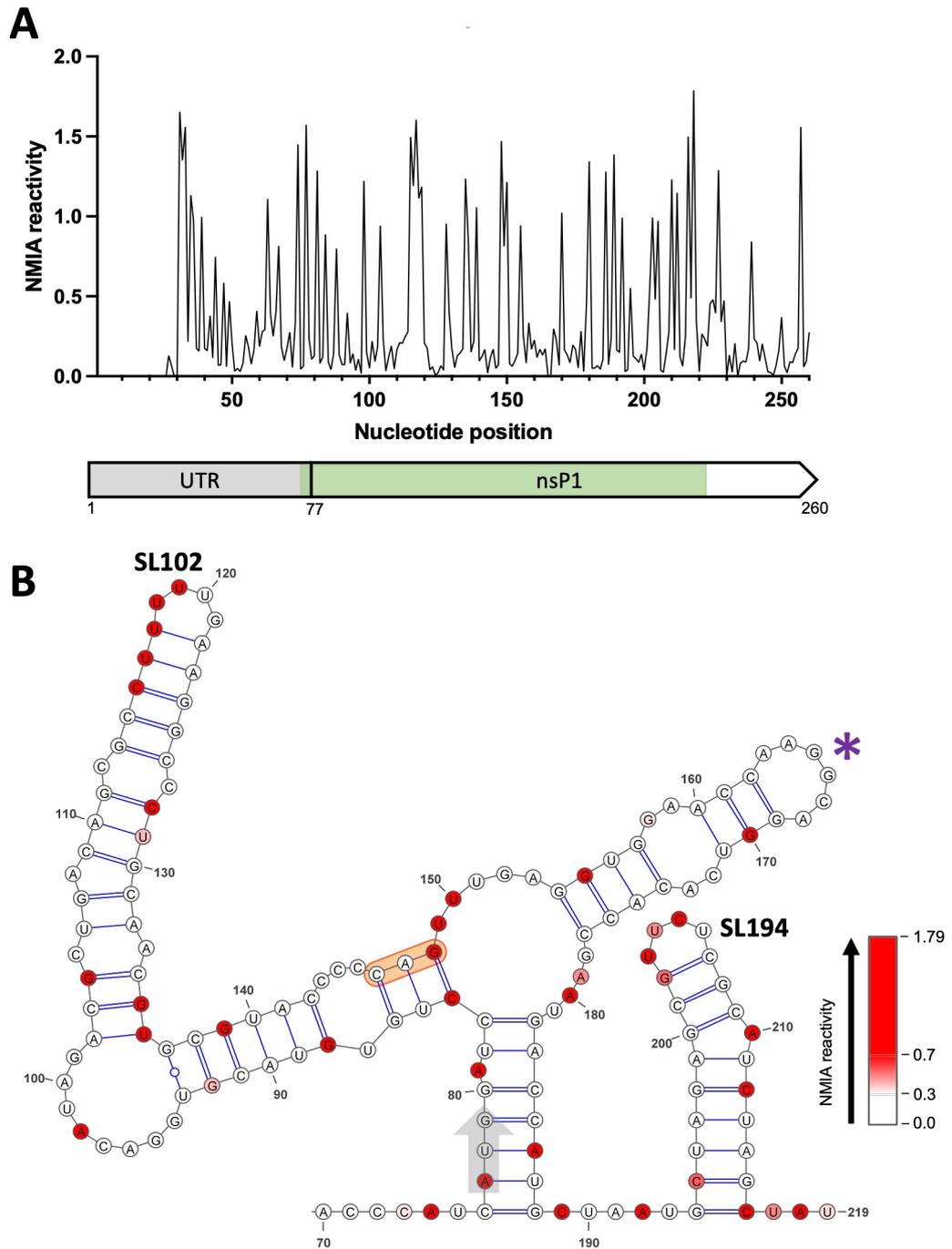


Figure 4.7: SHAPE constrained thermodynamic model of $M^{24}Q$ CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

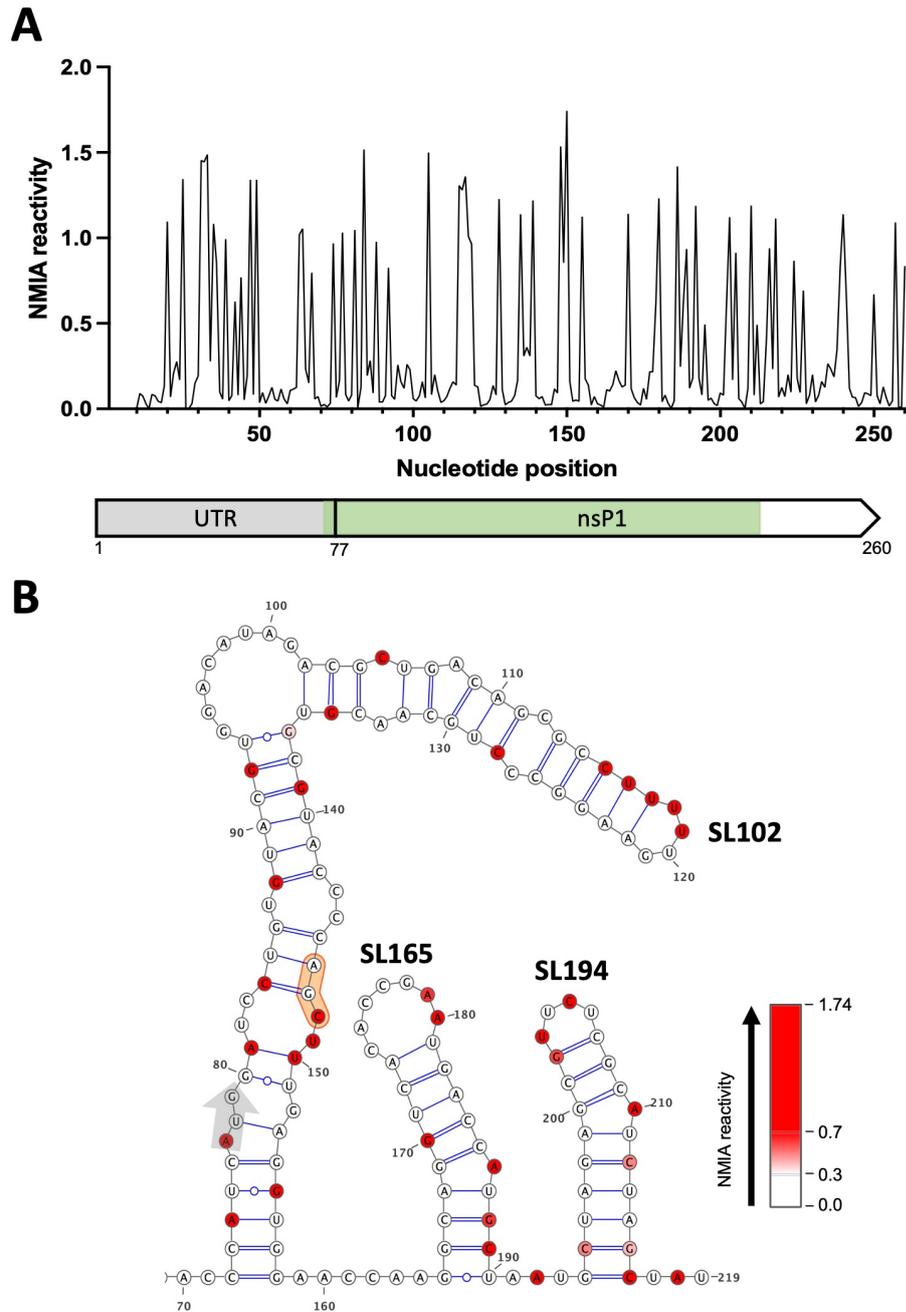


Figure 4.8: SHAPE constrained thermodynamic model of $M^{24}S$ CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

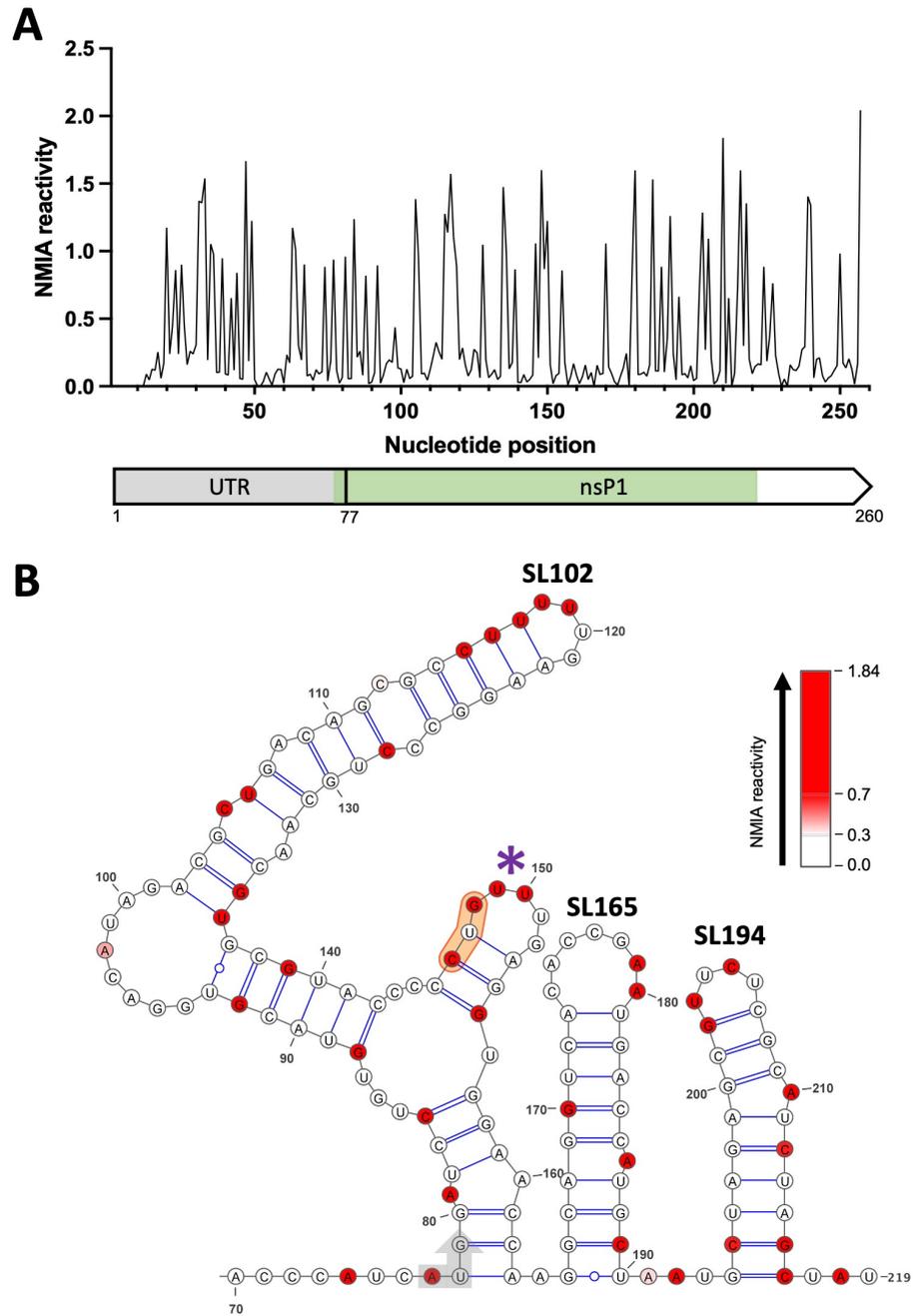


Figure 4.9: SHAPE constrained thermodynamic model of M^{24} LI CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

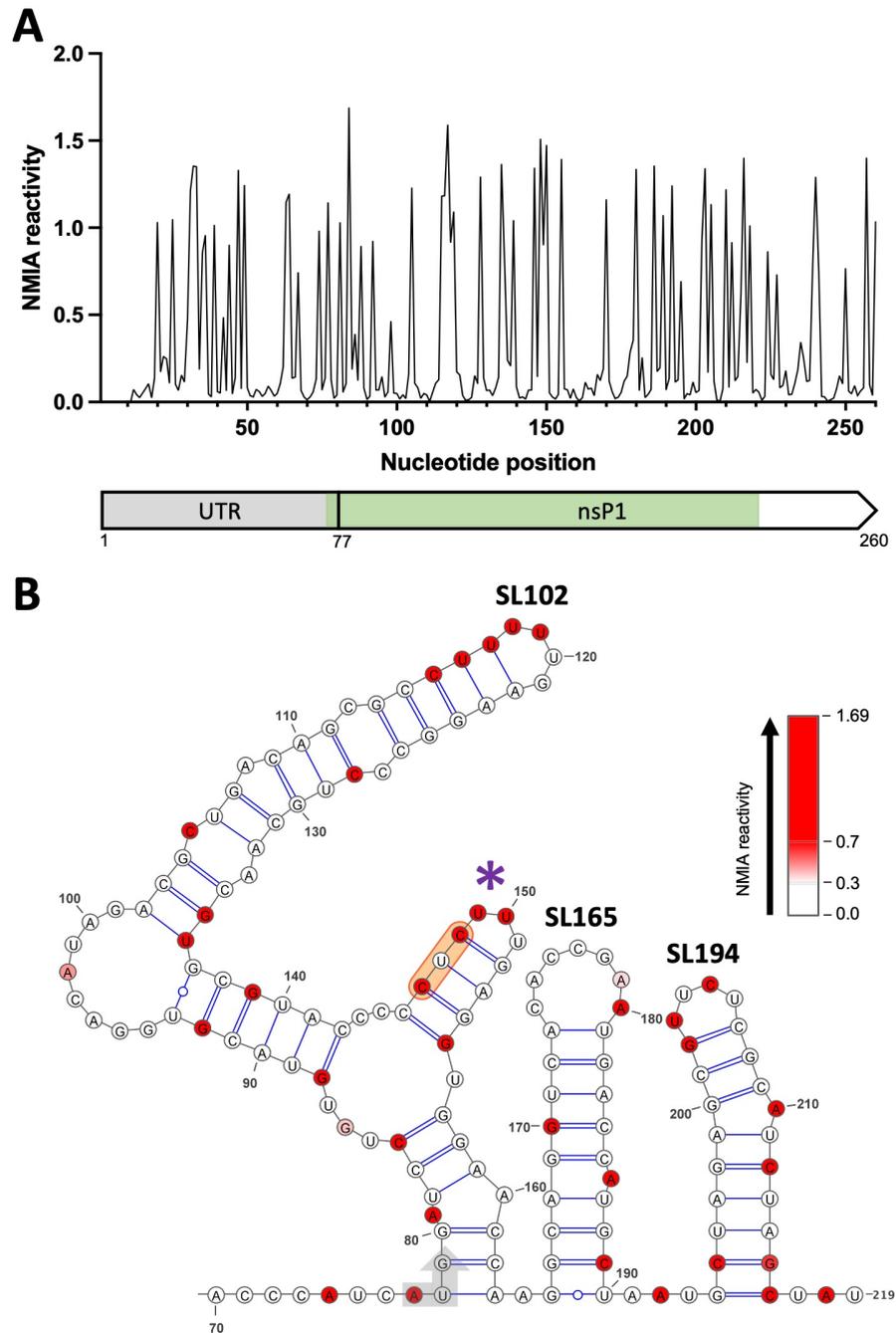


Figure 4.10: SHAPE constrained thermodynamic model of M^{24} LII CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=>3$.

4.2.2 Monitoring of pseudo-escape mutations through serial passaging of M²⁴A CHIKV

The previous section demonstrated that the RNA secondary structures in the 5' coding region of nsP1 were disrupted following introduction of mutants at the M²⁴ codon. In order to further investigate the importance of the change in RNA secondary structures a series of blind passaging experiments were performed. This approach allows for CHIKV to naturally select mutations that restore or rescue virus replication/infectivity and may provide insights into potential function/interactions of this region and/or the M²⁴ codon across the CHIKV lifecycle. The CHIKV genome sequence and viral titre were analysed after each passage to assay whether mutant CHIKV was able to rescue to wild-type titre levels of replication, and if this was associated with pseudo-escape/reversion mutations.

Throughout this chapter pseudo-escape mutants are defined as a base change associated with rescue of mutant CHIKV replication to that of wild-type during passaging. Reverse genetics experiments whereby detected pseudo-escape mutations are re-introduced into the full-length CHIKV infectious clone must be performed to confirm they are true escape mutants. Reversion mutations are defined as a base change from the original mutant sequence to wild-type sequence. Mutations detected in non-coding regions that alter the primary RNA sequence would allude to long-range RNA:RNA interactions between the M²⁴ codon and the mutated region. Mutations that alter the amino acid coding may suggest that the mutated protein interacts with the M²⁴ codon. It is also of note that pseudo-escape mutants may alter the RNA secondary structure surrounding the M²⁴ codon to rescue the RNA structure/interactions observed in wild-type CHIKV.

The M²⁴A mutant was used for this assay as it was previously shown to reduce infectious virus production, decrease CHIKV genome replication and disrupt the RNA secondary structure. Wild-type CHIKV was also subjected to parallel passaging alongside M²⁴A in order to control for cell culture adaptations throughout

the experiment.

Three individual passaging experiments were performed and are described in more detail in sections 4.2.2.2 - 4.2.2.4. Briefly, the appropriate cell line was infected with wild-type and M^{24} A mutant CHIKV at $MOI = 1$. Following a 48 hour incubation, the supernatant containing released CHIKV was harvested; 1 mL of supernatant was used to infect naïve cells and the remainder was retained in order to determine viral titre by plaque assay on BHK-21 cells. Furthermore, the monolayers were harvested with TRI-reagent in order to extract total RNA for downstream NGS analysis. This was repeated every 48 hours until passage 10 (P10) was reached.

4.2.2.1 Development of an NGS pipeline for detecting mutations in CHIKV

Prior to performing the passaging experiments, an NGS pipeline was designed and developed to allow for sequencing of viral RNA extracted from infected cells. Whilst Sanger sequencing continues to be used for validation in a clinical diagnostic setting, the emergence of NGS has allowed for a high-throughput approach and the analysis of a larger sample set (219). NGS allows for deeper analysis and higher mutation resolution. An NGS approach was utilised for this project due to the large sample size and the desire to detect mutations across the whole CHIKV genome, rather than across a short region the M^{24} codon. Furthermore, this approach would enable variant analysis in the population to monitor whether mutations detected increase within the population between passages.

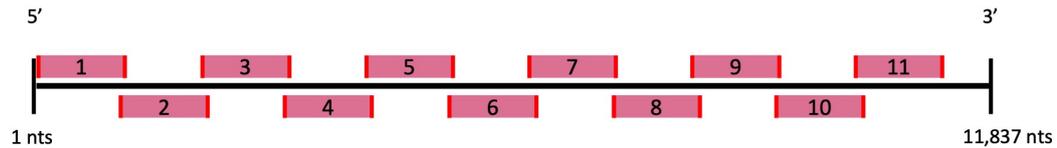


Figure 4.11: Schematic representation of the amplicons generated during the library preparation of CHIKV for NGS analysis

A schematic representation of the amplicons generated during library preparation for PCR-based target enrichment prior to whole genome sequencing of CHIKV. There are 11 amplicons generated that cover the majority of the CHIKV genome nucleotide 1 to 11,604. The red depicts primer binding sites. Amplicons 1, 3, 5, 7, 9 and 11 are generated from primer pool 1 primers. Amplicons 2, 4, 6, 8 and 10 are generated from primer pool 2 primers.

A multiplexed, amplicon based, whole genome sequencing approach was chosen to monitor for escape mutations within CHIKV, using the Oxford Nanopore Technology platform (figure 4.13). Modifications to the broadly used Midnight protocol for analysis of SARS-CoV-2 were made in order to cover the CHIKV genome. This protocol is highly adaptable as it uses a PCR-based target enrichment approach which results in the generation of amplicons that span the target viral genome. Primers were designed that are specific to the CHIKV genome and allow for coverage across 97.75% of the genome (nucleotides 1 - 11,604), generating 11 amplicons each with a length of 1,200 bp (figure 4.11). PrimalScheme was utilised to design the primers; this software optimises primer design for efficient multiplex PCR amplification for the monitoring of short term evolution of different virus lineages short (201). Due to using this algorithm the last 233 nts in the CHIKV 3' UTR are not covered in this sequencing analysis. The amplicons overlap with one another to ensure full genome coverage surrounding primer binding sites. Due to this slight overlap, the primer pairs are separated into two pools to avoid primer dimer formation between complementary nucleotides in different primers; primer pairs 1, 3, 5, 7, 9 and 11 are in primer pool 1 and primer pairs 2, 4, 6, 8 and 10 are in primer pool 2 and their numbers correspond to the amplicon produced (figure 4.11).

In order to confirm that the primers were suitable and that all 11 amplicons were successfully produced, a PCR reaction was performed using the CHIKV ICRES infectious clone DNA as a template, and individual PCR reactions for

each primer pair. The PCR products were analysed by agarose gel electrophoresis to confirm that the amplicon products were 1,200 bp in length. All 11 amplicons were successfully amplified and were all $\approx 1,200$ bp in length, with no other size fragments generated during PCR amplification, allowing us to confirm the primer pairs for each amplicon are specific (figure 4.12). Furthermore, all 11 amplicons were amplified to similar levels under the same PCR conditions, which would allow for equal coverage across the genome during NGS analysis.

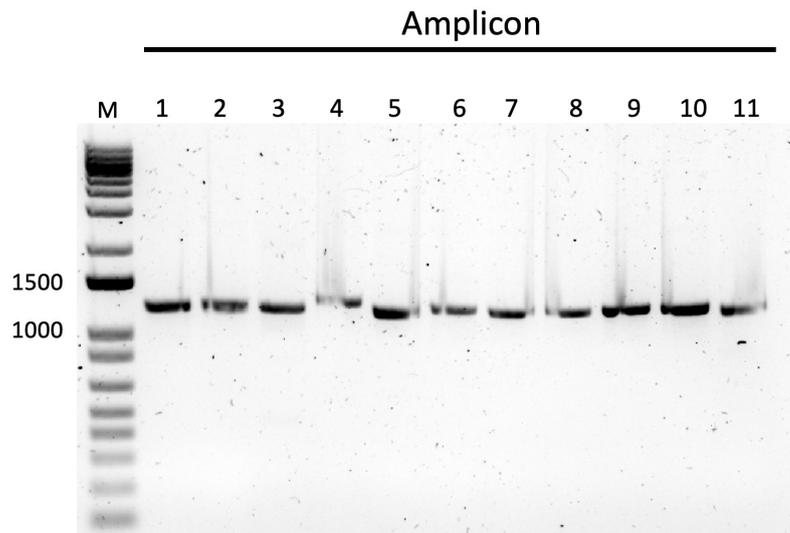


Figure 4.12: Confirmation that the primer pairs designed generate amplicon products of the correct length

Full length CHIKV ICRES cDNA was PCR amplified with primer pairs corresponding to the 11 amplicons that span the CHIKV genome. PCR products were analysed by agarose gel electrophoresis to confirm their size of 1,200 bp. M denotes the size marker.

Following confirmation that all 11 regions were successfully amplified with all primer pairs, an initial, proof of principle, library preparation was performed to confirm that full coverage of the CHIKV genome could be achieved using the Oxford Nanopore Technology platform and adapted Midnight sequencing protocol. BHK-21 cells were infected with M^{24} LII CHIKV at MOI 0.1 and RNA was extracted from the monolayer 24 hpi for NGS analysis.

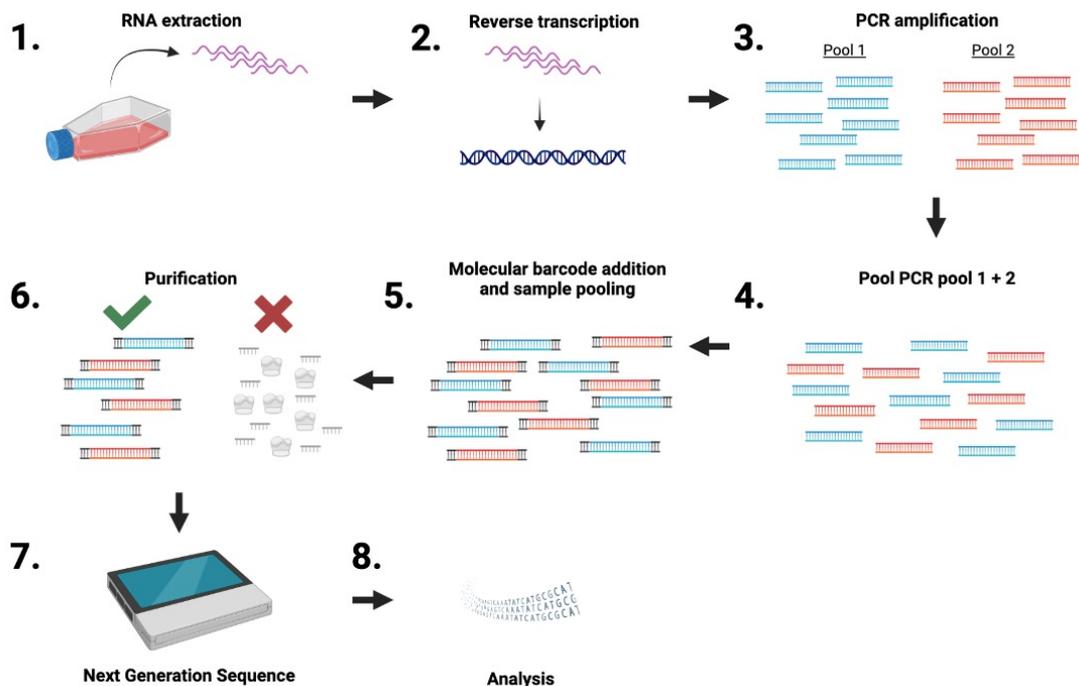


Figure 4.13: Schematic of the library preparation workflow for NGS
1 RNA was extracted from CHIKV infected cells with TRI-reagent and purified. **2** Total RNA was reverse transcribed with random hexamers. **3** PCR amplification of 1,200 bp amplicon fragments with primers designed to generate amplicons that covered the entire genome. Two separate PCRs were performed to prevent primer dimer formation. **4** Both of the primer pools were combined together for individual samples. **5** A molecular barcode was attached to each the amplicons generated for each sample to allow for multiplexing of samples. All samples were combined. **6** DNA was purified using magnetic SIPR beads that selectively bind to DNA. **7** A final ligation step was performed for the addition of the motor protein required for translocation of DNA through the Nanopore. The prepared library was analysed on an Oxford Nanopore Technology Flowcell. **8** Analysis was performed on the InterArtic web application to generate a consensus file.

The first stage in the library preparation workflow was random primer reverse transcription of extracted total RNA from CHIKV infected cells (figure 4.13). The cDNA pool was used as a template for PCR amplification of the amplicons discussed above; two multiplex PCR reactions were performed using primer pool 1 and primer pool 2. The products from the two sets of PCR reactions were combined and a sequencing “barcode” was ligated to each end of the cDNA fragment; this contained a unique sequence that was used to de-multiplex each sample from one another during analysis (up to 96 samples could be multiplexed and analysed on a single sequencing reaction). Following barcode-labelling, all

samples were pooled together and the cDNA was purified with SPRI beads, which are magnetic beads that selectively bind to nucleic acids. There was a final ligation step for the addition of the sequencing adapter. This adapter contains the motor protein that is required for sequencing; this motor protein is the proprietary factor of Oxford Nanopore Technology’s sequencing platform and controls the translocation of DNA through the pore during sequencing. The library was analysed on a flow cell and sequencing was performed. An in-depth library preparation workflow is described in Materials and Methods 2.4.7. Finally, the InterArtic web application was utilised to analyse the raw sequencing files, in order to ultimately produce a consensus file that could be aligned with the CHIKV reference genome (202). The InterArtic web application for analysis also outputs a schematic of the average read depth across the genome and highlights any nucleotides where a SNP was detected in >50% of the population (figure 4.14).

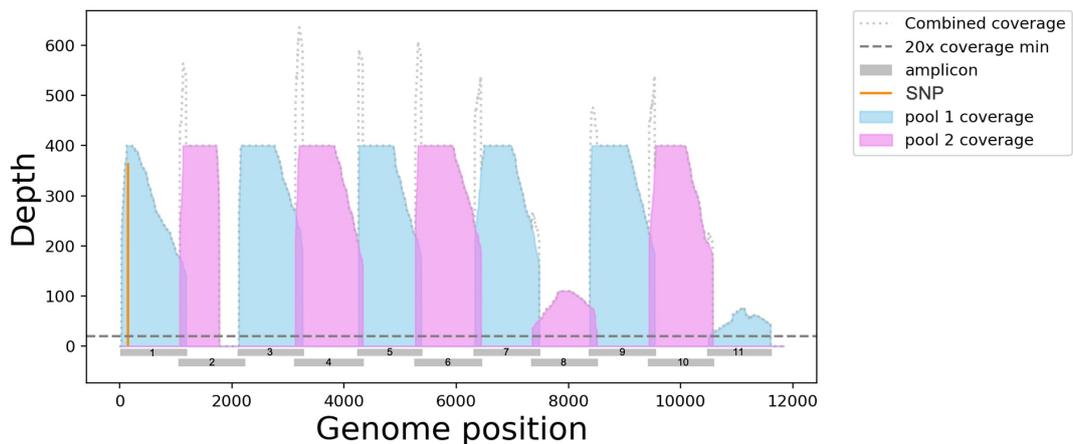


Figure 4.14: Genome coverage plot of CHIKV RNA

Total RNA was extracted from BHK-21 cells infected with M^{24} LII CHIKV. The RNA was analysed by NGS using the adapted Midnight protocol and Oxford Nanopore Technology platform. Results were analysed using InterArtic and a genome coverage plot for CHIKV was generated. Single nucleotide polymorphisms are denoted by an orange line.

A proof of principle sequencing experiment was performed on M^{24} LII RNA extracted from infected BHK-21 cells 24 hpi (infected at MOI 0.1). The majority of the M^{24} LII genome achieved > 200x read depth (figure 4.14). Read depth describes the number of reads that align to a known reference at a particular

location within the genome. Due to error rates of polymerases during PCR and sequencing reactions a higher read depth allows for increased confidence of the base called. The desired read depth varies depending on samples analysed but an average read depth of 20x is adequate for human whole genome sequencing. The regions of the M²⁴LII genome that were amplified by primer pairs 8 and 11 achieved a low sequencing read depth. There was also a lack of coverage achieved in the 3' region of primer pair 2, suggesting that the reverse primer did not bind and amplify properly. The concentration and sample quality of purified total RNA was measured on a NanoDrop Spectrophotometer, however, there was no correlation observed between low concentration/quality of RNA and a high sequencing read depth (data not shown). The *in-vitro* transcribed CHIKV RNA control achieved a high read depth across the whole genome, suggesting that the low coverage was not due to issues with library preparation (data not shown). It remained unclear as to why there was little to no amplification of amplicons 8 and 11, and a sudden drop off in coverage in primer pair 2. Figure 4.14 also successfully highlights the presence of a SNP at nucleotide 146, which results in an amino acid change from methionine to leucine in the M²⁴LII mutant. Given these results, we were confident that the chosen approach was suitable and would reliably detect any mutations, whether they were pseudo-escape or reversion mutants, throughout the following passaging experiments.

4.2.2.2 Pseudo-escape mutants observed at M²⁴ during passaging of M²⁴A in Huh7 cells

The first passaging experiment of wild-type and mutant M²⁴A CHIKV was performed in Huh7 cells, as this cell line was more physiologically relevant than BHK-21 cells and a reduced phenotype in productive virus replication and genome replication of M²⁴A was observed in Huh7 cells. Further, the observation of pseudo-escape or reversion mutants in this passaging experiment would allude to potential mechanistic interactions of the M²⁴ codon (nucleotides 146- 148) during infection of humans.

The viral titre of the supernatant from each passage was calculated. The viral

titre of M²⁴A increased gradually across the passaging experiment, from 4.7×10^4 to 5.6×10^6 pfu/mL (figure 4.1A). The wild-type viral titre increased slightly from P1 to P5, but then decreased by ≈ 2 logs at P6, and decreased further by P10. It was not possible to draw a conclusion whether the M²⁴A mutant virus has gained fitness and recovered to similar titre levels as wild-type from this data alone because the amount of virus transferred to naïve cells differed between wild-type and mutant at each stage due to a fixed volume of unknown titre being transferred at each passage.

Consequently, one-step growth curves were performed in Huh7 cells with a known MOI, in order to determine whether M²⁴A CHIKV had recovered to wild-type CHIKV viral titres. An MOI of 1 was used for virus harvested from P7 to P10; due to low titres of mutant virus at P5 and P6, an MOI = 0.1 was used. One-step growth curves could not be performed on for virus released from P1 - P4 as the M²⁴A titre was too low for MOI = 0.1 infections. Figure 4.15B shows that by P10, M²⁴A had a similar viral titre to that of wild-type CHIKV, suggesting that the virus has gained mutations that either recovered or compensated for the original 146^{AUG} > 148^{GCG} mutation.

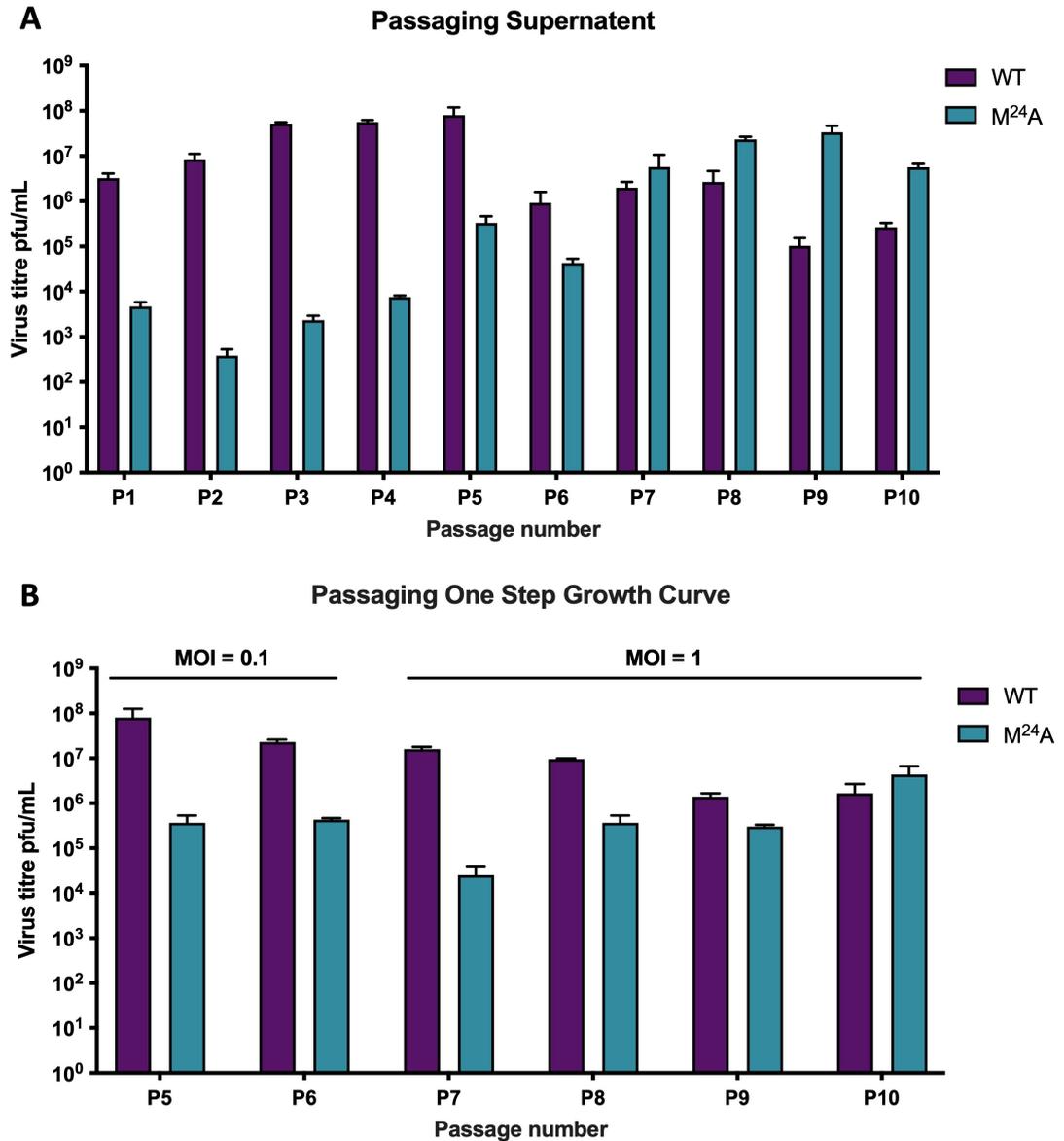


Figure 4.15: Viral titres of wild-type and M^{24} A CHIKV throughout passaging in Huh7 cells

Wild-type and M^{24} A mutant were passaged sequentially through Huh7 cells, up to passage 10. **A** The viral titre of the supernatant from each passaging step determined by plaque assay in BHK-21 cells. **B** A one-step growth curve was performed with virus released at each passage step (P5 & P6 at MOI 0.1, and P7 - P10 at MOI 1) and the viral titre calculated by plaque assay in BHK-21 cells. Error bars represent SEM between two technical repeats. $n=1$.

To investigate any mutations across the M^{24} A viral genome associated with rescue of wild-type phenotype, the RNA from the monolayer of cells at each passage step was extracted and analysed by NGS (as described in 4.2.2.1). There were no reads observed from P1 - P7 and as discussed in section 4.2.2.1 it was unclear

why some samples were unsuccessfully sequenced, therefore only sequencing data for P8 - P10 is presented in table 4.1. The results for nucleotide 147 are discussed in this section, a full list of mutations observed across the genome is presented in appendix table 7.1. An average read depth of 309x was achieved at nucleotide 147. As seen in table 4.1 there was a C₁₄₇^U SNP observed. Whilst this was a reversion to the wild-type sequence and nucleotide 147 it caused an amino acid change from alanine to valine at the M²⁴ codon. The C₁₄₇^U SNP was first detected in viral RNA harvested from P8; 13% of the population were observed to have C₁₄₇^U at this nucleotide. The presence of C₁₄₇^U increased at P9 to 42% and by P10 the majority of the population (71%) was observed to have C₁₄₇^U. The increase of C₁₄₇^U coincides with the recovery of M²⁴A viral titre to wild-type levels (figure 4.15B).

Table 4.1: Sequence observed at nucleotide 147 during passaging of M²⁴A CHIKV in Huh7 cells

The Huh7 monolayer infected with M²⁴A during passaging was harvested with TRI-reagent and total RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P8 - P10 is shown. The percentage of the population containing a cytosine or uracil at nucleotide 147 is presented in the table for each passage, alongside the dominant residue encoded at amino acid residue 24, and read depth for nucleotide 147. The highlighted row depicts where a change in residue was first observed amongst the majority population.

| | Population % C | Population % U | Dominant Residue | Nucleotide Read depth |
|-------------------|-------------------|-------------------|---------------------|--------------------------|
| Passage 8 | 87% | 13% | Alanine | 223x |
| Passage 9 | 54% | 43% | Alanine | 329x |
| Passage 10 | 26% | 71% | Valine | 374x |

During the serial passaging of M²⁴A through Huh7 cells the level of productive virus replication increased by P10 to a similar level to wild-type CHIKV. A reversion mutation was observed at C₁₄₇^U and it was initially shown that this mutation was sufficient to rescue M²⁴A replication to that of wild-type in mammalian cells, suggesting that this mutant is a pseudo-escape mutant and that this region is important for efficient replication. Further experiments were required to confirm that this mutant was a true escape mutant and could recover

replication to wild-type and to determine the potential function of this region.

4.2.2.3 Pseudo-escape mutants observed at M^{24} during alternate passaging of $M^{24}A$ in Huh7 and C6/36 cells

Following observations of the C_{147}^U pseudo-escape mutation in mammalian cells, a second passaging experiment was performed. This passaging experiment was undertaken with the infection of Huh7 cells with wild-type and $M^{24}A$ CHIKV (at MOI 1), and then alternated between C6/36 and Huh7 cells at every passage step. The C_{147}^U mutation observed earlier may not have been sufficient to rescue replication of $M^{24}A$ in mosquito cells and therefore this approach was taken to reflect that CHIKV is an arbovirus and cycles between infection in humans and mosquitoes hosts. If the same pseudo-escape mutations were observed during this passaging experiment it would allude to a function that is required in both mammalian and mosquito hosts.

As with the previous passaging experiment, the titre of virus released into the supernatant of each passage step was calculated in order to perform one-step growth curves at a known MOI. The titre of $M^{24}A$ CHIKV increased across the passaging experiment again, and by P5 the mutant virus had a similar titre to that of wild-type (figure 4.16A). It is of note that it was not possible to fully conclude from this data alone that $M^{24}A$ had recovered to wild-type levels due to a fixed volume of supernatant with unknown titre transferred at each passage.

Due to low titres of mutant CHIKV in P2 and P3, the one-step growth curves were performed at MOI 0.1, whilst the others (P3 - P10) were performed at MOI 1. Following the one-step growth curve analysis, $M^{24}A$ productive virus replication levels were observed to rescue to wild-type equivalent levels by P6, however, the titre of $M^{24}A$ was reduced again in comparison to wild-type by P7 - P8 (figure 4.16B). By P9 and P10, the replication of $M^{24}A$ had once again recovered similar to wild-type levels. The viral titre of $M^{24}A$ CHIKV was rescued to wild-type levels quicker than in passaging in Huh7 cells alone, suggesting that pseudo-escape mutations were gained at a faster rate when M^{24} was passaged alternately through C6/36 and Huh7 cells.

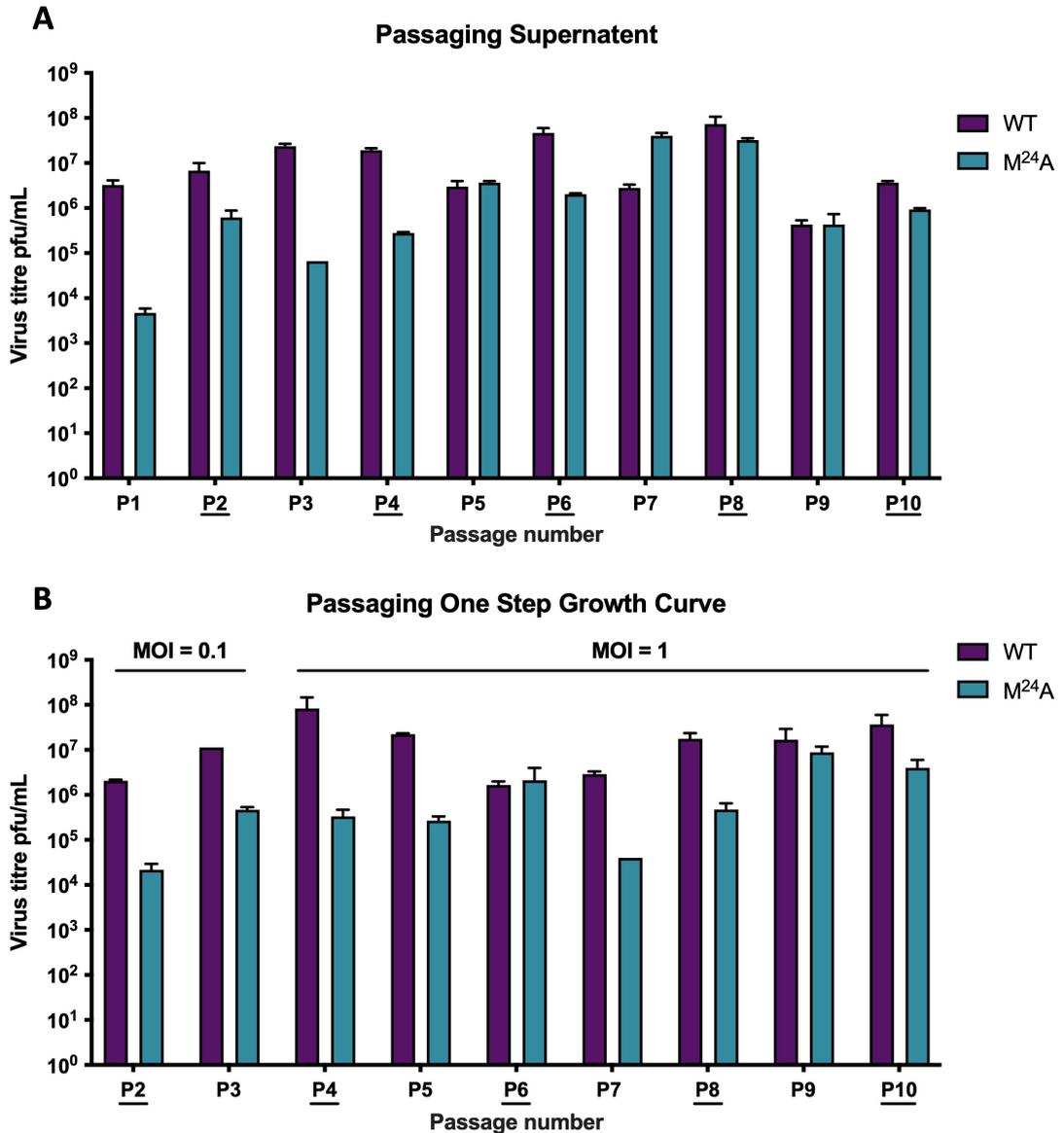


Figure 4.16: Viral titres of wild-type and $M^{24}A$ CHIKV throughout passaging in Huh7 and C6/36 cells alternately

Wild-type and $M^{24}A$ mutant were passaged sequentially through Huh7 cells and C6/36 cells, up to passage 10, alternating between cell line at each passage step. Underlined PX represents supernatant harvested from C6/36 cells. **A** The viral titre of the supernatant from each passaging step determined by plaque assay in BHK-21 cells. **B** A one-step growth curve was performed with virus released at each passage step (P2 & P3 at MOI 0.1, and P4 - P10 at MOI 1) and the viral titre calculated by plaque assay in BHK-21 cells. Error bars represent SEM between two technical repeats. $n=1$.

Viral RNA extracted from the monolayer of each passage step (P2 - P10) was analysed by NGS to monitor for any mutations across the CHIKV genome. This was compared to results obtained from the passaging of wild-type CHIKV to control for cell culture adaptations. The results for nucleotide 147 are discussed

in this section, a full list of mutations observed across the genome is presented in appendix table 7.2. The majority of samples had > 300x read depth at nucleotide 147. As with passaging of M²⁴A in Huh7 cells alone, a C₁₄₇^U SNP was observed (table 4.2). This mutation was observed in 49% of the population at P2, which was following one round of infection in Huh7 and then C6/36 cells. The majority of the population had C₁₄₇^U by P3 (95%) through to P10 (90%). Again, whilst C₁₄₇^U was a reversion to the wild-type base, it resulted in an amino acid change from alanine to valine at the M²⁴ codon. The presence of C₁₄₇^U was observed to be present in the majority of the population far quicker than when the mutant virus was passaged through Huh7 cells. One proposed hypothesis was that the mosquito host was able to act as a reservoir for mutagenesis generating a mixed population of virus, that is then selected for under pressure in humans. Alternatively, the increased rate of replication in mosquito cells could simply allow for more genome copies to be available on which selective pressure can act, or there is increased selective pressure in mosquitoes so the C₁₄₇^U mutant is gained quicker.

Table 4.2: Sequence observed at nucleotide 147 during passaging of M²⁴A CHIKV alternately between Huh7 and C6/36 cells

The Huh7 or C6/36 monolayer infected with M²⁴A during passaging was harvested with TRI-reagent and total RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P1 - P7 is shown. The percentage of the population containing a cytosine or uracil at nucleotide 147 is presented in the table for each passage, alongside the dominant residue encoded at amino acid residue 24, and read depth for nucleotide 147. The highlighted row depicts where a change in residue was first observed amongst the majority population.

| | Population % C | Population % U | Dominant Residue | Nucleotide Read depth |
|-------------------|-------------------|-------------------|---------------------|--------------------------|
| Passage 2 | 50% | 49% | Alanine | 304x |
| Passage 3 | 5% | 95% | Valine | 361x |
| Passage 4 | 5% | 92% | Valine | 372x |
| Passage 5 | 9% | 87% | Valine | 337x |
| Passage 6 | 2% | 95% | Valine | 317x |
| Passage 7 | 17% | 77% | Valine | 146x |
| Passage 8 | 9% | 82% | Valine | 57x |
| Passage 9 | - | 90% | Valine | 125x |
| Passage 10 | - | 90% | Valine | 342x |

Serial passaging of $M^{24}A$ alternately between Huh7 and C6/36 cells resulted in the selection of a C_{147}^U reversion mutant, which caused an amino acid change from alanine to valine at $M^{24}A$. This mutation was observed to be present in the majority of the population by P2, suggesting that there is greater selection pressure when $M^{24}A$ is passaged through Huh7 and C6/36 cells sequentially.

4.2.2.4 Escape mutants observed at M^1 during passaging of $M^{24}A$ in C6/36 cells

As discussed above, 49% of the population was observed to have C_{147}^U following one passage between Huh7 and C6/36 cells and we proposed that C6/36 cells allow for a higher incorporation rate of mutations within the CHIKV genome. A final passaging experiment was performed, this time in C6/36 cells alone, to observe whether mutations would be gained at a similar rate to the those observed across the alternate passaging experiment. The passaging of $M^{24}A$ in mosquito cells alone could have elucidated to functions of the M^{24} codon that are specific to replication within the mosquito vector.

The titre of wild-type CHIKV in the harvested supernatant decreased across the passaging experiment, and at P10 the titre was reduced by ≈ 1 log compared to P1 (figure 4.17A). The supernatant titre of $M^{24}A$ followed a similar pattern to that of wild-type, and gradually reduced from P1 through to P9, and then increased slightly again at P10. CHIKV does not induce CPE or cell death in C6/36 cells so it was surprising to see that the wild-type CHIKV titres decreased during passaging in this cell line, rather than continuing to gradually increase.

As with the previous two passaging experiments, one-step growth curves were performed in order to determine whether $M^{24}A$ titres were rescued to wild-type levels. C6/36 cells were infected at MOI 0.1 with the virus harvested from P1 - P10; P9 was excluded as the $M^{24}A$ titre was too low to perform an MOI 0.1 infection. $M^{24}A$ titres were not rescued to wild-type levels by P10 (figure 4.17B). At P3, there was only a 0.5 log reduction in $M^{24}A$ titre compared to wild-type, but the difference in titres increased again at P4 with a > 2 log reduction of mutant titre observed in relation to wild-type CHIKV. The sustained

reduced productive virus replication phenotype observed with the $M^{24}A$ virus after 10 rounds of passaging in C6/36 cells suggested that there were no pseudo-escape/compensatory mutations gained.

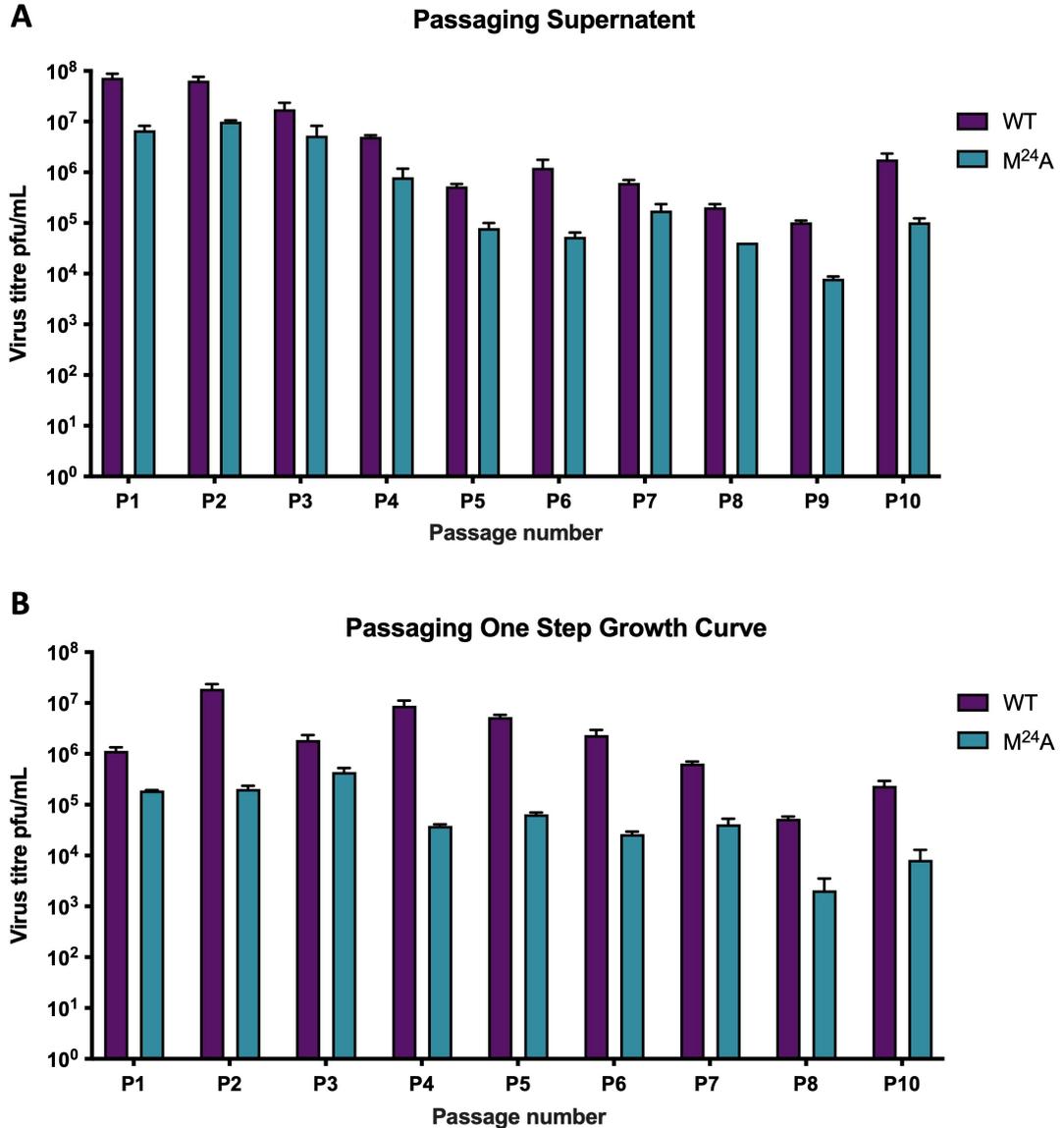


Figure 4.17: Viral titres of wild-type and M^{24} CHIKV throughout passaging in C6/36 cells

Wild-type and $M^{24}A$ mutant were passaged sequentially through C6/36 cells, up to passage 10. **A** The viral titre of the supernatant from each passaging step determined by plaque assay in BHK-21 cells. **B** A one-step growth curve was performed at MOI 0.1 with virus released at each passage step (apart from P9) and the viral titre calculated by plaque assay in BHK-21 cells. Error bars represent SEM between two technical repeats. $n=1$.

There were no reads observed from the RNA extracted from the P8 - P10

monolayers during passaging of M^{24} in C6/36 cells. The quality and concentration of purified extracted total RNA was measured with a spectrophotometer but there was no correlation observed between poor quality/concentration and low read depth. The results for nucleotide 77 are discussed in this section, a full list of mutations observed across the genome is presented in appendix table 7.3. NGS analysis showed that there was no observed change in sequence at nucleotides 146 - 148 (corresponding to $M^{24}A$ codon) from P1 - P7, with $M^{24}A$ still dominant across the population. However, there was a A^{77G} SNP observed (table 4.3) by P2 and maintained in the population through to P7. The SNP was first observed at P2, where 13% of the population encoded A^{77G} . By P3 the majority of the population (60%) had A^{77G} . Approximately 75% of the population encoded A^{77G} from P4 to P7. This nucleotide is part of the start codon for ORF-1 and therefore it was surprising that CHIKV would select for a mutation that would cause an amino acid change from methionine to valine. A lack of methionine at this site would inhibit translation of the non-structural proteins; however the percentage of population containing this SNP increased from P1 - P4 and remained present in the majority of the population across 5 rounds of passaging.

During the passaging of virus at high titres it is possible that defective interfering virions/genomes are produced. These defective interfering genomes can be truncated genomes that are recognised and replicated by the wild-type replicase complex in *trans* but do not generate infectious virions. It should be noted that the A^{77G} mutant was only present in a defective interfering genome. This hypothesis would support the reduction in viral titre during passaging and explain why the appearance of A^{77G} never reached the high percentages observed with $C_{147}U$.

Table 4.3: Sequence observed at nucleotide 77 during passaging of M²⁴A CHIKV in C6/36 cells

The C6/36 monolayer infected with M²⁴A during passaging was harvested with TRI-reagent and total RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P8 - P10 is shown. The percentage of the population containing an adenine or guanine at nucleotide 77 is presented in the table for each passage, alongside the dominant residue encoded at amino acid residue 1, and read depth for nucleotide 77. Highlighted row depicts where a change in residue was first observed amongst the majority population.

| | Population % A | Population % G | Dominant Residue | Genome Read depth |
|-----------|-------------------|-------------------|---------------------|----------------------|
| Passage 1 | 99% | 1% | Methionine | 396x |
| Passage 2 | 86% | 13% | Methionine | 386x |
| Passage 3 | 39% | 60% | Valine | 387x |
| Passage 4 | 19% | 81% | Valine | 376x |
| Passage 5 | 21% | 79% | Valine | 381x |
| Passage 6 | 23% | 77% | Valine | 383x |
| Passage 7 | 27% | 73% | Valine | 377x |

4.2.3 Reverse genetic confirmation of CHIKV M²⁴A pseudo-escape mutants

To further understand the impact of the potential escape mutants observed throughout the passaging experiments, they were introduced into the full-length CHIKV ICRES DNA infectious clone. A summary of the mutations introduced can be seen in table 4.4. The aim was to generate mutant virus in order to compare the replication and infectivity of mutant virus with wild-type CHIKV, through one-step growth curves. If the replication of these pseudo-escape mutants was similar to that of wild-type it would confirm that mutations observed during passaging of M²⁴A were sufficient to rescue replication to that of wild-type and elucidate the function of the M²⁴ codon (either alteration of RNA secondary structures or RNA:protein interactions with nsP1 and nsP3).

Table 4.4: Summary of pseudo-escape mutants introduced into the CHIKV ICRES infectious clone

Mutations that were detected by NGS in CHIKV during passaging experiments were cloned into the CHIKV infectious clone. A summary of which nucleotide regions were mutated within individual mutants alongside the mutant names. The base change at the given nucleotide is underlined.

| Mutant name | Nucleotide position | Codon sequence | Amino acid residue |
|--------------------------------------|---------------------|----------------|---------------------------|
| Wild-type | 77 - 79 | AUG | M ¹ |
| | 146 - 148 | AUG | M ²⁴ |
| | 425 - 427 | G <u>G</u> G | G ¹¹⁷ (nsP1) |
| | 4424 - 4426 | G <u>G</u> G | G ¹¹⁷ (nsP3) |
| M ²⁴ V | 146 - 148 | <u>G</u> UG | M ²⁴ V |
| M ²⁴ V combination | 146 - 148 | <u>G</u> UG | M ²⁴ V |
| | 425 - 427 | <u>A</u> GG | G ¹¹⁷ R (nsP1) |
| | 4424 - 4426 | <u>A</u> GG | G ¹¹⁷ R (nsP3) |
| M ¹ V | 77 - 79 | <u>G</u> UG | M ¹ V |
| M ¹ V & M ²⁴ A | 77 - 79 | <u>G</u> UG | M ¹ V |
| | 146 - 148 | <u>G</u> CG | M ²⁴ A |

Virus containing the mutants observed during passaging (table 4.4) was generated in BHK-21 cells as described in Materials and Methods 2.4.1.1. The viral titres of the sequencing mutant panel were reduced in comparison to wild-type CHIKV (figure 4.18). Whilst there was no virus recovered from the M¹V & M²⁴A mutant, there was 2.6 x10³ pfu/mL of M¹V virus generated. There was a higher titre of M²⁴V combination mutant generated in comparison to the M²⁴V standalone mutant, suggesting that the two additional mutants in nsP1 and nsP3 within the combination mutant were required for more efficient CHIKV replication.

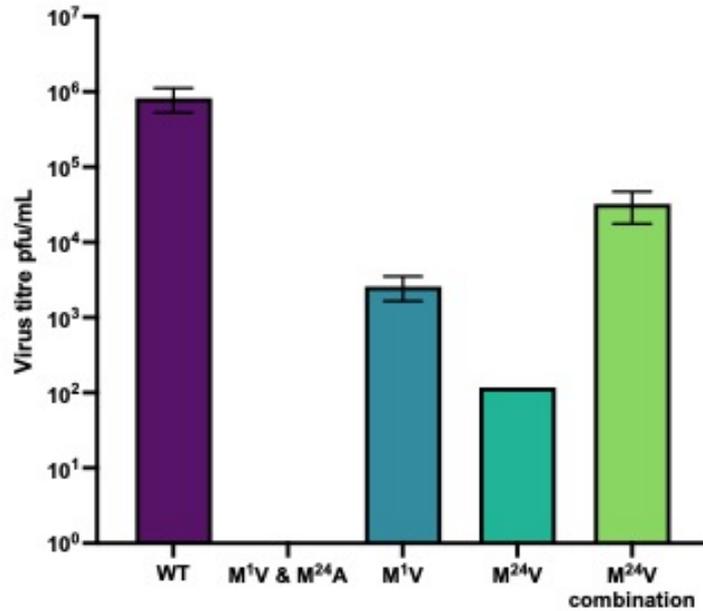


Figure 4.18: Titre of escape mutations following transfection of BHK-21 cells

BHK-21 cells were transfected with wild-type and mutant CHIKV RNA. Supernatant containing released virus was harvested 24 hpt and productive CHIKV replication measured by plaque assay in BHK-21 cells. Error bars represent SEM between two technical repeats. $n=1$

4.2.4 M^{24V} CHIKV replication competed with wild-type CHIKV replication

A competition assay was performed to compare the fitness pseudo-escape mutants (table 4.4) with that of wild-type CHIKV, and whether the wild-type CHIKV genome would be favoured over mutant genomes. Cells were transfected with equal amounts of both wild-type ICRES RNA and mutant CHIKV ICRES RNA (M^{24V} , M^{1V} & M^{24A} and M^{24A}). Control transfections were also performed, with yeast tRNA and wild-type or mutant CHIKV ICRES RNA, to monitor the replication of the mutants in the absence of wild-type CHIKV RNA and to normalise the nucleic acid levels during transfection. The inclusion of yeast tRNA alongside CHIKV RNA allows for the regulation of transfection efficiency with a normalised amount of RNA. Virus that was released into the supernatant following transfection (P0) was subjected to two rounds of passaging in C6/36 cells, as described in 4.2.2. Productive CHIKV replication was calculated by

plaque assay for supernatant harvested from all three passaging steps and total RNA was harvested from the monolayers for NGS analysis to determine whether the wild-type CHIKV genome was able to out compete the mutant genomes.

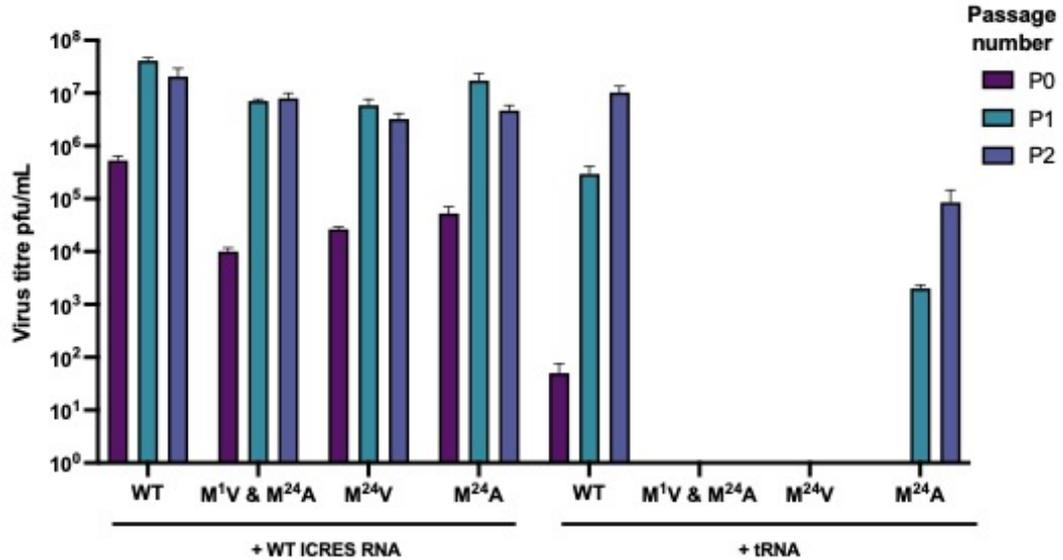


Figure 4.19: Titre of virus generated across passaging the competition assay of mutant CHIKV against wild-type CHIKV

Cells were transfected with 1 μ g wild-type ICRES RNA and the relevant mutant CHIKV RNA. Control transfections using yeast tRNA were performed. Supernatant was harvested 48 hpt and transferred to naïve cells for 2 rounds of passaging. Productive CHIKV infection was measured by plaque assay for supernatants harvested at each passaging step. Error bars represent SEM between two technical repeats. $n=1$.

As shown in figure 4.19 there was no virus generated for the mutants M¹V & M²⁴A and M²⁴V when the mutant ICRES RNA was co-transfected with yeast tRNA. However, when the mutant ICRES RNA was co-transfected with wild-type CHIKV ICRES RNA productive mutant virus infection was observed. Both of these mutants had a ≥ 1 log reduction in viral titre compared with wild-type CHIKV at P0, but the titres had increased by P2 and were only ≈ 0.5 log lower than the wild-type titre. These results suggested that either the mutants required the wild-type CHIKV genome in *trans* to efficiently replicate, or that it was only the wild-type ICRES RNA input that produced infectious CHIKV.

Similarly, when M²⁴A ICRES RNA was co-transfected with wild-type ICRES RNA there was a 1 log reduction in M²⁴A titre observed compared to wild-type only at P0, that recovered to a 0.5 log reduction vs wild-type by P2. There was

no virus detected at P0 when M²⁴A ICRES RNA was co-transfected with yeast tRNA. However, by P2 there was 8.5 x 10⁴ pfu/mL virus detected. This was a 2 log reduction compared to virus generated when the M²⁴A CHIKV ICRES RNA was co-transfected with wild-type CHIKV ICRES RNA. This data suggested that the wild-type genome is able to out compete the M²⁴A mutant genome or that the wild-type is acting as a helper virus. In order to confirm whether wild-type functioned as a helper virus NGS analysis of cellular viral RNA was required.

When wild-type RNA was co-transfected with yeast tRNA there was a 4 log reduction in virus titre at P0 compared to virus generated when double the amount of wild-type RNA was transfected into C6/36 cells, suggesting that the co-transfection of CHIKV RNA with yeast tRNA had reduced efficiency and that the yeast tRNA competed against CHIKV ICRES RNA during transfection.

To understand what was happening at the genome level during the passaging of mutant ICRES RNA co-transfected with wild-type CHIKV ICRES RNA, NGS analysis was performed on extracted RNA from the monolayers of each passage step. An NGS approach was chosen over a Sanger sequencing approach so that we could determine what percentage of the population contained wild-type or mutant genomes.

There was very low read depth achieved for all the samples across P0 and P1, which meant no conclusion could be drawn about the relative replication efficiencies of the wild-type and mutant viruses. A read depth of > 500x was achieved for the cDNA sequenced from the P2 samples. Only sequencing data is shown in figure 4.20 from RNA extracted from the P2 monolayers across the competition assay. Following transfection and passaging of wild-type + M¹V & M²⁴A ICRES RNA, 96% of the population was observed to encode M²⁴ (figure 4.20). This demonstrated that the wild-type CHIKV genome was favoured for efficient replication of CHIKV over the M¹V & M²⁴A mutant genome. Alternatively it suggested that the M¹V & M²⁴A mutant reverted to wild-type sequence at both mutated residues. The same result was observed for the M²⁴A mutant when it was co-transfected with wild-type RNA. Interestingly, when the M²⁴V mutant CHIKV ICRES RNA was co-transfected alongside wild-type

CHIKV ICRES RNA the 50:50 ratio of M^{24} V virus:wild-type virus was maintained over 2 rounds of passaging. This data suggested that the M^{24} V genome was either as efficient for CHIKV genome replication as wild-type CHIKV or it may suggest that the M^{24} V mutant utilised the wild-type genome/machinery in *trans* to allow for replication of the mutant genome. Due to previous results demonstrating that M^{24} V had reduced replication compared to wild-type CHIKV (figure 4.18) the latter is more likely.

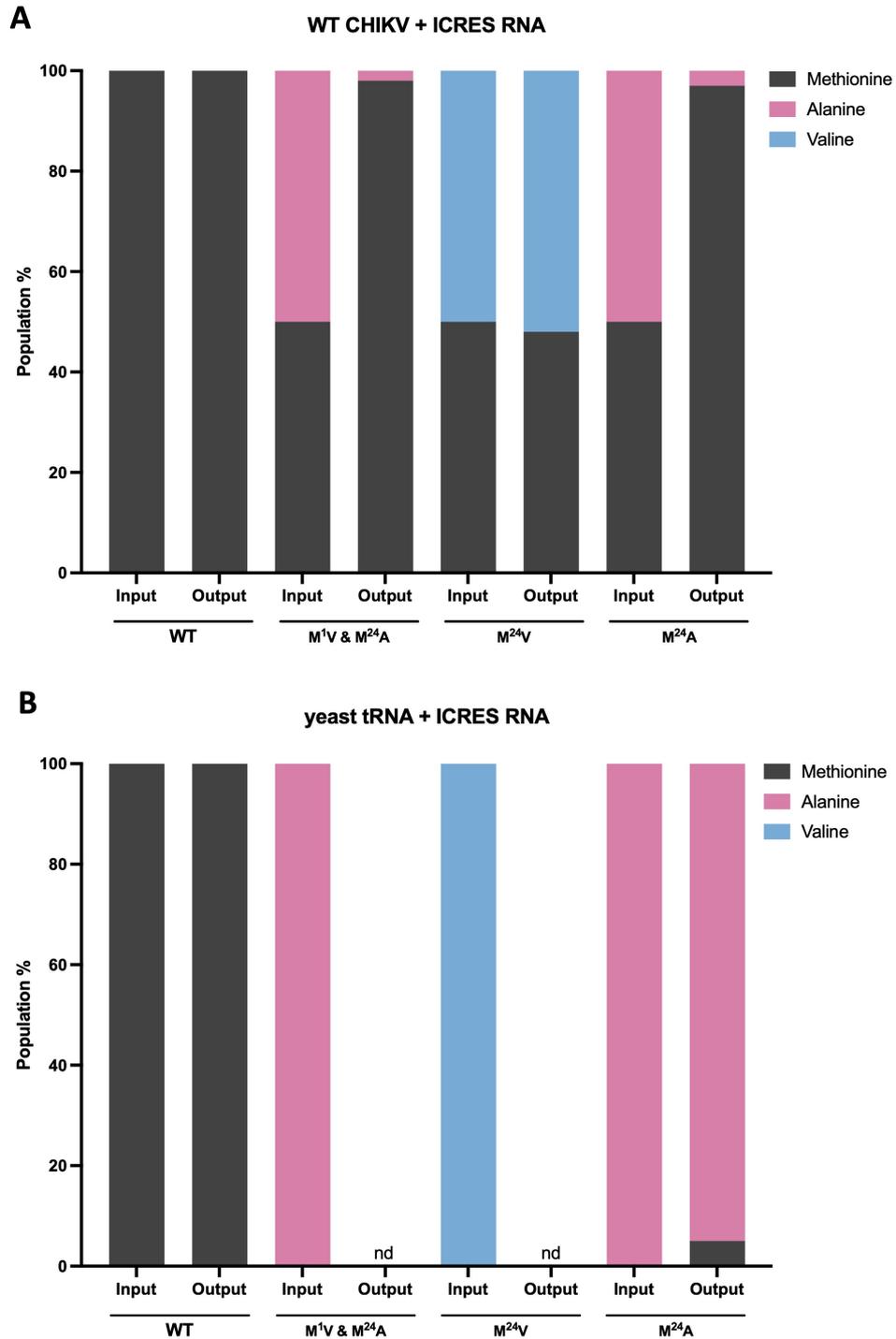


Figure 4.20: NGS analysis of RNA extracted from cells co-transfected with wild-type and mutant CHIKV RNA

RNA from the monolayers of infected cells from the competition assay was harvested at P3 and analysed by NGS to determine whether the wild-type sequence was able to out compete the mutant sequence. The input ratio of wild-type to the mutant amino acid 24 (in nsP1) is shown alongside the percentage of population with either a methionine, alanine or valine following three rounds of passaging. **A** The population percentage of three different amino acids observed at residue 24 (in nsP1) when mutant RNA was co-transfected with wild-type CHIKV RNA. **B** The population percentage of three different amino acids observed at residue 24 (in nsP1) when mutant RNA was co-transfected with yeast tRNA as a transfection control.

4.2.5 SNPs detected during passaging of mutant CHIKV result in RNA secondary structure changes

In order to understand the mechanism behind the development of pseudo-escape mutants, we investigated how the SNPs observed following passaging of M²⁴A affected the RNA secondary structure of the virus genome (M²⁴V and M¹V & M²⁴A mutants). Initial RNA structures were predicted *in-silico* using the Vienna RNAfold server, which predicts the minimum free energy secondary structures from the single stranded RNA sequence. The ^{A77G} mutation, alongside the original M²⁴A mutation (deemed M¹V & M²⁴A), rescued the RNA secondary structure to that of the structure published by Kendall et al (173) (figure 4.21A). This rescue of wild-type secondary structure was also observed with the ^{G147U} mutation (deemed M²⁴V) (figure 4.21B).

To further confirm that the SNPs observed during sequencing were altering the RNA secondary structure, SHAPE analysis was performed on M²⁴V and M¹V & M²⁴A from *in-vitro* transcribed RNA, as described in Materials and Methods 2.2. Secondary structures were predicted using the Vienna RNAprobing server, which predicts the RNA structure using SHAPE reactivity constrained minimum free energy.

The *in-silico* predicted structure for M¹V & M²⁴A was also observed when using the SHAPE constrained approach, and aligns with the published structure from Kendall et al with SL85 present (figure 4.22) (173). The model predicted using SHAPE constraints for the M²⁴V mutant differed to that predicted from minimum free energy. There was a longer arm structure predicted with SL102 at the apex and SL85 was not present; this secondary structure is similar to that published by Madden et al (figure 4.23B) (174). In both mutant structures SL102, SL165 and SL194 were formed. Furthermore, pairing between the first two bases of M¹ and M²⁴ was observed in both structures.

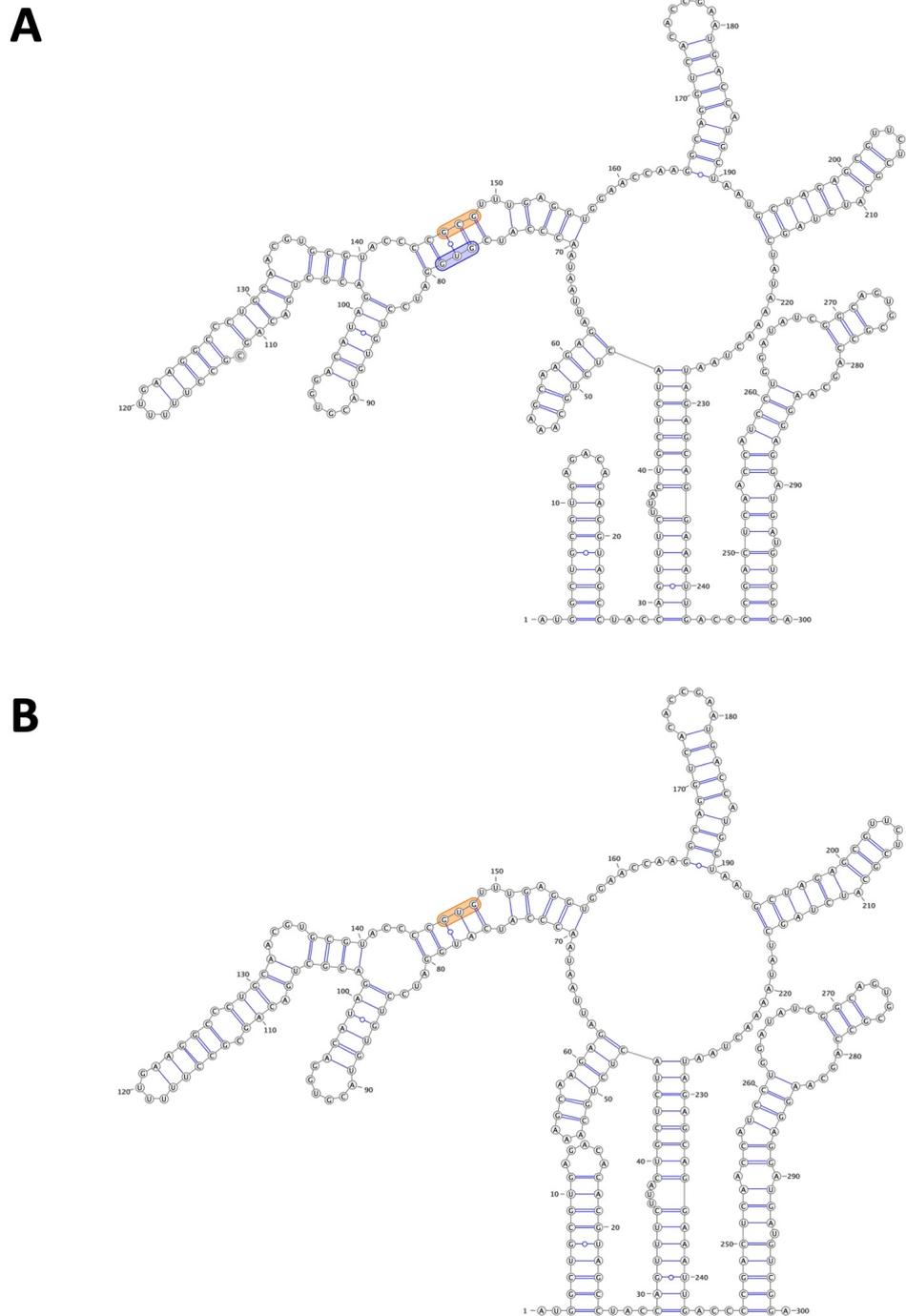


Figure 4.21: *In silico* minimum free energy RNA secondary structure prediction of M^{1V} & M^{24A} and M^{24V} CHIKV

The sequence of CHIKV containing mutants observed through passaging experiments were input into the Vienna RNAfold server to predict the minimum free energy secondary structure. Structures were visualised using VARNA. **A** M^{1V} & M^{24A} mutant predicted structure. The orange highlight denotes amino acid residue 24 in nsP1. The blue highlight denotes amino acid residue 1. **B** M^{24V} mutant predicted structure. The orange highlight denotes amino acid residue 24.

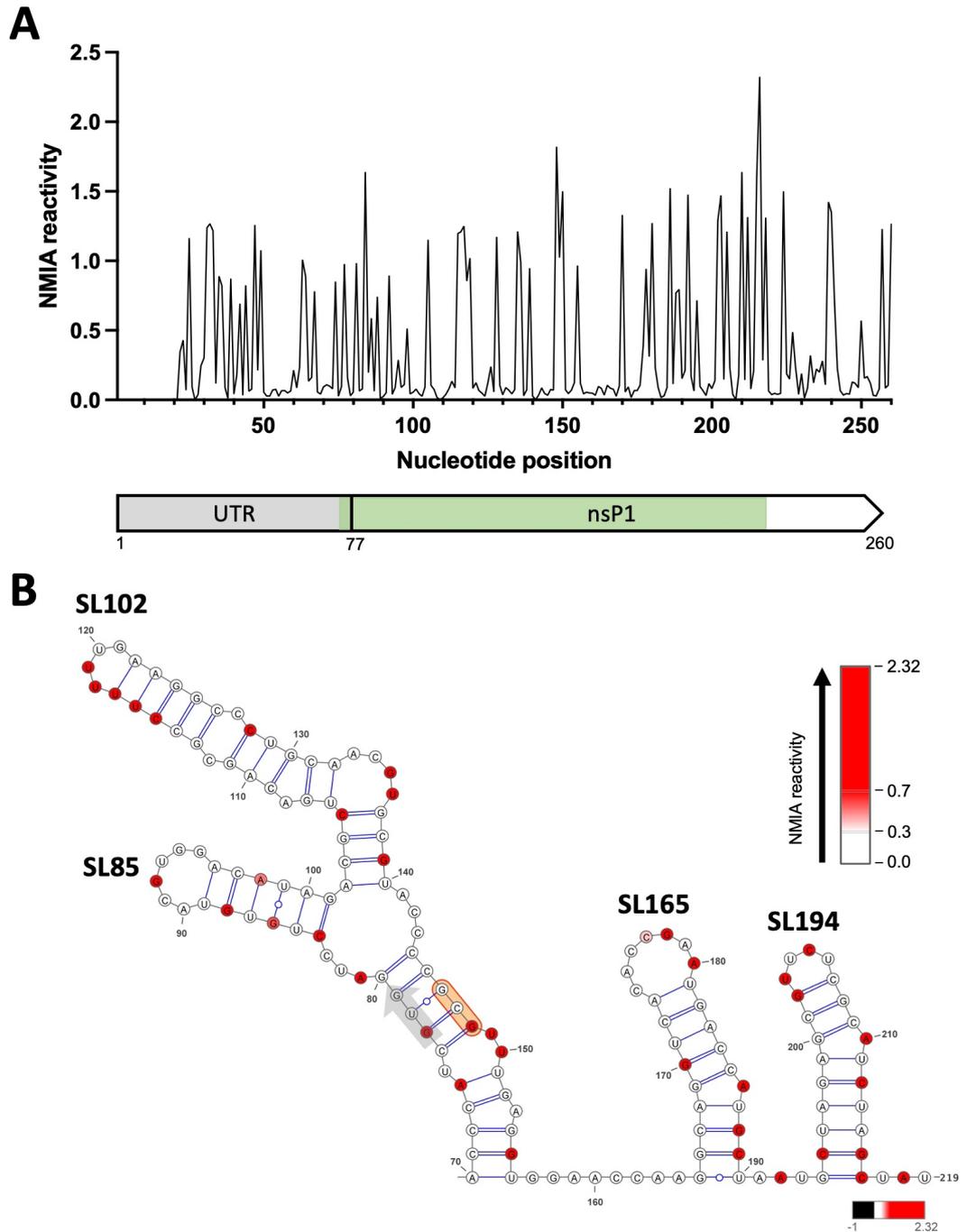


Figure 4.22: SHAPE constrained thermodynamic model of M^{1V} & M^{24A} CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAProbing server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. N=6.

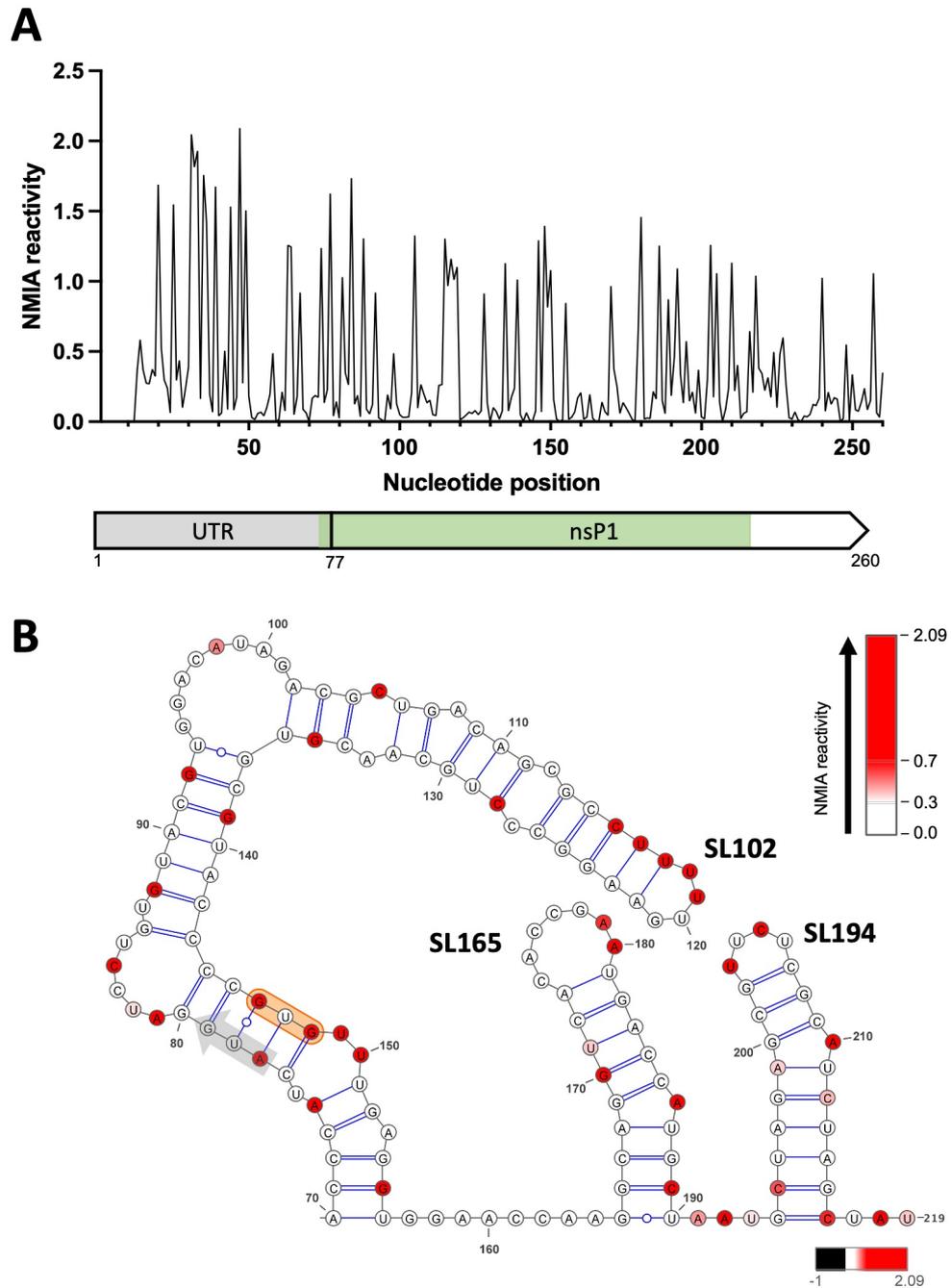


Figure 4.23: SHAPE constrained thermodynamic model of $M^{24}V$ CHIKV

A SHAPE reactivity for individual nucleotides. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3. **B** SHAPE reactivity constrained minimum free energy model. Base-pairing was predicted by the Vienna RNAPROBING server. The structure was visualised using VARNA. NMIA reactivities were overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). The start codon for ORF-1 is depicted by a grey arrow. The orange highlight depicts amino acid 24. $N=6$.

The data presented in this section has demonstrated that the pseudo-mutants

observed during passaging of $M^{24}A$ in mammalian and mosquito cells (deemed M^{1V} & $M^{24}A$ and $M^{24}V$) were sufficient to rescue the RNA secondary structures to those observed in published models of CHIKV (173, 174), confirming that the presence of these replication elements was essential for efficient CHIKV replication.

4.3 Discussion

4.3.1 Mutations in the M^{24} codon disrupted RNA secondary structures

This chapter has shown that introducing mutations (either a single base change or a 2 nucleotide substitution) at the M^{24} codon disrupted the RNA secondary structure of CHIKV and that the alteration in structure was unique to each mutant (table 3.1). Published data has previously shown the importance of stem loops within the 5' region of the CHIKV genome, and that introducing mutations within this region to disturb the structures results in reduced CHIKV replication efficiency (173, 174). Previous work from our group utilised locked nucleic acid oligonucleotides (LNAs) to further demonstrate the importance of SL165 for efficient CHIKV replication, and more specifically that this stem loop is important for viral genome replication opposed to ORF-1 translation (Oliver Prosser, University of Leeds). Therefore the disruption of stem loops (SL85, SL165 and SL194) observed across the M^{24} mutant panel (figures 4.5 - 4.10) was likely to impact replication of mutant CHIVK (figure 3.7). The $M^{24}S$ mutant was the only mutant from the original mutant panel that displayed a similar structure to the wild-type structure (figure 4.8B and 4.5B respectively). This suggests that both the structure and sequence of the 5' coding region of nsP1 were important for efficient CHIKV replication.

The wild-type structure that was predicted following SHAPE analysis in this chapter differed from the structure that was previously published from our group (173). The published Kendall et al structure showed an additional stem loop

(SL85) prior to SL102, whereas the structure predicted here displayed a long arm from nucleotide 72 that extended into SL102. This longer arm structure is seen in the model published by Madden et al (174), however, the structure presented in this thesis differs slightly to theirs and has a larger single-stranded bulge halfway through the arm than the Madden structure.

It is clear that the 5' region of CHIKV is dynamic surrounding nucleotides 80 - 100. The two current published structures differ with the presence or absence of SL85, and from the passaging experiments we detected the M^1V & $M^{24}A$ mutant that reverted to the Kendall structure (173) (SL85 present) and the $M^{24}V$ mutant which reverted to the Madden model (174) (SL85 absent). Furthermore, the *in-silico* predicted structure for CHIKV predicted the presence of SL85, whereas the SHAPE reactivity constrained thermodynamic predicted structure showed a long arm terminating in SL102. Kendall et al discuss that their SHAPE mapping of SL85 was not consistent with their *in-silico* predicted structure but were able to confirm the presence and importance of SL85 for CHIKV genome replication by structure-led reverse genetic analysis (173). They go on to state that this region may be highly dynamic and suggest that the loop region at the end of SL85 is able to potentially form a pseudoknot with the area of the genome upstream of and including the ORF-1 start codon. It is possible that SL85 is at the thermodynamic threshold of the free energy states between structures that have the presence or absence of SL85 and that it only takes a minute change in the thermodynamics for SL85 to be formed. Alternatively, SL85 may be stabilised by intracellular conditions (such as pH or salt levels) or co-transcriptional folding.

It is of note that the reactivity scores for the unpaired nucleotides present in the single-stranded bulges of the wild-type structure presented in figure 4.5B are reduced in comparison to those published by Kendall et al. This could be due to a lower concentration of NMIA used during SHAPE mapping experiments performed throughout this thesis (a final concentration of 5 mM NMIA was used in this study whereas Kendall et al used a final concentration of 10 mM). An increase in NMIA concentration may potentially cause the SHAPE constrained predicted structure to include SL85 and work should be carried out to investigate whether altering the NMIA concentration alters the predicted RNA structure.

A consistent pattern was observed across all eight RNA structures presented in this chapter, whereby residues 115 - 119 (at the apex of SL102) displayed high (> 0.7) reactivity levels, suggesting that these are unpaired bases. However, the structural predictions all displayed residues 115 - 117 as paired and residues 118 and 119 were presented as unpaired.

It would be interesting to perform SHAPE analysis on mammalian and mosquito cell lines infected with wild-type and all 7 mutants discussed throughout this chapter (M²⁴A, M²⁴Q, M²⁴S, M²⁴LI, M²⁴LII, M¹V & M²⁴A and M²⁴V). The work presented in this chapter was performed on *in-vitro* transcribed RNA and it would be important to also investigate how mutations are able to alter the RNA structure within a cellular environment. RNA-RNA interactions, RNA-protein interactions and the cellular conditions (such as pH and ion concentrations) could alter the RNA structure (220). These future experiments would further confirm that mutations in the M²⁴ codon disrupt the RNA secondary structures at the 5' region known to be essential for efficient replication and that passaging of M²⁴A in both mammalian and mosquito cell lines results in selection for pseudo-escape mutants to rescue wild-type RNA structures.

4.3.2 Passaging of M²⁴A CHIKV leads to pseudo-escape mutants to rescue wild-type replication

A novel workflow for NGS analysis of CHIKV has been presented in this chapter. Successful whole genome sequencing was performed using an amplicon-based approach on the Oxford Nanopore Technology platform. This allowed us to be able to detect mutations associated with passaging of M²⁴A CHIKV in different cell lines. As well as detecting mutations at the initial site of interest (residues 146 - 148) we were able to detect SNPs in both ORFs and in the 3' UTR (table 7.1 - 7.3 in appendix). An NGS approach, rather than a Sanger sequencing approach, also allowed us to be able to monitor the population percentage that contained a SNP. Therefore, we were able to observe an increase in the presence of SNPs throughout the passaging experiments at residue 147, from them being in a sub-section of the population to being present in the majority of the population.

It was surprising to discover that during passaging of M²⁴A in C6/36 cells mutations were gained in the start codon for ORF-1 (A⁷⁷G) which resulted in an amino acid change from methionine to valine. This mutation would lead to an inhibition in ORF-1 translation of the non-structural proteins, which in turn would limit genome replication, and ultimately viral replication. The sequencing data also showed that this mutation was present in 70 - 80 % of the population but did not increase beyond this, and instead the percentage of population with A⁷⁷G actually decreased by 8% from P4 to P7. Unfortunately, there were no reads observed following NGS of P8 - P10 and hence work should be repeated to determine whether the presence of the SNP continued to decrease to P10. It would be interesting to continue to take the passaging further to investigate whether the presence of this mutant would disappear from the population over time. Importantly, there was no observed recovery of M²⁴A replication to that of wild-type levels following 10 rounds of passaging in C6/36 cells (figure 4.17B) and there was no virus recovered when the M¹V & M²⁴A mutants were re-introduced into the CHIKV genome.

Previously published studies have demonstrated that CHIKV was capable of genomic RNA replication in *trans* and even demonstrated that the CHIKV replicase complex was able to replicate genomic RNA of viruses within the SFV complex in *trans* (215). It is therefore possible that the M¹V & M²⁴A mutant virus is able to replicate (despite lacking the AUG start codon for ORF-1 translation) through the use of the wild-type genome in *trans*. In this model, presence of wild-type genome sequence would allow for the efficient translation of the CHIKV non-structural replicase proteins (nsP1-4). SHAPE analysis in this chapter showed that A⁷⁷G was sufficient to revert the RNA structure of M²⁴A to the wild-type structure published by Kendall et al (figure 4.22B). Given that the stem loop structures present in the 5' region of CHIKV are known to be essential for virus genome replication, it was likely that the replication complex translated from the wild-type template was able to recognise and replicate the M¹V & M²⁴A mutant genome in *trans*. In order to confirm this hypothesis, a competition assay was performed in which equal amounts of wild-type and M¹V & M²⁴A ICRES RNA were transfected into the same cells. If the mutant

RNA was equally favoured for replication we would have expected to be able to continue to detect the M^1V & $M^{24}A$ mutations by NGS analysis after 2 rounds of passaging. This however was not the case, and only the wild-type sequence was detected in cells after transfection and passaging of M^1V & $M^{24}A$ in C6/36 cells (figure 4.20), suggesting that the wild-type genome sequence is preferentially replicated. Due to time constraints this experiment could only be performed once and therefore it is essential that this work is repeated at least twice more in order to confirm this result.

A second hypothesis was that a valine at residue 1 is still able to function as a start codon. It has been shown that amino acids with the codon XUG (where X can be any nucleotide) are able to initiate translation, although they are far less efficient at initiating translation in comparison to an AUG (216, 221). For example, the eukaryotic initiation factor 4G2 (eIF4G2) is known to initiate solely from a GUG start codon (222). It is therefore possible that CHIKV is still able to translate the non-structural proteins in ORF-1 by initiation through a GUG (valine), albeit at a lower efficiency. This reduced translation efficiency of a GUG start codon would also explain why the titres of the mutant virus did not recover to wild-type levels.

A final hypothesis was that the M^1V mutation was incorporated within defective interfering genomes that are generated as a side effect of passaging virus at a high titre. Whilst the chosen NGS workflow was successful at detecting mutations across the entire genome and gave insight into the increase in population percentage of mutations throughout the passaging experiments, the amplicon approach did not allow us to determine whether genomes were full-length or defective interfering genomes. Furthermore, this approach is unable to determine whether mutants detected were on the same genome strand or as a part of a greater mixed population. Utilising a direct RNA sequence approach may have meant we could confirm whether the M^1V mutant was a defective interfering or full-length genome and whether mutations detected were incorporated within the same genome.

Despite the M^1V mutant persisting in the majority of the population when

passaging M²⁴A in C6/36 cells, attempts to rescue M¹V & M²⁴A were not successful. On the other hand, the introduction of the M¹V mutant without the presence of mutant M²⁴A generated virus in BHK-21 cells. *In-silico* prediction of the M¹V & M²⁴A mutant and M¹V mutant structures both displayed the same RNA secondary structure, so it was surprising that only virus was generated with the M¹V mutant. The M¹V & M²⁴A was observed during passaging of M²⁴A in C6/36 cells and therefore it was likely that the mutations gained were only viable in mosquito cells. Attempts were made to generate virus through transfection of C6/36 cells but transfection efficiency was significantly reduced and low titres of wild-type CHIKV were reported and no recovery of mutant virus (data not shown). This reduced transfection efficiency could have been due to the high passage number of the insect cells and therefore additional work should be carried out to determine whether we are able to generate M¹V & M²⁴A mutant virus in mosquito cells rather than mammalian cells. It would be interesting to investigate whether the mutants gained through passaging of M²⁴A in C6/36 cells are able to replicate as efficiently in the *Ae. aegypti* cell line, Aag2. Furthermore, it is important to include the original M²⁴A mutant as a control in future repeats of this experiment.

Following 10 rounds of passaging in both Huh7 and alternately between Huh7 and C6/36 cells, M²⁴A had recovered to viral titres similar to that of wild-type CHIKV. However, when these mutations (M²⁴V and M²⁴V combination) were re-introduced into the CHIKV genome for reverse genetic analysis there remained a reduced viral titre phenotype in comparison to wild-type CHIKV. It is of note that when the M²⁴V mutant was introduced into the CHIKV genome alone there was a reduced viral titre in comparison to the M²⁴V combination mutant virus (which contained additional single base changes in nsP1 and nsP3). This set of experiments was only performed once and therefore would need to be repeated to at least $n = 3$ for any conclusions to be drawn. Furthermore, there were additional SNPs observed across the CHIKV genome and whilst they were not present in the majority of the population it is possible that they are important in conjunction with the discussed mutants for efficient replication. These mutations were not taken forward for further analysis as only mutations present in >50% of

the population were investigated. A more in-depth analysis could be performed on the passaging NGS data sets to design follow up experiments where a number of mutants are introduced across the genome.

All of mutant viruses detected during passaging (table 4.4) were generated by transfection in BHK-21 cells. The mutants should be interrogated in both mammalian and mosquito cell lines. If the M¹V & M²⁴V mutant virus is attenuated in mammalian cells but can be efficiently generated in C6/36 cells this may provide important information in the development of attenuated vaccines in the future and elucidate to mechanistic interactions of the replication elements in the 5' coding region of nsP1 during replication CHIKV in the mosquito vector. Furthermore, there is a possibility that the mutations observed throughout all three passaging experiments were due to tissue culture adaptations, however this was controlled for through the passaging and analysis of wild-type CHIKV in parallel and the mutations observed and discussed in M²⁴A were not observed in wild-type CHIKV.

Furthermore, it would be interesting to perform another set of passaging experiments using the mutant virus generated in this chapter (table 4.4). This could be performed once again in Huh7 and C6/36 cells, and alternately between the two cell lines. If the M¹V & M²⁴A and M²⁴V mutants were sufficient to rescue replication to that of wild-type levels we may not expect to see any new SNP gained throughout these set of passaging experiments. On the other hand, M²⁴V would only require a single base change to revert to the full wild-type sequence (146^{GUG} > 148^{AUG}) and therefore M²⁴V may select for reversion mutants more readily.

In addition, the *trans*-replicase assay system should be utilised to further characterise all pseudo-escape mutations observed during the passaging experiments. This system would allow for the M¹V mutant to be incorporated into both the replicase and reporter strand. As discussed above, it would be important to perform the *trans*-replicase assay experiments in both mammalian and mosquito cell lines to determine whether any effects observed are host specific.

The competition assay experiments were performed with yeast tRNA as the

transfection control but it would be ideal to repeat the experiments again with the CHIKV infectious clone that has an inactive RdRp (there is a GAA mutation in the active site of nsP4) as the RNA for the transfection controls. It also needs repeating to $n = 3$ to confidently conclude that the wild-type genome was preferentially replicated over the M¹V & M²⁴A and M²⁴A mutant genomes but that the M²⁴V mutant genome was as equally favoured as wild-type/utilised the wild-type genome in *trans*.

An alternative approach to performing NGS on extracted total RNA throughout the competition assay is to utilise CHIKV with fluorescent proteins, such as eGFP and mCHERRY, incorporated within the genome. For example, pseudo-escape mutations would be introduced into the mCHERRY strain and both the wild-type eGFP and mutant mCHERRY RNA would be co-transfected into cells (as described for the competition assay performed in this thesis). Live cell imaging using an IncuCyte would allow you to determine the proportion of wild-type to mutant virus throughout infection cycles and whether mutant virus is able to compete with wild-type virus. If the mutant virus replicated as efficiently as wild-type you would observe a consistent 1:1 ratio of eGFP and mCHERRY signal. If the mutant virus was less efficient at replicating or a defective interfering genome only the eGFP signal would be detected.

In addition it is important to take into account the MOI of wild-type and mutant virus throughout a competition assay. At an increased MOI there is competition within individual cells in addition to the competition for cell entry. At a low MOI this competition does not exist and instead virus that is able to release progeny at early time points has an advantage over a slower replicating virus.

Finally, work was undertaken to introduce the mutants detected across the passaging experiments into the *trans* replicase system utilised in section 3.2.5 in order to investigate the mutants in the context of genome replication to confirm whether they are able to replicate as efficiently as the wild-type genome. Unfortunately, due to time constraints this work could not be completed fully.

Chapter 5

Purification of recombinant CHIKV nsP1

5.1 Introduction

When the project began in 2018 there was no published molecular structure available for nsP1 across the alphaviruses, and a large amount of functional knowledge of CHIKV nsP1 was based on research carried out on SFV and SINV nsP1. Whilst these viruses are within the same genus and are expected to perform the same gross function, there are differences between the sequences, notably the inclusion of a second methionine as described previously. CHIKV is an emerging virus that is a potential threat to global health, therefore it is important to study nsP1 within the specific context of CHIKV. This project was initially interested in the importance of the residue M²⁴, and its role within the CHIKV lifecycle as a potential ORF-1 alternative start codon. A methionine at this position is conserved in CHIKV and ONNV but not seen in any other alphavirus sequence and therefore it is essential to further investigate nsP1 specifically to CHIKV. Preliminary results prior to my project starting indicated towards the expression of an N-terminally truncated version of nsP1 and the initial hypothesis was that nsP1 had previously uncharacterised functions during CHIKV replication. This project aimed to purify recombinant CHIKV nsP1 from *E. coli* in order to determine a molecular structure through X-ray crystallography to further understand the importance of the M²⁴ residue and its function of a potential truncated form of nsP1. A structure for nsP1 would help improve the understanding of previously characterised functions of nsP1 (such as the addition of a 5' cap on nascent viral RNA and anchoring the replicase complex to the host plasma membrane), and

the comparison of full-length and truncated nsP1 structures may allude to novel functions in the truncated version.

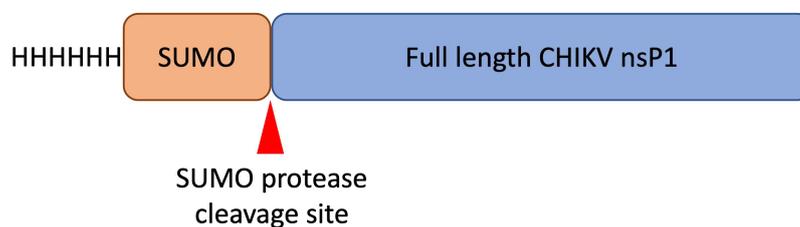


Figure 5.1: Schematic representation of the 6xHis-SUMO nsP1 construct

The *E. coli* expressed 6xHis-SUMO tagged CHIKV nsP1. SUMO protease cleaves off the SUMO tag and poly-histidines for purification of CHIKV nsP1 with no additional residues.

Previous work by our group had initiated efforts in the purification of recombinant CHIKV nsP1, utilising bacterial expression systems. Codon optimised full-length wild-type CHIKV nsP1 was cloned into a pET-28a-6xHis-SUMO expression vector (figure 5.1). This vector allowed for the addition of a N-terminal 6xHis-SUMO tag to the protein of interest and subsequent purification of nsP1 by nickel column chromatography (immobilised metal affinity chromatography (IMAC)). Another advantage of this tag was that SUMO protease could be used to cleave off the tag, leaving no residual amino acids and allowing for the native protein to be purified. Small-scale purifications were successfully performed and they demonstrated that CHIKV nsP1 could be purified through Ni-NTA chromatography. However, the protocol required further optimisation as they were not able to purify a high enough concentration for further X-ray crystallography trials following size exclusion chromatography.

5.1.1 Aims

The aim of this chapter was to purify recombinant CHIKV nsP1 from *E. coli*, in order to determine a molecular structure through X-ray crystallography. It further aimed to utilise a structural approach to gain a greater understanding

into the importance of the M²⁴ residue and function of a potential truncated form of nsP1.

5.2 Results

The nsP1-pET28a-6xHis-SUMO plasmid was transformed into Rosetta 2 cells (*E. coli*) and a glycerol stock of transformed bacterial cells was used to inoculate a starter culture. This culture was then used to inoculate 1 L of LB medium and incubated at 37°C in an orbital incubator. When the OD₆₀₀ reached 0.4-0.6 the culture was induced with IPTG and incubated overnight at 30°C. Bacterial cells were pelleted and resuspended in lysis buffer and either lysed by sonication or by a cell disrupter. The insoluble and soluble fractions were separated and filtered supernatant was applied to a HisTrap column, allowing 6xHis-nsP1 to bind to the column.

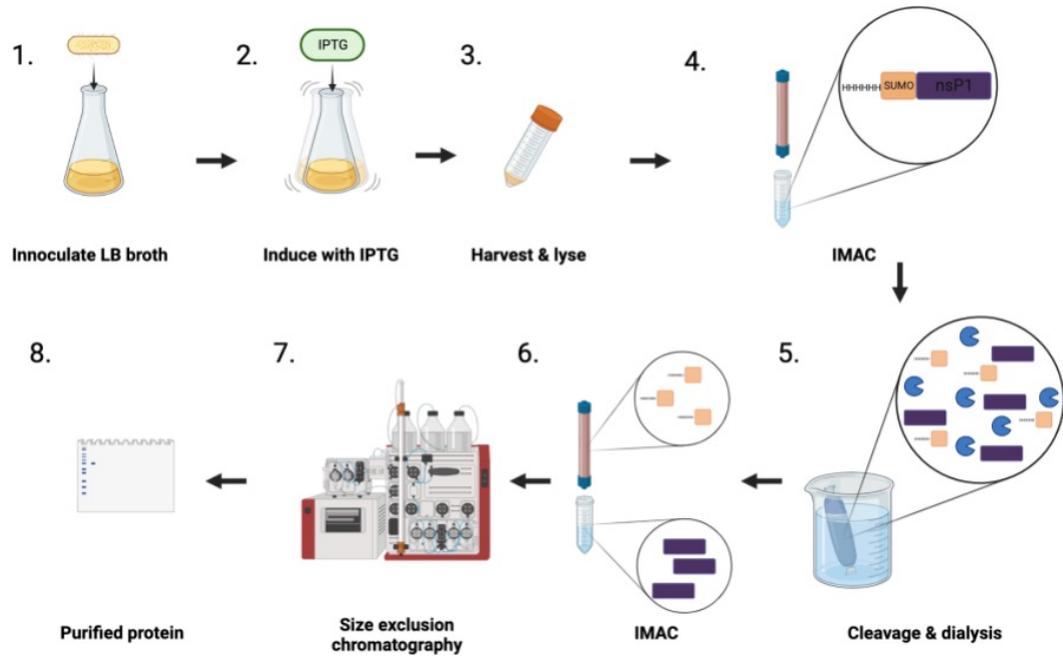


Figure 5.2: Schematic of the protein purification workflow for CHIKV nsP1

1 LB broth was inoculated with 6xHis-nsP1 transformed *E. coli* and incubated until an $OD_{600} = 0.4$ was reached. **2** Bacteria were induced with IPTG for 6xHis-nsP1 protein expression and incubated for 16 hrs at 30°C. **3** Bacterial cells were pelleted and resuspended in lysis buffer and lysed through sonication or a cell disrupter. **4** The supernatant was applied to a HisTrap column (IMAC), where the 6xHis tag bound to column and protein was eluted with increasing concentrations of imidazole. **5** Eluted protein was dialysed out of high imidazole buffer and SUMO protease was used to cleave the 6xHis-SUMO tag from CHIKV nsP1. **6** Sample was applied to a second HisTrap column, where the 6His-SUMO tag was bound and CHIKV nsP1 was eluted in the flow-through. **7** Size exclusion chromatography was performed to further purify CHIKV nsP1 from larger/smaller contaminant proteins. **8** Purity and integrity of protein was analysed by SDS-PAGE and Coomassie stain/western blot.

The first protein purification was performed using the protocol that had been optimised by previous members of the group. The column was washed with increasing concentrations of imidazole and a high salt wash to elute any bound RNA. 6xHis-nsP1 was then eluted with 500 mM imidazole, and a 1 M imidazole elution was used to elute any remaining bound protein (figure 5.3A). The four 500 mM elution fractions were pooled, cleaved and dialysed, to remove imidazole from the buffer and to cleave the 6xHis-SUMO tag for further purification steps. A band at ≈ 70 kDa was observed in the 500 mM imidazole lanes in figure 5.3A, showing that the elution of 6xHis-nsP1 was successful. It is of note that there

were other bands in these lanes suggesting there were other protein contaminants eluted as well. Despite this, all of the 500 mM imidazole fractions were taken forward further purification because they contained a high concentration of nsP1. A secondary HisTrap column was used to purify cleaved nsP1 from the 6xHis-SUMO tag and any other contaminants. Following cleaving of the 6xHis-SUMO tag from nsP1 with SUMO protease, 3 different sized bands were observed by SDS-PAGE and Coomassie staining: a band at ≈ 70 kDa was uncleaved 6xHis-nsP1, a band at 60 kDa was successfully cleaved nsP1 and a third band at 17 kDa was the SUMO-6xHis tag (figure 5.3B). The second HisTrap purification was a success as there was no band corresponding to SUMO-6xHis tag in the flow-through. However, there was a small amount of uncleaved 6xHis-nsP1 also present in the flow-through showing that this did not bind to the column.

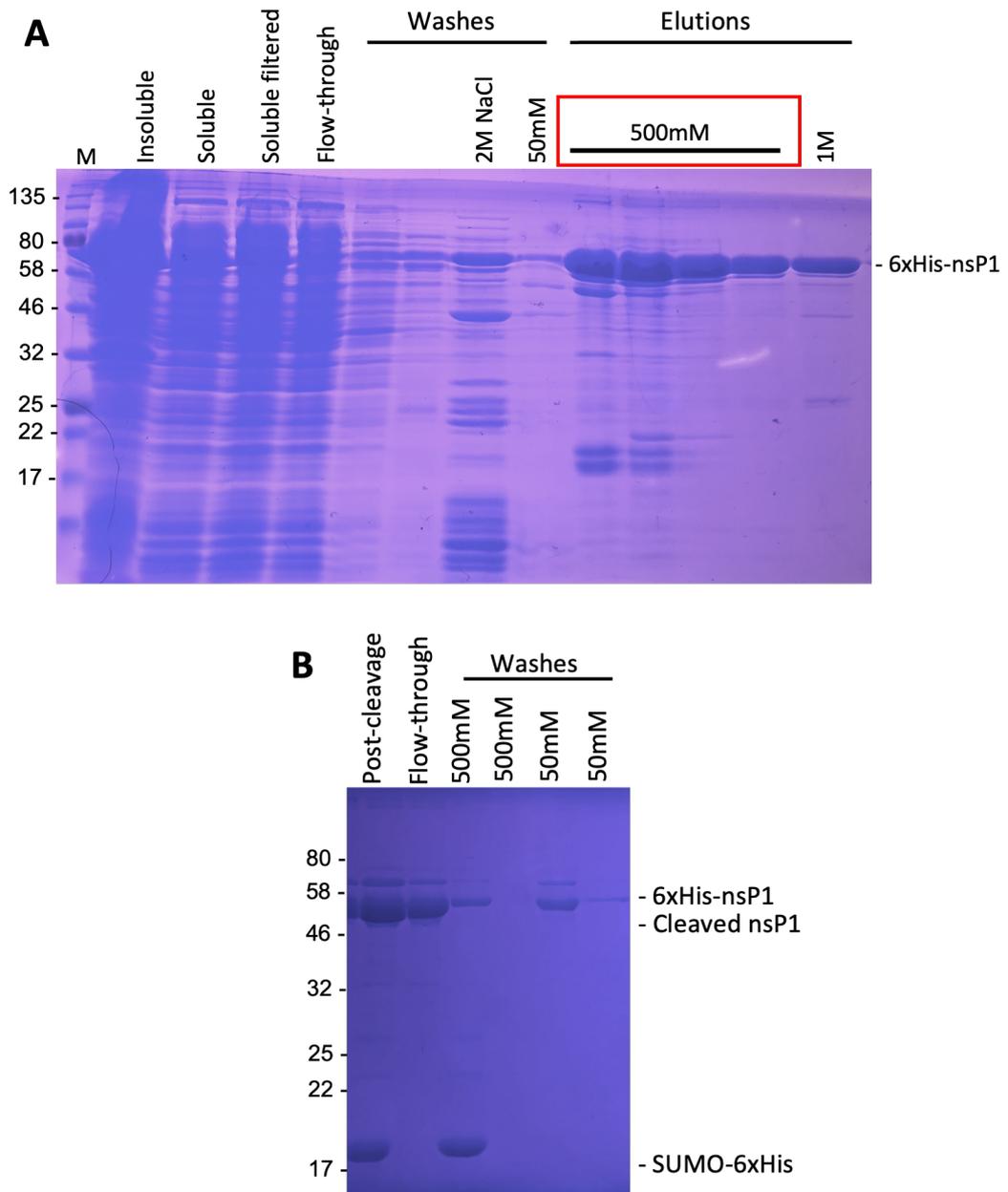


Figure 5.3: Primary and secondary HisTrap column fractions from the first purification.

A Fractions from the primary HisTrap column purification step were analysed by SDS-PAGE and Coomassie stain. The four 500 mM elutions were pooled together for dialysis and cleavage of the SUMO-6xHis tag. The cleaved product was applied onto a secondary HisTrap column (**B**) and native nsP1 was eluted in the flow-through and analysed by SDS-PAGE and Coomassie stain.

The flow-through from the secondary HisTrap column (figure 5.3B) was concentrated prior to injection into a HiLoad 75 Superdex pg column for size exclusion chromatography, to purify cleaved nsP1 from the uncleaved 6xHis-nsP1. Samples from the fractions correlating to the trace peak were analysed by SDS-PAGE

and Coomassie stain to confirm the purity of CHIKV nsP1 (figure 5.4). The un-cleaved 6xHis-nsP1 was eluted in the same fractions as nsP1, as there was a band present at ≈ 70 kDa and 60 kDa (representing 6xHis-nsP1 and nsP1 respectively). There was a much greater concentration of cleaved nsP1 within the fractions than 6xHis-nsP1. There were other bands also present in the eluted fragments that increased and decreased in concentration along with eluted nsP1 suggesting that they were aggregation/degradation products of nsP1. The larger protein at ≈ 200 kDa should have been separated from nsP1 during the size exclusion chromatography suggesting that the aggregation occurred after elution.

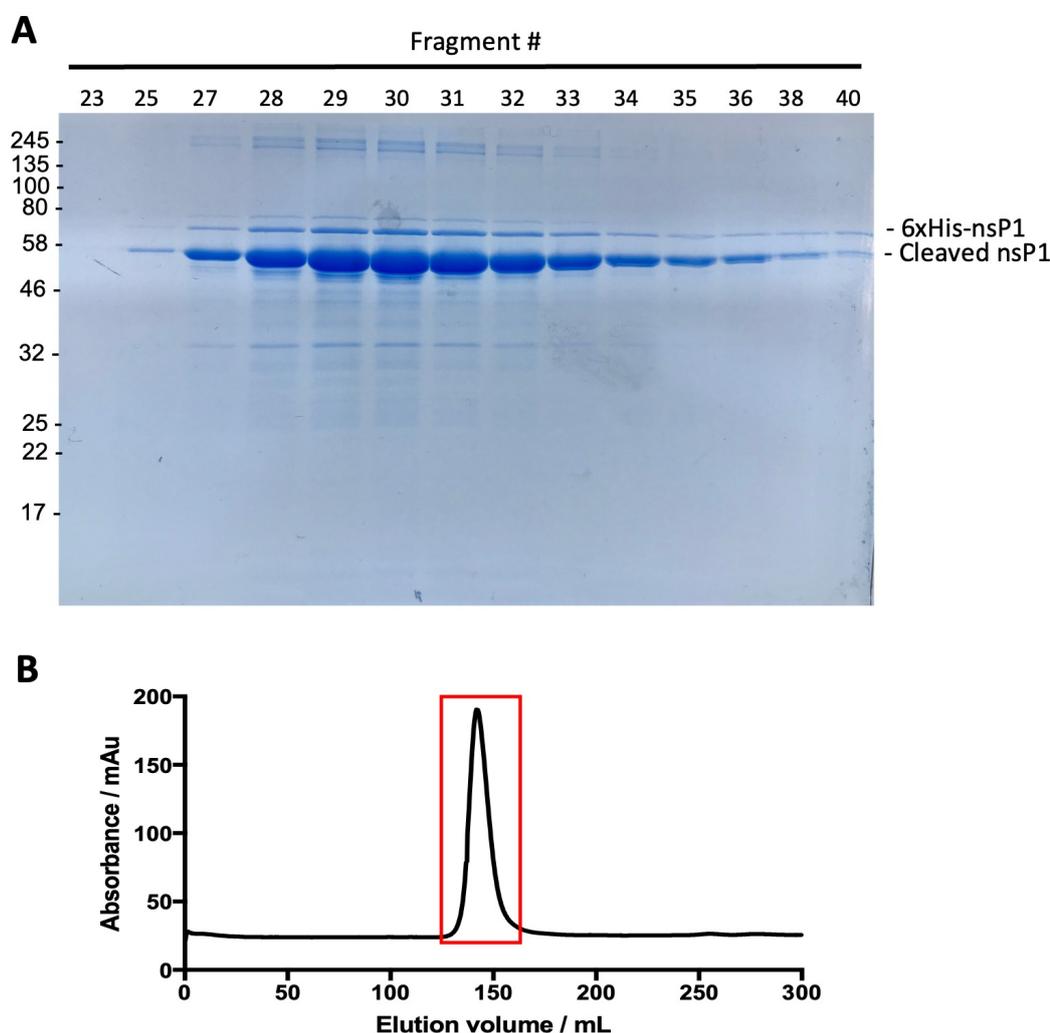


Figure 5.4: Size exclusion chromatography of the first purification.

The flow-through from the secondary HisTrap column purification step was purified by size exclusion chromatography. **A** The purity of nsP1 in the peak elution fractions was analysed by SDS-PAGE and stained with Coomassie. **B** The elution of CHIKV nsP1 is indicated by the peak in absorbance.

The purified nsP1 was screened for initial crystallisation conditions using the com-

mercially available Joint Centre of Structural Genomics Core (JCSG) suite screens (I-IV). These screens were designed to test the most successful and reproducible conditions from the JCSG and they sample a wide-range of precipitants, salts, buffers, pH, polymers and organic molecules in a high-throughput 96 well format, which allows for the testing of 384 different conditions. Importantly, these screens were compatible with an automated robot system which was utilised in order to efficiently and reproducibly set up sitting-drops with various protein:mother liquor ratios across all 384 conditions. Unfortunately there were no conditions that resulted in initial 'hits'. This may have been due to the number of other contamination bands present in the final purified sample.

Despite the presence of contamination products, circular dichroism (CD) was carried out to investigate the secondary structure of purified nsP1. The α -helices, β -sheets or random sheets within a protein have a unique absorbance spectrum which allows for the differentiation between different proteins, and to determine whether a protein has folded correctly. Analysis of the CD spectrum for purified CHIKV nsP1 displayed the characteristic troughs at 208 nm and 222 nm that are indicative of predominantly α -helical protein.

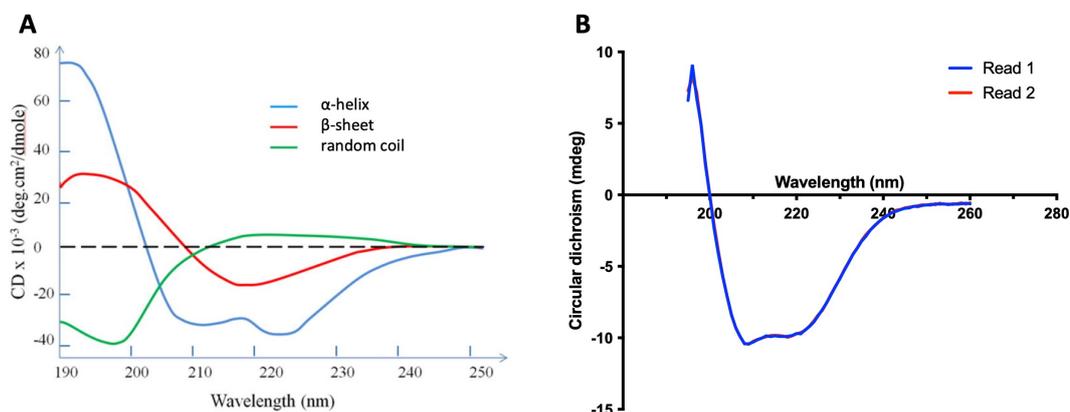


Figure 5.5: Circular dichroism spectrum of purified CHIKV nsP1

A Characteristic CD spectra. Adapted from (223) **B** CD spectrum of CHIKV nsP1 protein at 0.6 mg/mL. Troughs at 208 nm and 222 nm indicate a predominantly α -helical conformation.

To decrease the contaminants eluting alongside 6xHis-nsP1 on the primary HisTrap column, the concentration of imidazole in the washes prior to elution was increased, as well as the number of washes. In addition, 6xHis-nsP1 was initially eluted

with 300 mM of imidazole rather than 500 mM imidazole. There were far fewer contamination products in the 300 mM and 500 mM imidazole elutions from the second purification (figure 5.6A) compared to the 500 mM elutions in figure 5.3A. However, a band at 70 kDa in 100 mM and 120 mM imidazole washes showed that a lot of 6xHis-nsP1 was eluted in these washes and therefore the overall concentration of protein eluted from the primary HisTrap column was significantly reduced. The 300 mM and 500 mM elution fragments were dialysed and cleaved, prior to a second HisTrap column to purify nsP1 from the cleaved SUMO-6xHis tag (figure 5.6B). The second HisTrap column was successful in separating nsP1 from the cleaved SUMO-6xHis tag, as a band ≈ 17 kDa was seen in the post-cleavage lane, but not the flow-through lane. However, cleavage of the tag from nsP1 was not completely successful as there was a faint band at 70 kDa (indicative of 6xHis-nsP1) also present in the flow-through lanes for both the 300 mM fraction and 500 mM (figure 5.6B).

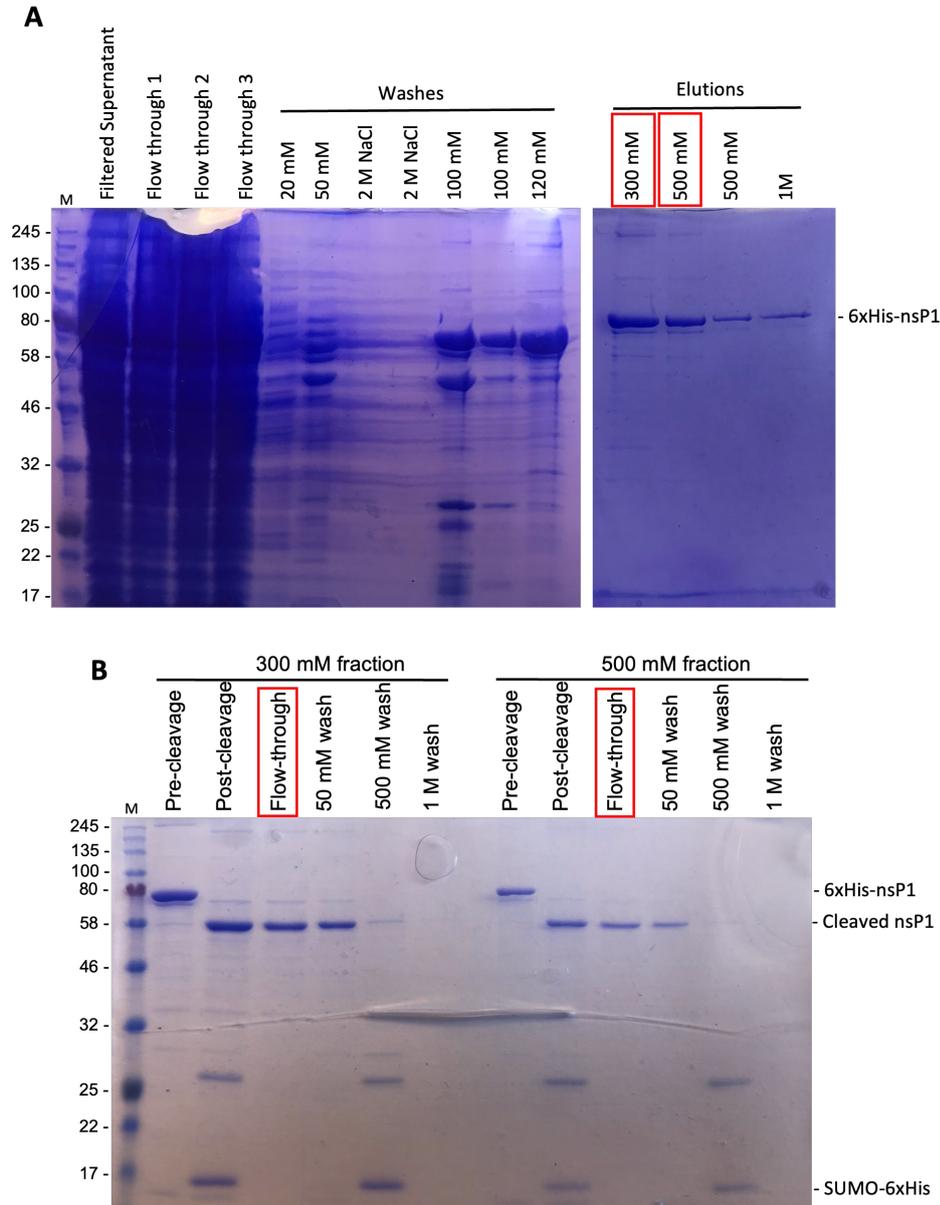


Figure 5.6: Primary and secondary HisTrap column fractions from the second purification.

Fractions from HisTrap column purification steps were analysed by SDS-PAGE and Coomassie stain. **A** Fractions from the primary HisTrap column purification step. The 300 mM and 500 mM elution fractions were dialysed and cleaved. **B** Fractions from the secondary HisTrap column purification step. The flow-through from the 300 mM and 500 mM elution fractions were carried forward for size exclusion chromatography.

The flow-through from both 300 mM and 500 mM fractions following the second HisTrap column were pooled together and concentrated prior to injection into a HiLoad 75 Superdex pg column for size exclusion chromatography. Samples correlating to the trace peak were analysed by SDS-PAGE and Coomassie stain to confirm the purity of CHIKV nsP1 (figure 5.7). There was an overall reduced

concentration of nsP1 purified following size exclusion chromatography. There were less contamination bands observed by SDS-PAGE and Coomassie analysis. However, there was a faint band at ≈ 70 kDa that would correspond to the size of uncleaved 6xHis-nsP1 (figure 5.7A). There was also a band at ≈ 200 kDa that is observed in fractions 13-15, which coincides with the fractions that contained the majority of eluted nsP1. Similarly to before, this band may correspond to aggregate products of nsP1 and this aggregation must have occurred following elution as a protein $\approx 3x$ the size of nsP1 should have been excluded during size exclusion chromatography. It is of note that SDS-PAGE should have denatured any aggregated forms, so it possible that this is a different protein contaminant.

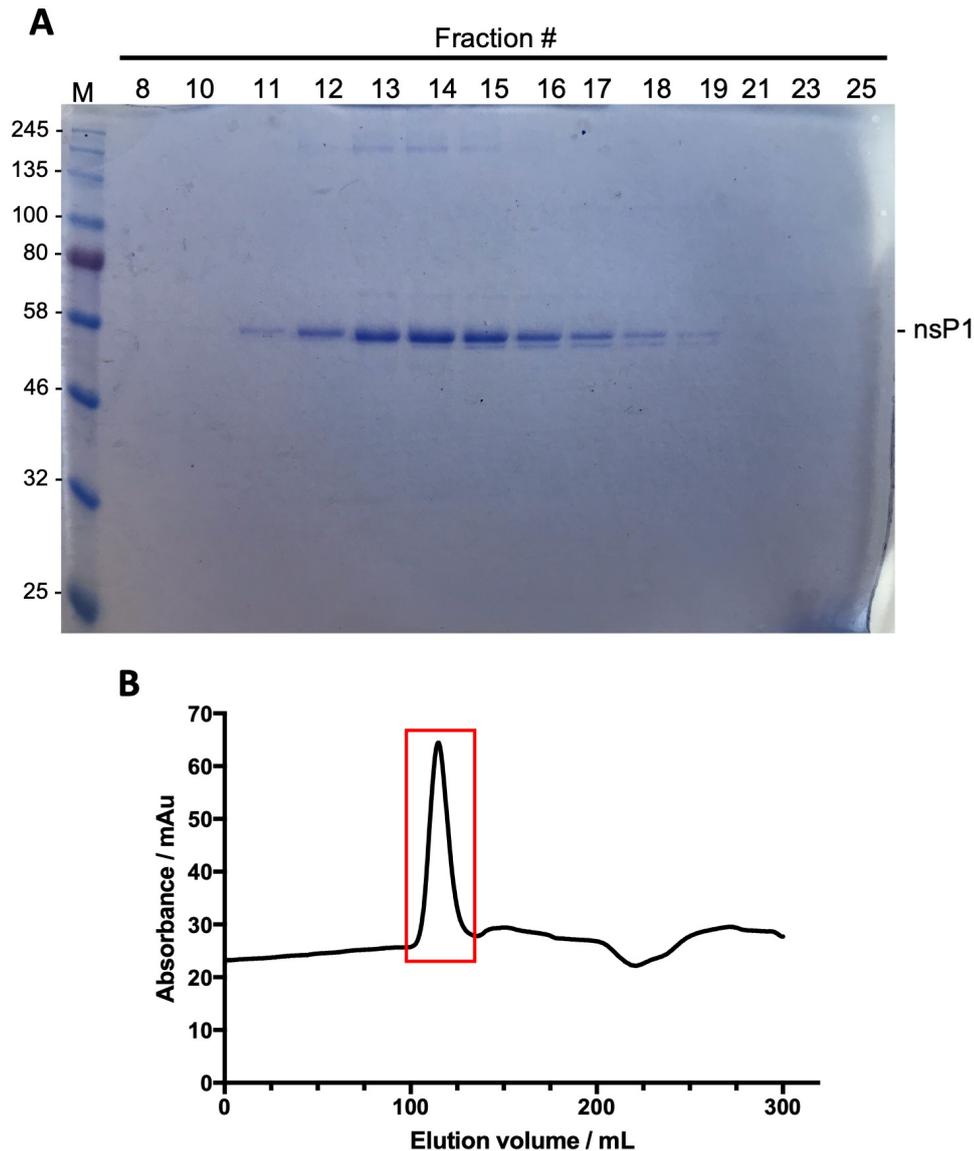


Figure 5.7: Size exclusion chromatography of the second purification
 Size exclusion chromatography was performed to further purify nsP1. **A** The purity of nsP1 in the peak elution fractions was analysed by SDS-PAGE and stained with Coomassie. **B** The elution of CHIKV nsP1 is indicated by the peak in absorbance.

Fractions 12-16 from figure 5.7A were pooled and concentrated to 1 mL, at a concentration of 0.8 mg/mL. To confirm that the protein was still stable following concentration, the sample was analysed again by SDS-PAGE and visualised by both Coomassie stain and western blot (figure 5.8). A neat sample was analysed alongside a 1:10 dilution. The western blot in figure 5.8B confirmed that the protein purified was CHIKV nsP1. The Coomassie stain (figure 5.8A) shows a single band at 58 kDa, suggesting that nsP1 was successfully purified. However,

lane 1 of figure 5.8B contains multiple contaminant bands, both bigger and smaller than nsP1. The majority of the contaminant bands were not visible in the 1:10 diluted sample (lane 2), this may be because the concentration was too low for detection by western blot.

Furthermore, throughout each purification stage a proportion of the desired protein is lost and this factor, combined with the minimal elution of 6xHis-nsP1 from the primary HisTrap column (figure 5.6), resulted in a low amount of nsP1 in the final product. The overall amount of nsP1 purified during this purification was under 1mg, whereas over 10 mg of nsP1 was purified following the first purification attempt. A higher concentration would be required for future experiments to determine whether purified protein was enzymatically active and for structural analysis.

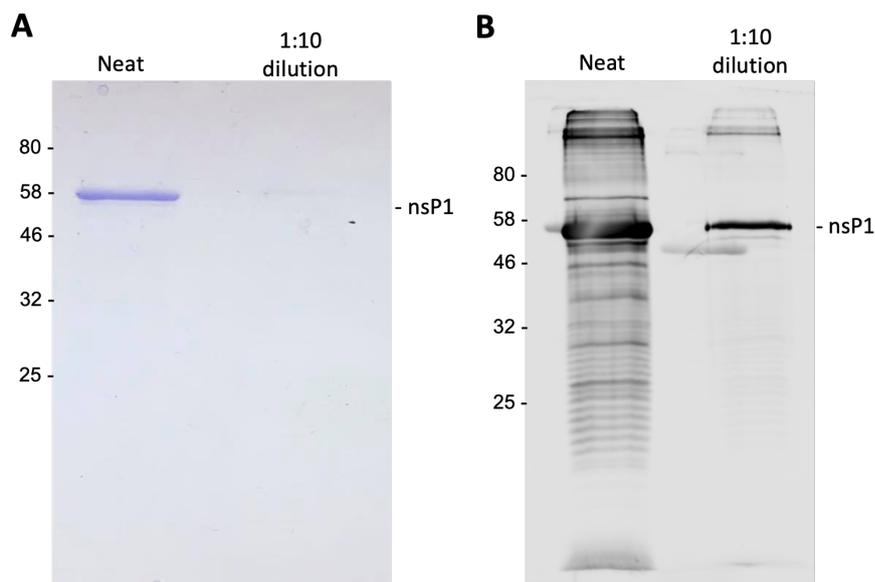


Figure 5.8: SDS-PAGE analysis of purified CHIKV nsP1

Purified recombinant CHIKV nsP1 was analysed by SDS-PAGE. **A** Protein was visualised by Coomassie stain, and **B** Protein was visualised by western blot.

To avoid the loss of nsP1 in higher concentration imidazole washes, another protein purification was performed following a similar methodology to the first, but with an increased number of 50 mM imidazole washes and the elution at 300 mM imidazole, (figure 5.9A). This was successful and a high amount of 6xHis-nsP1 was eluted within the 300 mM elutions, however, there was still

non-specific proteins being eluted alongside the desired nsP1. The 300 mM eluted fractions were pooled together, dialysed and cleaved, alongside both the 500 mM and 1 M elution. The cleaved and dialysed products of all fractions were run through a secondary HisTrap column in order to separate uncleaved 6xHis-nsP1 and the SUMO-6xHis tag from cleaved nsP1 (figure 5.9B). This was successful, as can be seen by a band at ≈ 17 kDa not being present in the flow-through of the 300 mM and 500 mM fractions.

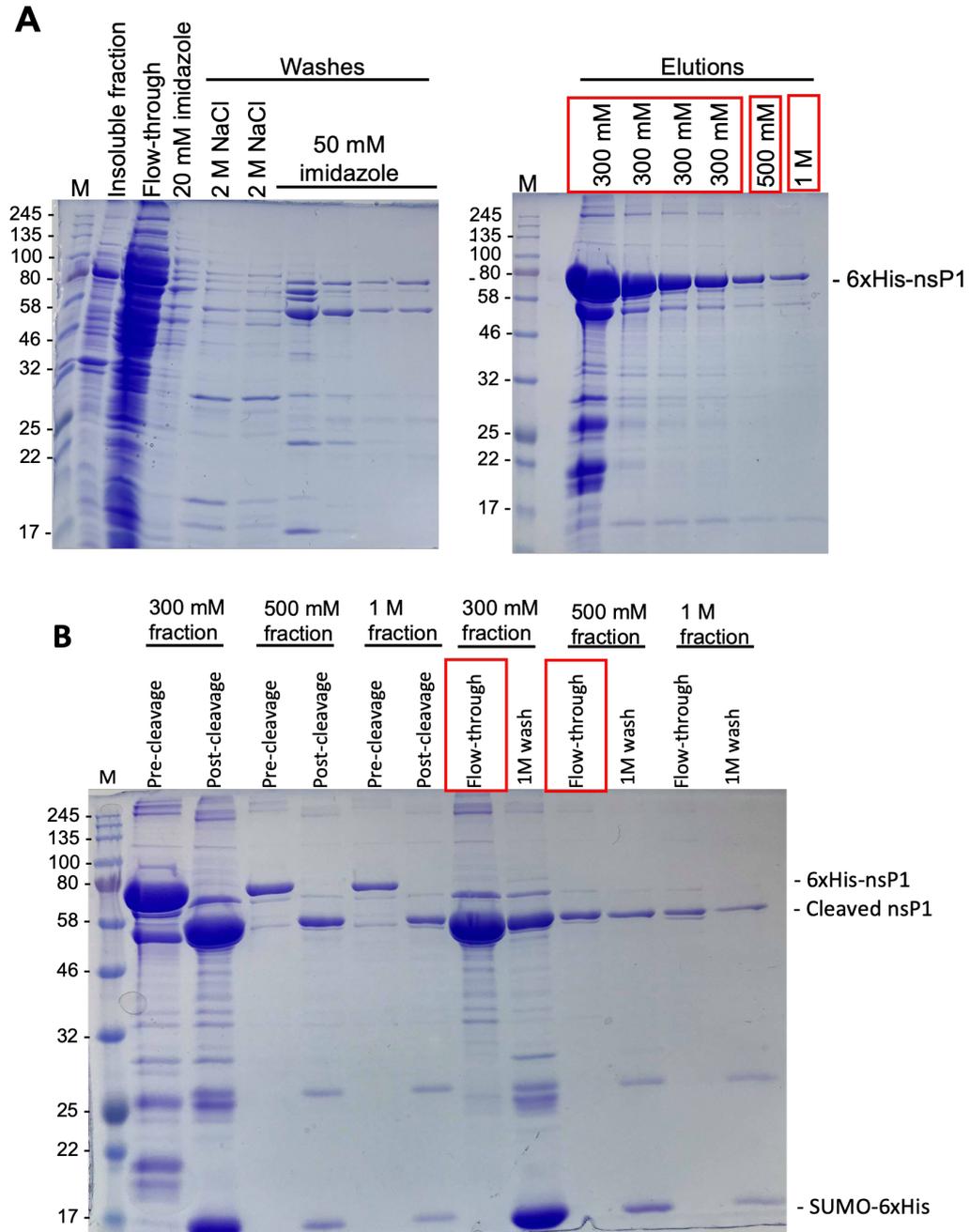


Figure 5.9: HisTrap column fractions from the third purification.

A The fractions from the primary HisTrap column purification step were analysed by SDS-PAGE and Coomassie stain. The four 300 mM eluted fractions were pooled together, and along with the 500 mM and 1 M eluted fractions, were cleaved and dialysed for further purification by a secondary HisTrap column. **B** Fractions from the secondary HisTrap column purification.

The flowthrough from the secondary HisTrap column of the 300 mM fraction and 500 mM fraction were pooled together and concentrated and then injected into a HiLoad 75 Superdex pg column for size exclusion chromatography. Samples

from the fractions correlating to the trace peak were analysed by SDS-PAGE and Coomassie stain to confirm the purity of CHIKV nsP1 (figure 5.10). The overall amount of protein purified was 20 mg, which is significantly more than the previous two purifications. Unfortunately, the total purified sample was not pure enough to go forward for crystallography trials. Similar to the pattern seen during the first purification (figure 5.4), the contaminant bands increased and decreased in concentration along with nsP1 (figure 5.10A). Furthermore, nsP1 was eluted from the size exclusion column approximately 50 mL earlier than observed in the first purification (figure 5.4B). The elution volume observed during this purification was much closer to the void volume and suggests that nsP1 was beginning to aggregate during concentration after the secondary HisTrap column as a lower elution volume correlates to a larger protein size.

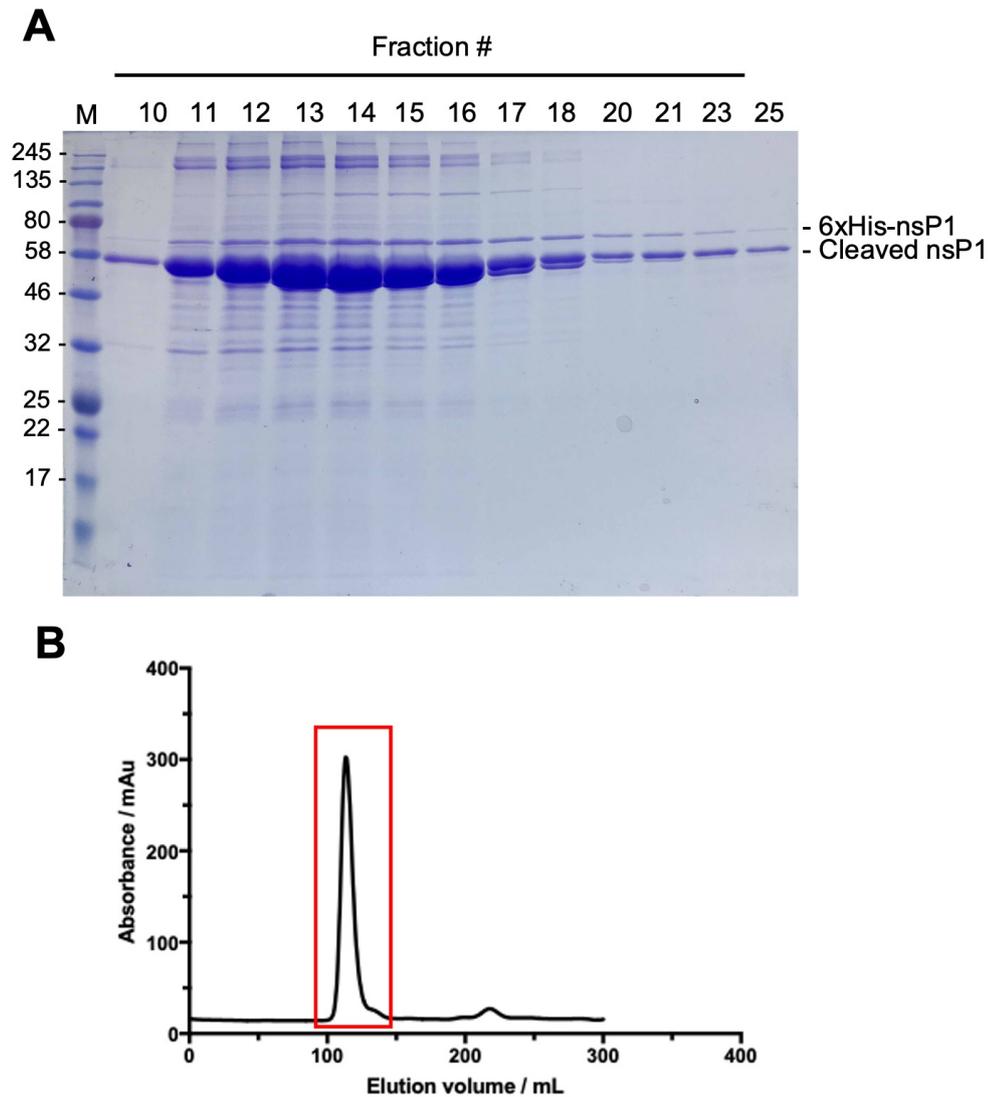


Figure 5.10: Size exclusion chromatography of the third purification. The flow-through of the 300 mM fraction and 500 mM fraction from the secondary HisTrap column purification step were pooled together and further purified by size exclusion chromatography. **A** The purity of nsP1 was analysed by SDS-PAGE and stained with Coomassie. **B** The elution of CHIKV nsP1 is indicated by the peak in absorbance.

5.3 Discussion

The results from this chapter have demonstrated that CHIKV nsP1 was successfully expressed and purified utilising an *E. coli* system and a 6xHis-SUMO tagged nsP1. The system was optimised to produce a high concentration of nsP1 following two rounds of IMAC and size exclusion chromatography. Unfortunately,

contamination bands continued to be observed following size exclusion chromatography, suggesting that nsP1 was potentially aggregating or degrading following elution from the column. It is also of note that later size exclusion chromatography attempts saw the elution of nsP1 shortly after the void volume, suggesting that nsP1 may have begun to become aggregated during this purification method (figure 5.7B and 5.10B). There is a possibility that the shift in elution of nsP1 closer to the void volume could have been due to the oligomerisation of nsP1 into the dodecameric form presented by Zhang et al. and Jones et al. (113, 205). In order to confirm this a calibration curve would be required of proteins with known sizes. We were able to determine that nsP1 had a mostly α -helical structure by CD analysis 5.5.

Further experiments could be performed to determine whether the purified protein contained MTase and/or GTase activity. The enzymatic activity of nsP1 could be compared between different purification attempts and also to compare the protein that was eluted in different fractions across the purification workflow. Furthermore, purified nsP1 should have been analysed on a native gel to gain an deeper understanding into whether nsP1 was aggregating/forming dodecamers during different steps of the purification workflow.

Over the course of this study, two publications that determined the structure of CHIKV nsP1 became available (113, 205). The findings from these studies are discussed in more detail below. Of note, the publications expressed the protein in either mammalian or insect cells (whereas a bacterial expression system was utilised in this study) and both studies used a cryo-electron microscopy (cryo-EM) approach, rather than X-ray crystallography. Cryo-EM has greatly improved in recent years and it is no longer limited to the elucidation of macromolecular structures, and has been used to determine the structure of a 64 kDa protein (224). One of the major advantages to cryo-EM is that it is able to bypass the crystallisation step required for X-ray crystallography. However, it is important to mention that X-ray crystallography is able to generate a higher-resolution molecular structure than cryo-EM.

Jones et al. reported the first structure of CHIKV nsP1 in 2020, utilising cryo-

EM (113). As with this study, they used a bacterial expression system to express and purify recombinant nsP1 but also compared this to a Baculovirus expression vector system in High Five™ insect cells. They demonstrated that the protein purified from High Five™ cells was membrane-associated. They also demonstrated that nsP1 purified from insect cells had GTase activity as the cap intermediate m^7G was observed to covalently bind to purified oligomeric nsP1 when incubated with GTP and SAM. The monomeric nsP1 form purified from bacterial cells did not display GTase activity. A bacterial expression system may not produce oligomeric/enzymatically active nsP1 as bacteria do not have a nucleus/endoplasmic reticulum/golgi apparatus which may be required for the correct post-translational processing of CHIKV nsP1. The structure reported showed that 12 monomers of nsP1 form a ring structure that is membrane-associated; with three regions they refer to as the crown, the waist and the skirt. The crown region is shown to be comprised of the 12 capping domains of each monomer; the inner walls of the waist are shown to be negatively charged and they suggest that this would allow for proteins <70-90 kDa to pass through; and the skirt contains the membrane-binding domains. They suggest the structure is acting as viral nuclear pore complex, whereby it controls the entry and exit of molecules into and out of the replication complex.

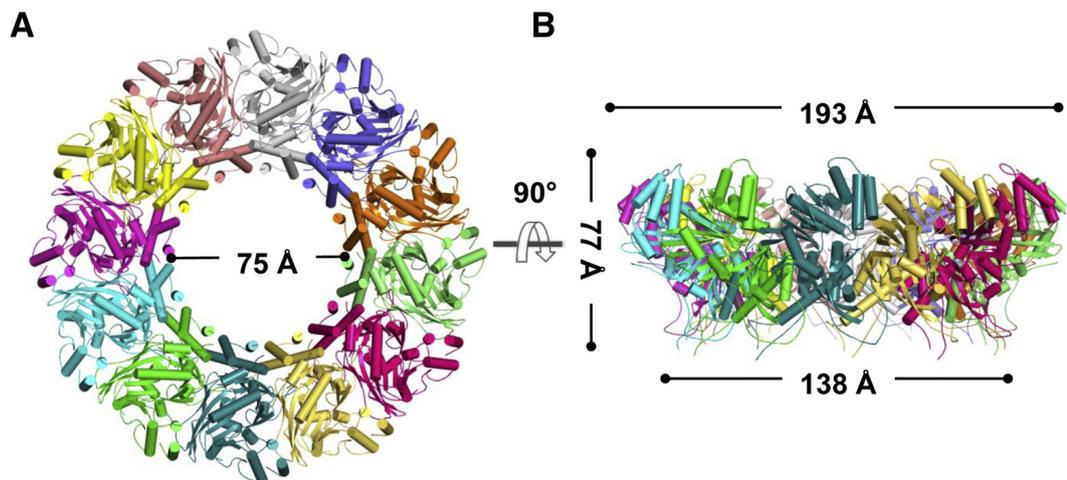


Figure 5.11: Crown-shaped ring structure of CHIKV nsP1

The structure of nsP1 reported by Zhang et al. 12 copies of nsP1 form a crown-shaped ring that is known to bind to the host plasma membrane. **A** The “top-view” of the ring structure, each monomer is a different colour. **B** The “side-view” of the ring structure. Image adapted from (205).

Zhang et al. also presented a structure of CHIKV nsP1 by cryo-EM in 2021 (205). Similarly to Jones et al. (113) they showed that 12 copies of nsP1 form a crown-shaped ring structure (figure 5.11). They began by purifying nsP1 from *E. coli* and human Expi293F cells; they observed 3 size peaks during size exclusion chromatography of protein expressed in *E. coli* but only 1 peak with protein expressed in Expi293F cells. Gel-based MTase/GTase assays were utilised to confirm that nsP1 purified from Expi293F cells and the higher molecular weight protein from *E. coli* were enzymatically active, and they went on to use protein expressed in Expi293F cells for their structural studies.

Both structures published show nsP1 forming a dodecameric ring structure with C12 symmetry, and both groups highlight the importance of nsP1 in the formation of the replication complex. The structure proposed by Zhang et al. is larger than the Jones et al. structure, with the most significant difference between the two models being in the membrane associated loops situated in the skirt region (113, 205). Zhang et al. produced a model of the transition between the structures and suggest that the closing of the membrane association loops is coordinated with the opening of the top/crown region of the ring; further supporting the current understanding that membrane association of alphavirus nsP1 is able to modulate the MTase/GTase activities of the protein (225).

As mentioned, both of these studies noted that the protein purified from bacterial cells was not enzymatically active. If future experiments are to be performed to further confirm that introducing a different amino acid at residue 24 does not impact the catalytic function of nsP1, the expression and purification of nsP1 should be performed using an alternative expression system to bacterial cells. It would also be interesting to express a truncated nsP1 (removal of residues 1-24) to determine whether this form would still be able to form the dodecameric ring, or if the loss of the first 23 nucleotides is detrimental to the interactions between monomers.

Given that a crystal structure of nsP1 was published throughout this study, the work in this chapter was not taken further in order to focus on the virus and molecular biology aspect of the project. Furthermore, the two published studies

demonstrated that bacterially expressed CHIKV nsP1 was not enzymatically active and therefore the protein I had successfully managed to purify would not have been viable for *in-vitro* enzymatic studies.

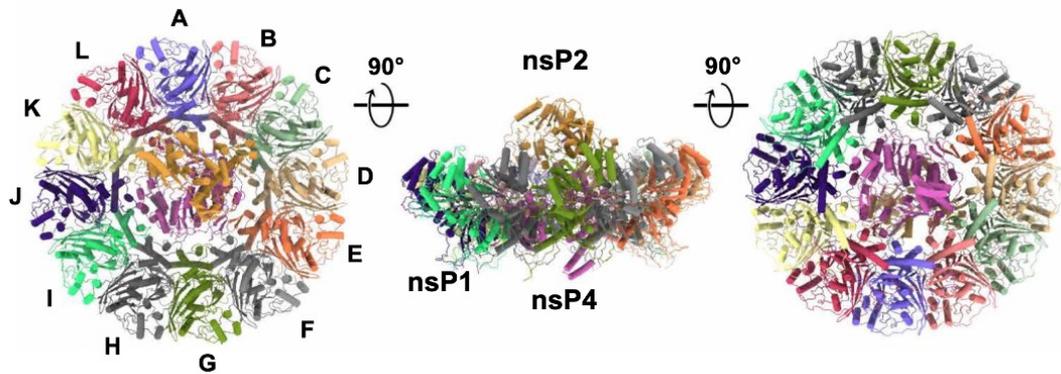


Figure 5.12: Molecular structure of CHIKV replicase complex

The structure of the CHIKV replicase complex reported by Tan et al. 12 copies of nsP1 form a crown-shaped ring with nsP4 situated in the centre 'pore' and interacts with nsP2. The top, side and bottom view are presented respectively. (73).

A third report was published in December 2022 which determined a high-resolution image of the replication complex in its cellular context (73). The newly published structure showed that an nsP4 monomer sits in the pore formed by the nsP1 dodecameric ring (figure 5.12). This structure also demonstrated that nsP4 and nsP2 are in complex together within the replicase complex. Tan et al further discuss how the enzymatic activity of nsP4 is only fully activated following correct folding within the central pore of the nsP1 dodecameric ring, and that interactions with nsP2 further increase nsP4s enzymatic activity.

Chapter 6

Discussion

My initial primary focus was to determine whether the second, in-frame methionine (M²⁴) of nsP1 was functioning as an alternative start codon for the synthesis of truncated nsP1, which may have had previously undetermined roles in the CHIKV lifecycle. The combination of using a full-length infectious clone alongside a *trans*-replicase assay allowed me to determine that whilst virus replication was reduced (figure 3.7), mutating M²⁴ was not detrimental to the function of nsP1 and the replicase complex (figure 3.14). Instead, the *trans*-replicase assay demonstrated that mutations at M²⁴ in the RNA template decreased replication of the 49S (full-length) and 26S (sub-genomic) RNA (figure 3.14). This data suggested that this region of the RNA molecule was important for interactions during genome replication. This finding correlated to published literature, whereby it was demonstrated that a reduction in virus replication was due to the alteration in the 5' CSE RNA structures and not due to amino acid substitutions in the nsP1 coding region (174). Initially Madden et al. disrupted the CSE structures through broad range mutagenesis across the region and observed a reduced viral titre phenotype. A second set of mutations were then introduced that would maintain the CSE replicatin elements but alter the amino acid sequence, and they observed no reduction in viral replication. The results from chapter 3, alongside the literature, concluded that nsP1 is able to tolerate amino acid changes at the N-terminus but alterations to the RNA sequence are deleterious.

Whilst the *trans*-replicase assay clearly demonstrated that genome replication was inhibited in the M²⁴ mutant in relation to wild-type, the sub-genomic and ss qRT-PCR results did not show a reduction in mutant genome replication when compared to wild-type CHIKV. The levels of positive- and negative-sense RNA were measured by ss qPCR in both mammalian and mosquito cells, whereas the *trans*-replicase assay and sub-genomic replicon experiments were only performed

in mammalian cells. It would be wise to repeat the replicon work in mosquito cell lines (ideally both C6/36 (*Ae. albopictus*) and Aag2 (*Ae. aegypti*)) to confirm whether genome replication is reduced in a host-dependent manner as there are a variety of published studies across the genus demonstrating how mutagenesis of the 5' coding region of nsP1 is tolerated in a host-specific manner (173, 174, 184, 226).

Collectively the results from chapter 3 demonstrated that M²⁴ does not function as a non-canonical, alternative start codon for the synthesis of a truncated nsP1, and a new hypothesis was proposed that it was the alteration of primary RNA sequence, rather than amino acid sequence, that inhibited CHIKV replication. This could have been due to the disruption of RNA secondary structures that are known to be essential for efficient CHIKV generation (173, 174), or inhibition of sequence specific RNA:protein, or RNA:RNA, interactions during genome replication.

To further investigate how the mutations introduced were affecting the RNA template, the RNA secondary structures were analysed by SHAPE-constrained thermodynamic models. The wild-type structure presented in this thesis more closely matched the model proposed by Madden et al., with SL85 reported by Kendall et al. not observed (figure 4.5). The presence of SL85 was confirmed to be true and functional in this study through i) reverse genetics and ii) compensatory mutations to restore the formation of SL85. It was proposed that this region is highly dynamic and/or is involved with tertiary interactions, which may explain why SL85 was not formed in the model predicted in this thesis.

The only mutant not observed to alter the RNA structure of the 5' region was M²⁴S, which matched very closely to the predicted wild-type structure (figure 4.8). This suggests that the RNA sequence and presence of RNA structures are both important for efficient CHIKV replication. RNA is understood to undergo complex changes depending on cellular conditions (such as pH, protein:RNA interactions, temperature) (227–229) and therefore analysis of the altered RNA secondary structures should be performed in cell culture. It is of note that Kendall et al. saw no significant difference in RNA structure when the *in-vitro* transcribed

RNA was folded at 37°C or 28°C so it is likely that differences in RNA secondary structures will not be observed between mammalian and mosquito cells.

Across RNA secondary structure models of the M²⁴ mutant panel, all 5 mutants conserved at least SL102 and SL194. Despite the M²⁴LII mutant displaying a greater reduction in virus replication in comparison to M²⁴LI, relative to wild-type (figure 3.7), the predicted RNA secondary structures were almost identical (figure 4.9 and 4.10). Both of these mutants were observed to contain SL102, and SL164 and SL195 of the 5' CSE, and a previously uncharacterised short stem loop was observed from nucleotide 145 (SL145). The presence of the genus-wide conserved structures (SL102, SL165 and SL194) may have suggested that the M²⁴LI and M²⁴LII mutants should replicate to wild-type levels, but perhaps the novel SL145 was capable of masking these structures, through interactions with long-range sequences to form tertiary structures which would make SL102, SL165 and SL194 inaccessible. This novel SL145 was 11 nts long in both mutants. However, the M²⁴LI apical loop contained 5 exposed nts where the M²⁴LII apical loop only contained 3 nts.

Previously published studies have utilised passaging of mutations known to inhibit viral replication to gain insights about protein:RNA and RNA:RNA interactions in genome replication (171, 180, 181, 184). This approach was used to determine mechanistic interactions of the M²⁴ codon during CHIKV genome replication. The M²⁴A mutant was taken forward for passaging experiments as it was known to i) inhibit virus replication, ii) decrease genomic replication, iii) disrupt RNA secondary structures.

There was an observed increase in M²⁴A mutant replication after 10 rounds of passaging in Huh7 cells and alternately between Huh7 and C6/36 cells, with levels of replication being similar to that of wild-type by P10 (figure 4.15 and 4.16). Extracted total RNA was analysed by NGS to determine whether any mutations were gained across the CHIKV genome during passaging of M²⁴A mutant and wild-type CHIKV. In both the mammalian and alternate passaging experiments, a single nucleotide C¹⁴⁷U mutation was observed in M²⁴A CHIKV. This resulted in an amino acid change of M²⁴V, however results from this thesis and the literature

strongly suggest that this amino acid alteration would not affect the function of nsP1. The *trans*-replicase assay in chapter 3 clearly determined that the presence of M²⁴A in the replicase complex did not impact genome replication (figure 3.14). Experiments performed in published studies have demonstrated that reduction in virus replication was due to disruption in the RNA secondary structure and not due to changes to the amino acid sequence (174).

In-silico analysis predicted minimum free energy models showed that C₁₄₇^U would allow base pairing to reform between the A nucleotide in the ORF-1 AUG start codon (nucleotide 77) and M²⁴ (nucleotide 147), which was observed by Kendall et al., and Madden et al (figure 4.21A). The reformation of this base-pairing restored wild-type RNA secondary structure. A SHAPE-constrained thermodynamic model confirmed that the C₁₄₇^U mutant was sufficient to restore wild-type base-pairing (figure 4.23). This supports the understanding that the structures within the nsP1 coding region were essential for efficient CHIKV replication. Furthermore, it demonstrated that introducing mutations at the M²⁴ codon resulted in alterations to the RNA structures that were rescued through passage of M²⁴A in C6/36 cells by the selection of a reversion mutation at C₁₄₇^U.

As well as the discussed C₁₄₇^U mutant, two further mutations were selected in nsP1 and nsP3 (nucleotides G⁴²⁵A and G⁴⁴²⁴A respectively) encoding regions during serial M²⁴A CHIKV passage. To investigate whether these mutations were required in conjunction with the C₁₄₇^U mutant they were re-introduced into the full-length CHIKV infectious clone (deemed M²⁴V combination), as well as the C₁₄₇^U alone (deemed M²⁴V). There was a greater titre of virus recovered following transfection of BHK-21 cells with RNA in the M²⁴V combination virus than M²⁴V, suggesting that mutations in nsP1 and nsP3 were required for the M²⁴V to enhance CHIKV replication. It is of note that there was also reduced replication of the M²⁴V combination mutant in comparison to wild-type, but the data presented is n=1 and it is therefore essential that further repeats are carried out before drawing a conclusion as to whether these observed mutants are sufficient to recover wild-type levels of replication.

When M²⁴A was passaged in C6/36 cells a SNP (nucleotide A⁷⁷G) was observed

(which caused an amino acid change of M²⁴V at the ORF-1 AUG start codon). This was surprising as this amino acid change would remove the ORF-1 start codon and thus presumably prevent translation of the non-structural proteins. It is of note that the M²⁴A mutations were conserved across this entire passaging experiment. The A⁷⁷G substitution was only observed in $\approx 75\%$ of the population and the remaining population was the wild-type sequence. The A⁷⁷G mutant was re-introduced into the full-length ICRES clone alongside M²⁴A (deemed M¹V & M²⁴A), in order to determine whether this was sufficient to restore replication to wild-type levels. However, there was no virus recovered following transfection of BHK-21 cells with RNA. This mutant was only observed during passaging in mosquito cells and therefore was likely to be a host-specific mutation that was not viable in mammalian cells. Further work could be performed to determine whether this mutant replicates to wild-type levels in mosquito cells.

It was hypothesised that the continued presence of wild-type sequence amongst the M¹V & M²⁴A virus was used in *trans* by the mutant virus to allow for more efficient replication. However, this was disproved when both wild-type and M¹V & M²⁴A mutant RNA were transfected into the same cells and only wild-type sequence was detected by NGS following 2 rounds of passaging (figure 4.20). Suggesting that either the wild-type genome was preferentially replicated over mutant, or that the mutant genome reverted to wild-type. Finally, a SHAPE-constrained minimal free energy predicted map of M¹V & M²⁴A CHIKV revealed that the SNP at A⁷⁷G was sufficient to restore wild-type base pairing for the SL85, SL102, SL165 and SL194 replication elements known to be critical for efficient CHIKV replication (173, 174). This supports previous results demonstrating that mutations at M²⁴ disrupt the RNA secondary structures in the 5' coding region of nsP1 which causes a reduction in CHIKV replication.

6.1 Concluding remarks

The work comprised in this thesis has furthered our understanding into the importance of RNA secondary structures within the 5' region of CHIKV. The

introduction of mutations at the M²⁴ codon impacted genomic replication due to alterations in the RNA molecule, rather than alterations to the amino acid sequence of nsP1. This confirmed that the second in-frame start codon (M²⁴) does not act as an alternative start codon to produce a truncated nsP1. Furthermore, substitutions at M²⁴ altered the RNA structure in the 5' coding region of nsP1. The wild-type structure was recovered by passaging of M²⁴A in mammalian and insect cells through the selection of SNPs at either M²⁴ or elsewhere in the genome.

Chapter 7

Appendix

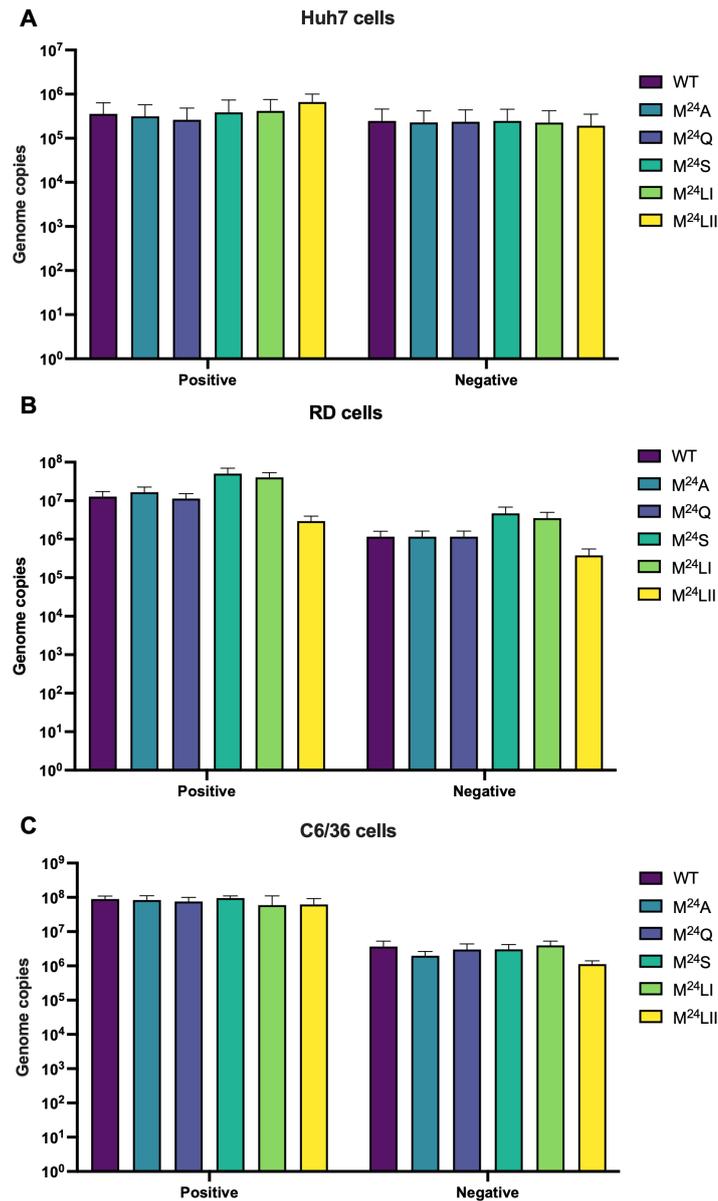


Figure 7.1: Strand specific qPCR of WT and M²⁴ mutations

Huh7, RD and C6/36 cells were infected with WT CHIKV and the M²⁴ mutant panel, at an equal MOI = 0.1. RNA was extracted from the monolayers 24 hpi with TRI-reagent. The RNA was reverse transcribed using primers specific to positive and negative sense CHIKV RNA. **A** Data from Huh7 cells. **B** Data from RD cells. **C** Data from C6/36 cells.

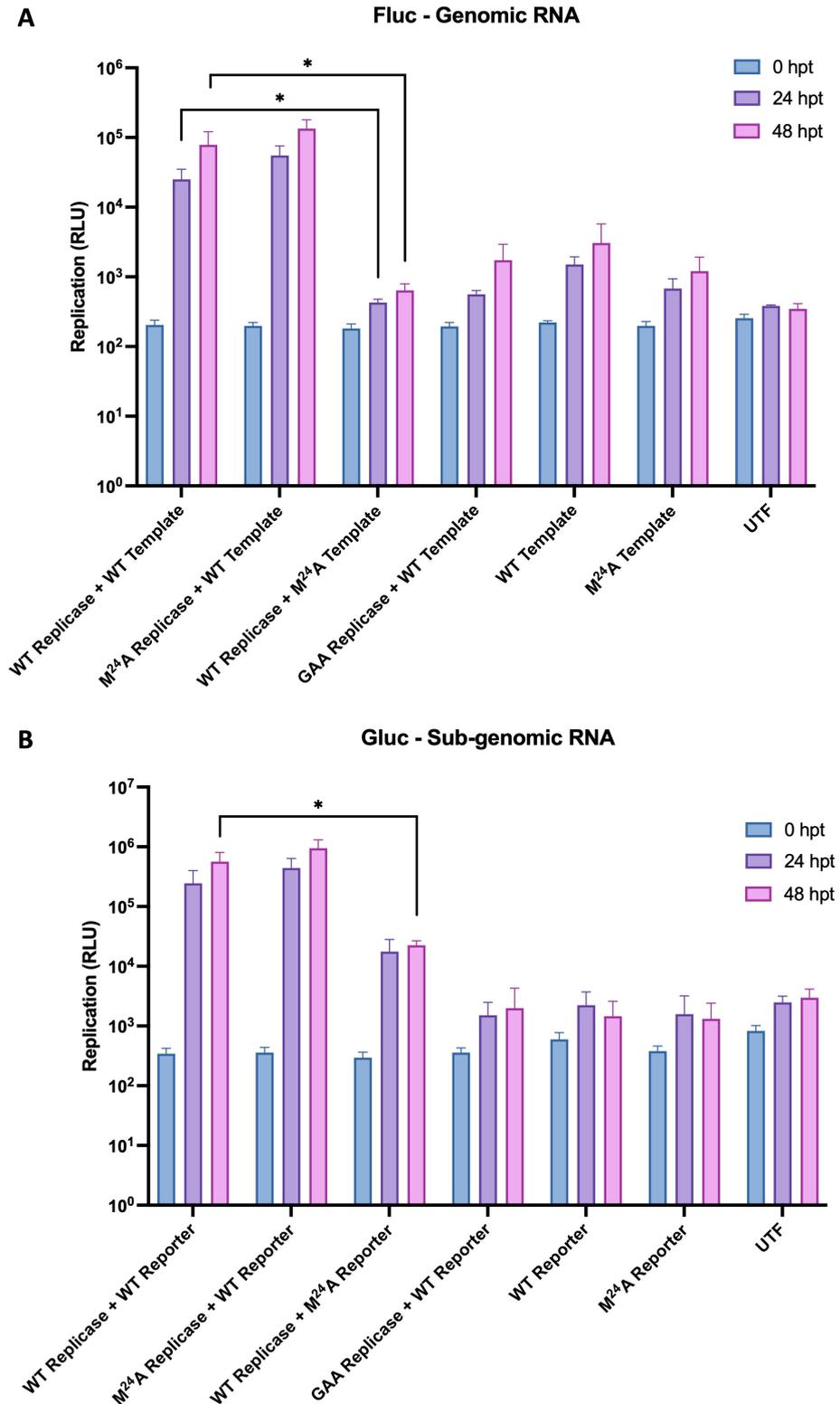


Figure 7.2: Introduction of M²⁴A mutant in the *trans*-replicase assay BHK-21 cells were transfected with the replicase + reporter constructs and lysed 48 hpi. Genome replication was measured by Fluc and Gluc expression. **A** Fluc levels, corresponding to viral genomic RNA replication. **B** Gluc levels, corresponding to sub-genomic transcription. Error bars represent SEM and significance is shown relative to wild-type. * P<0.05. n=3.

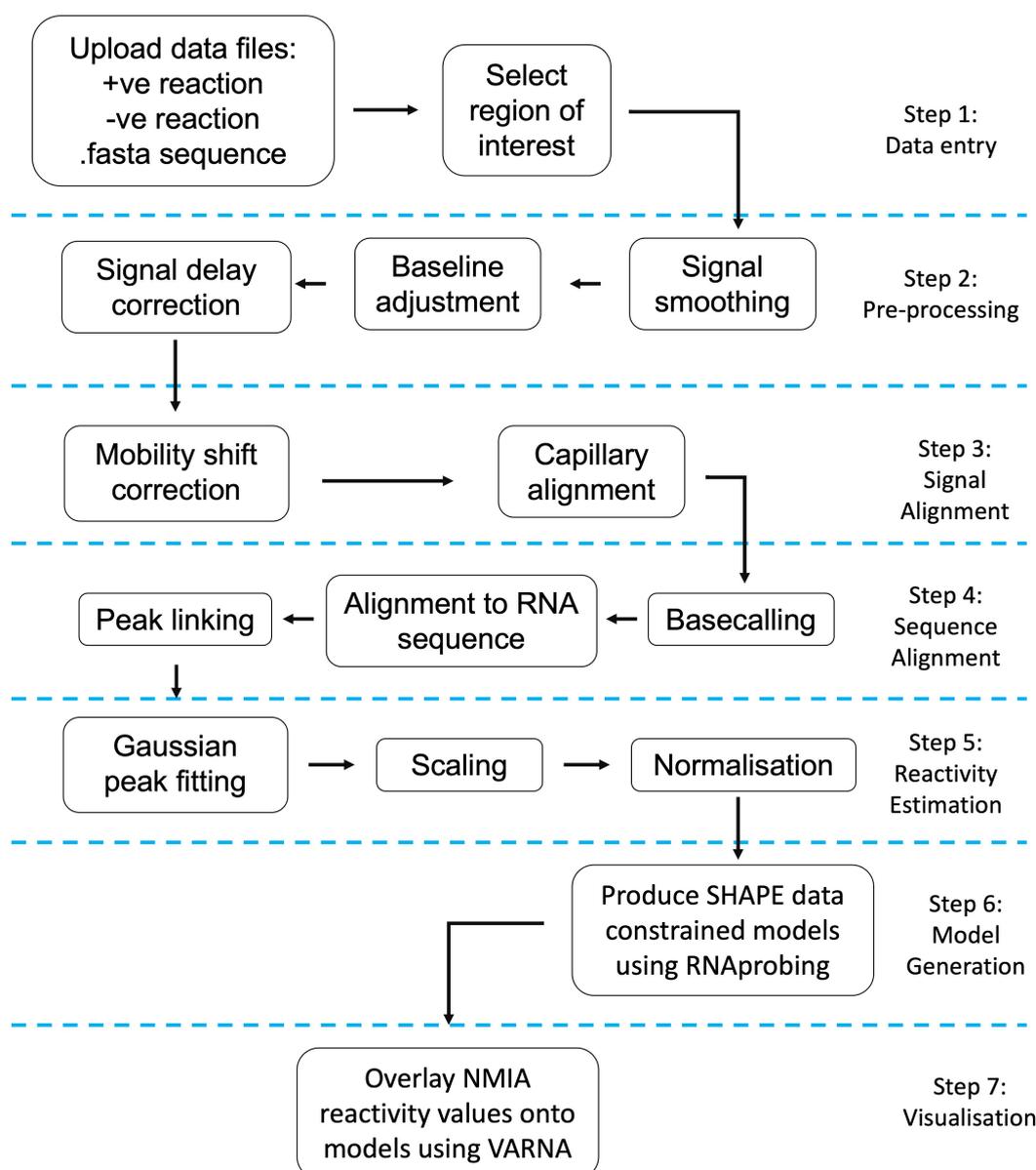


Figure 7.3: SHAPE data analysis workflow

1 Data from NMIA treated (+ve) and control DMSO (-ve) reactions, and a .fasta file containing the RNA sequence, are uploaded to QuSHAPE and the region of interest is selected. **2** Pre-processing operations remove high-frequency noise and baseline offset. Signal delay correction converts the fluorescent signal intensities to probabilities of primer termination. **3** The use of different fluorescent labels (FAM and HEX), and use of different capillaries, results in different retention times. The mobility shift correction aligns pairs of signals within each capillary. Capillary alignment aligns the signal across the two capillaries. **4** Basecalling classifies all of the peaks in the sequencing signal from the DMSO (-ve) capillary as specific peaks produced by ddATP and non-specific peaks corresponding to the other bases. The signal from the DMSO channel (FAM) is then aligned with the RNA sequence signal (HEX). Peak linking links the DMSO sequencing data to the corresponding peaks in the NMIA (+ve) and DMSO (-ve) signals. **5** Reactivity estimation is determined by Gaussian integration; the area of each peak is correlated with the primer termination probability for the corresponding nucleotide in the sequence. Scaling determines the magnitude of the scaling parameter, which accounts for the spread of the distribution and establishes units of measurement.

Figure 7.3 continued. Normalisation computes the probability of adduct formation at each nucleotide and is normalised so that the average intensity for highly reactive nucleotides is 1, and 0 indicates no reactivity. Data was capped so that reactivities >3 were given a score of 3 and nucleotides with no reactivity were given a score of 0. **6** RNA structural models are generated using the Vienna RNAprobing prediction algorithm, where the prediction is constrained by the SHAPE reactivities. A dot-bracket file is generated, representing the predicted base-pairing. **7** Dot-bracket files are used to create RNA structure maps with VARNA, and SHAPE reactivity values are overlaid for visualisation of SHAPE reactivity at individual nucleotides.

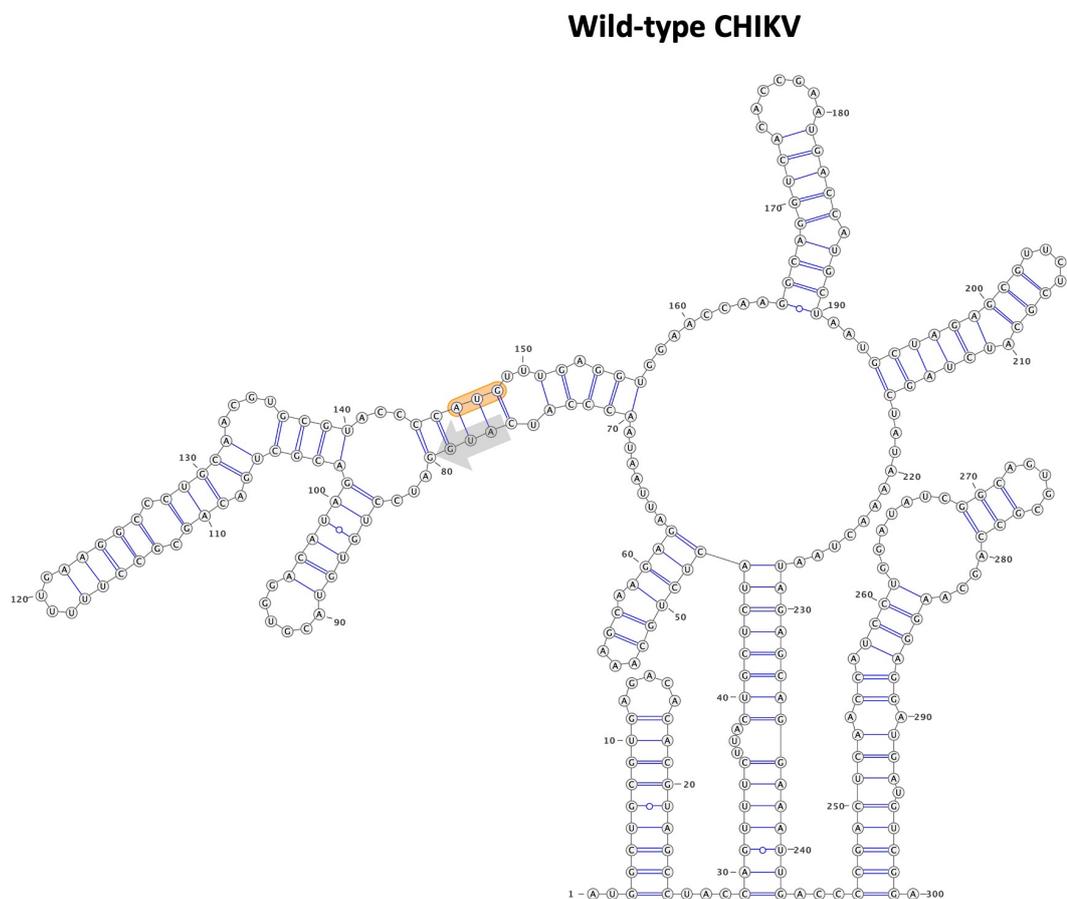


Figure 7.4: *In-silico* minimum free energy RNA secondary structure prediction of wild-type CHIKV

An *in-silico* generated minimum free energy model for wild-type CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

M²⁴A CHIKV

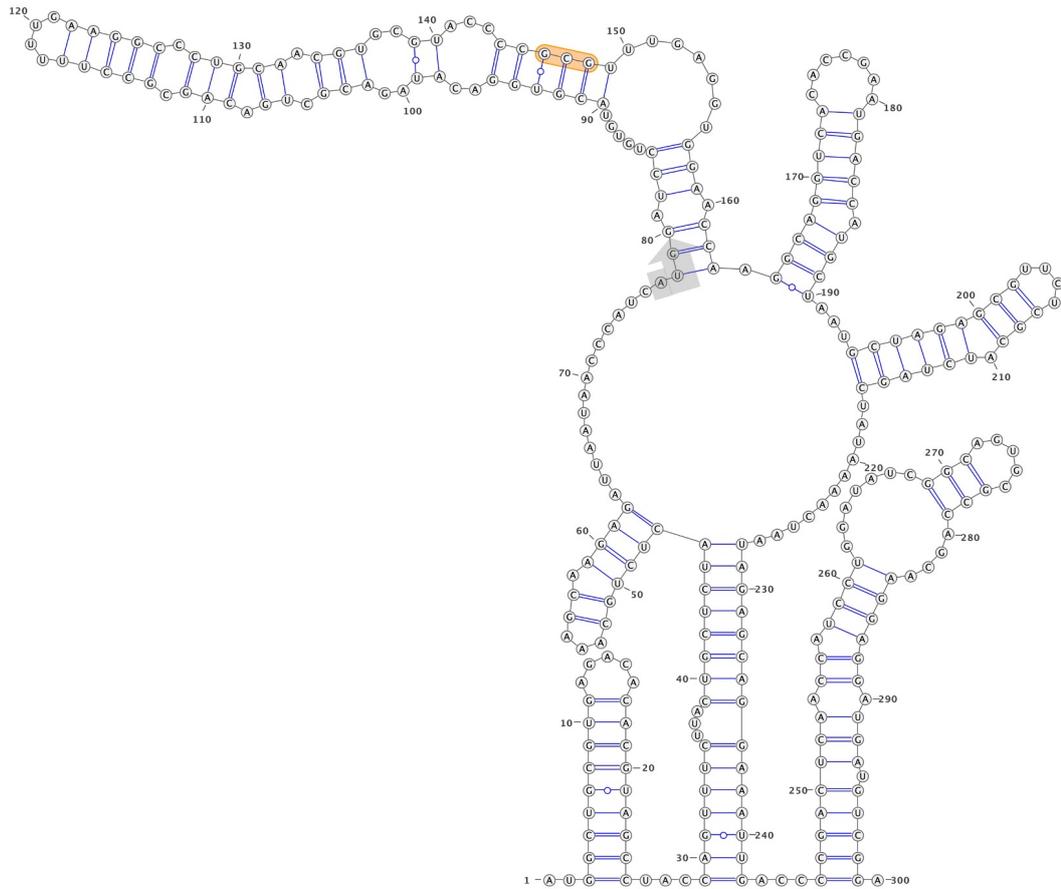


Figure 7.5: *In-silico* minimum free energy RNA secondary structure prediction of M²⁴A CHIKV

An *in-silico* generated minimum free energy model for M²⁴A CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

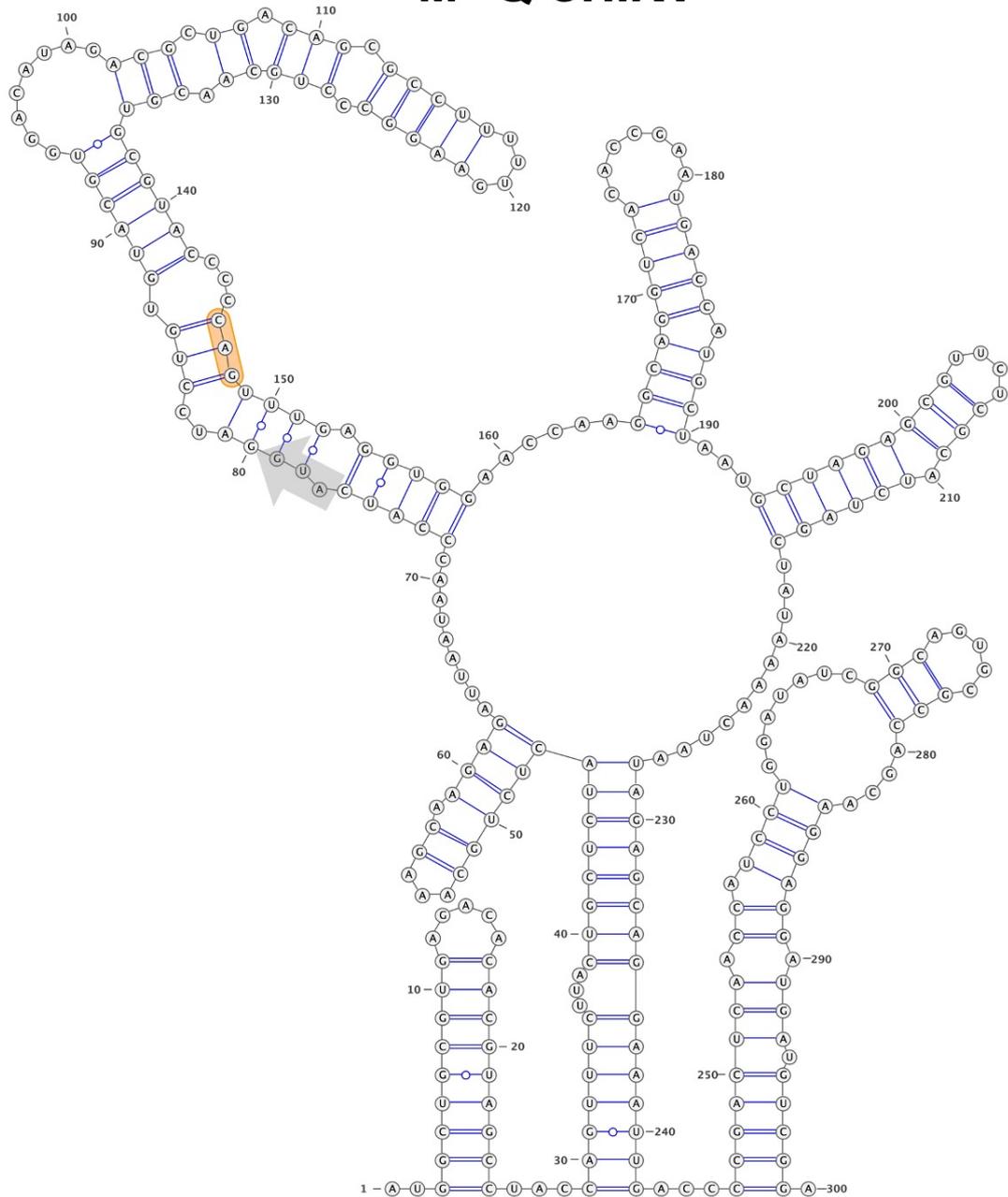
M²⁴Q CHIKV

Figure 7.6: *In-silico* minimum free energy RNA secondary structure prediction of M²⁴Q CHIKV

An *in-silico* generated minimum free energy model for M²⁴Q CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

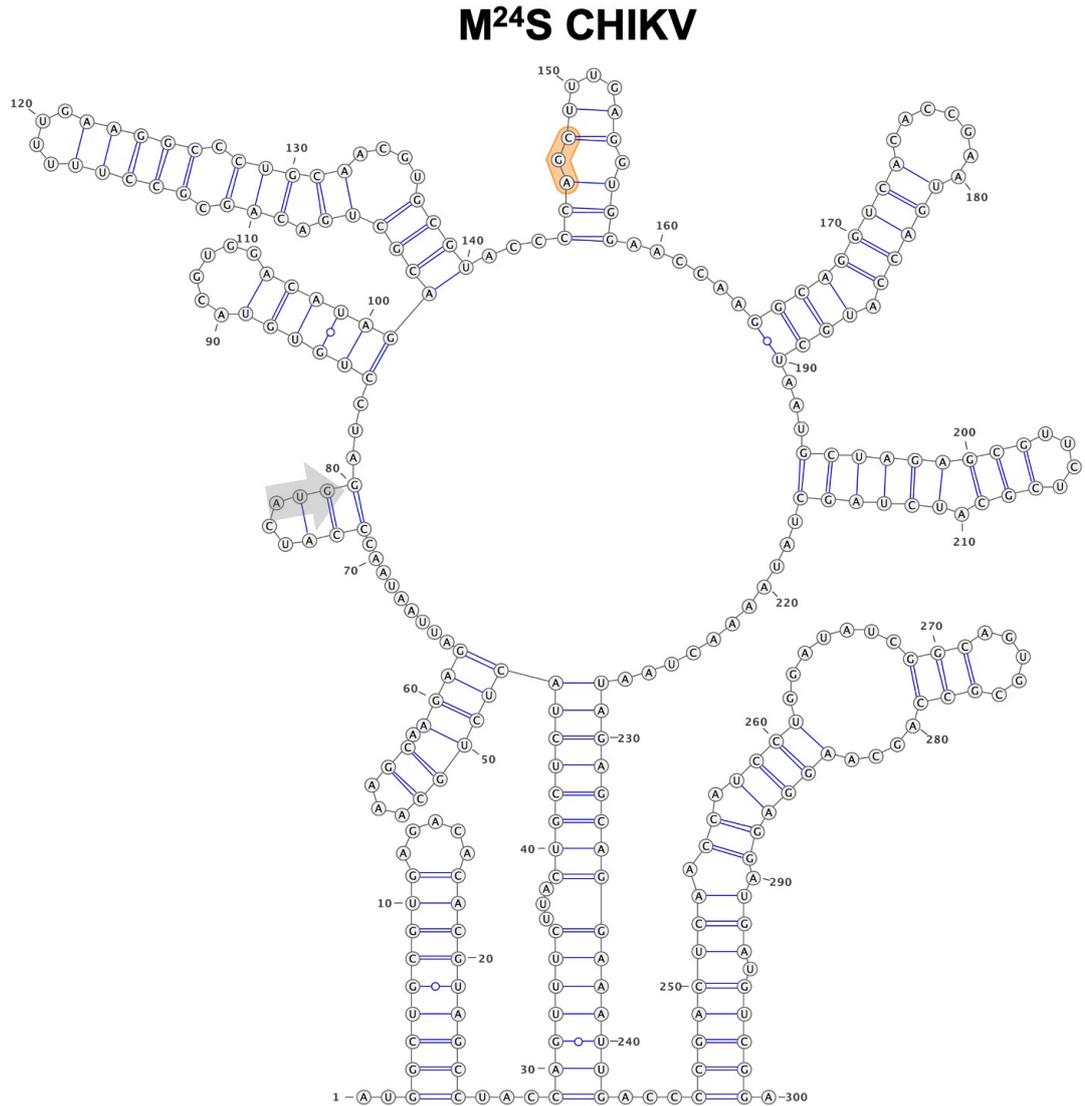


Figure 7.7: *In-silico* minimum free energy RNA secondary structure prediction of M²⁴S CHIKV

An *in-silico* generated minimum free energy model for M²⁴S CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

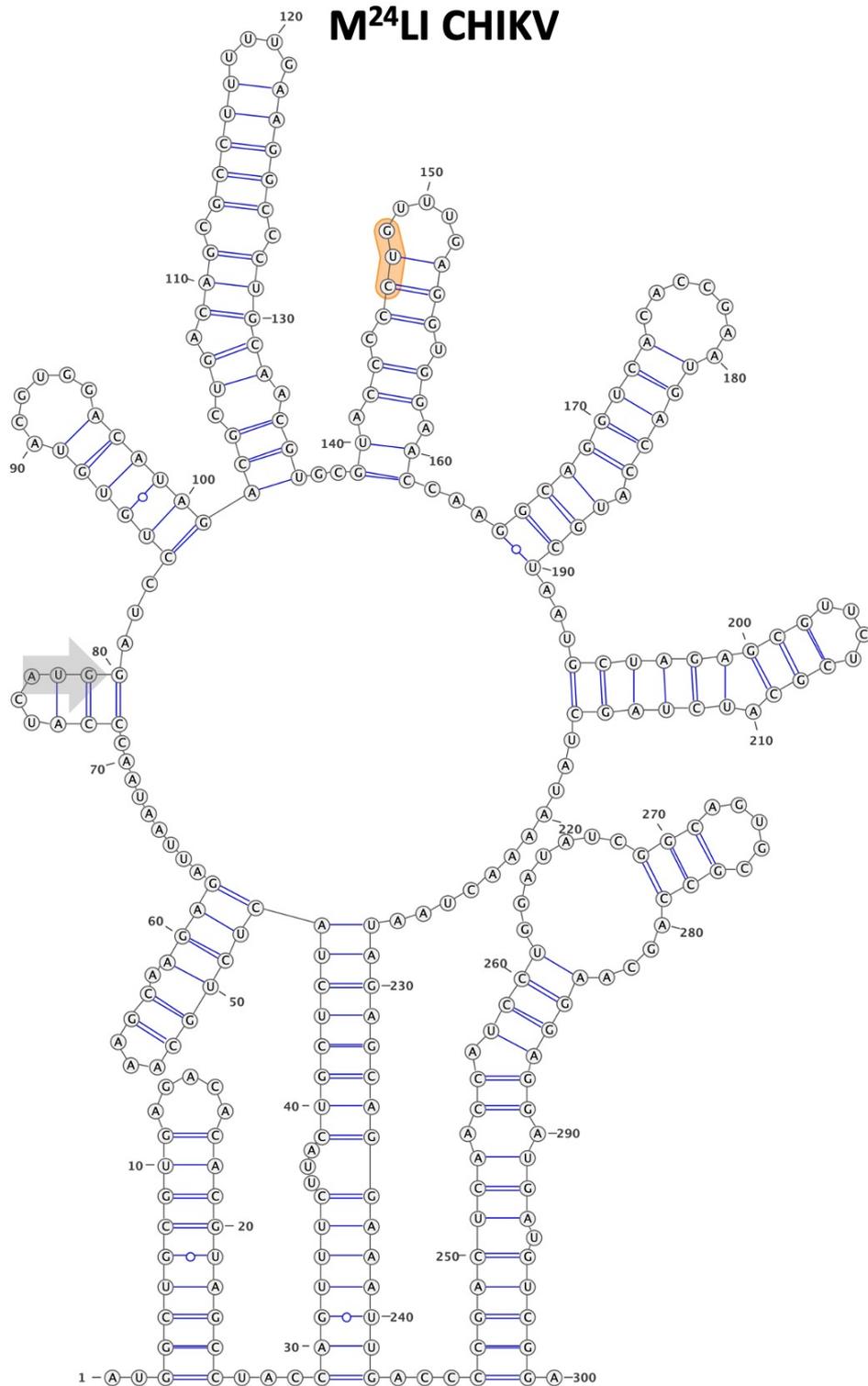


Figure 7.8: *In-silico* minimum free energy RNA secondary structure prediction of M²⁴LI CHIKV

An *in-silico* generated minimum free energy model for M²⁴LI CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

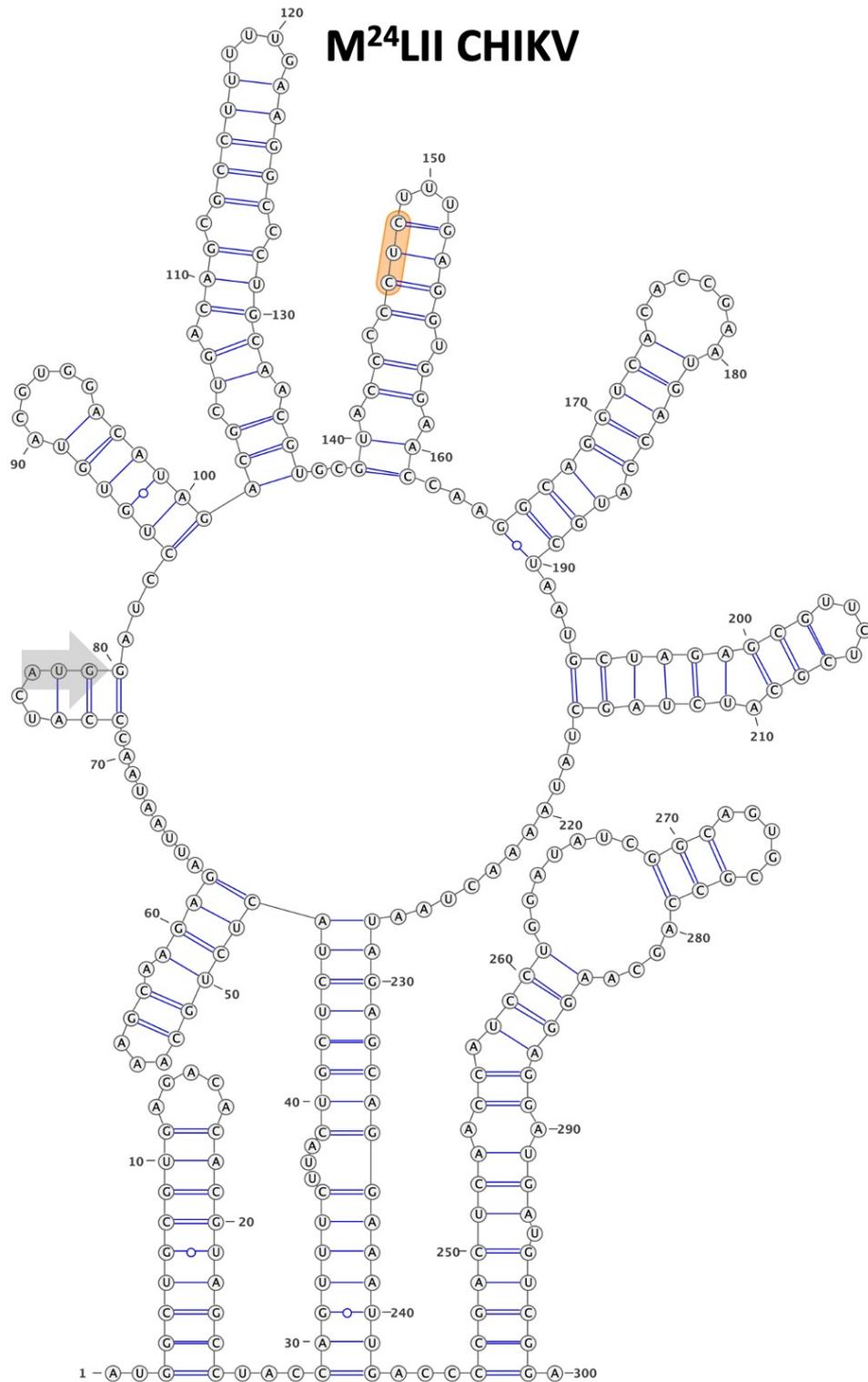


Figure 7.9: *In-silico* minimum free energy RNA secondary structure prediction of M²⁴LII CHIKV

An *in-silico* generated minimum free energy model for M²⁴LII CHIKV, using the Vienna RNAfold server to predict base pairings. Structure was visualised using Varna. Orange denotes residue M²⁴, grey arrow denotes the AUG start codon for nsP1.

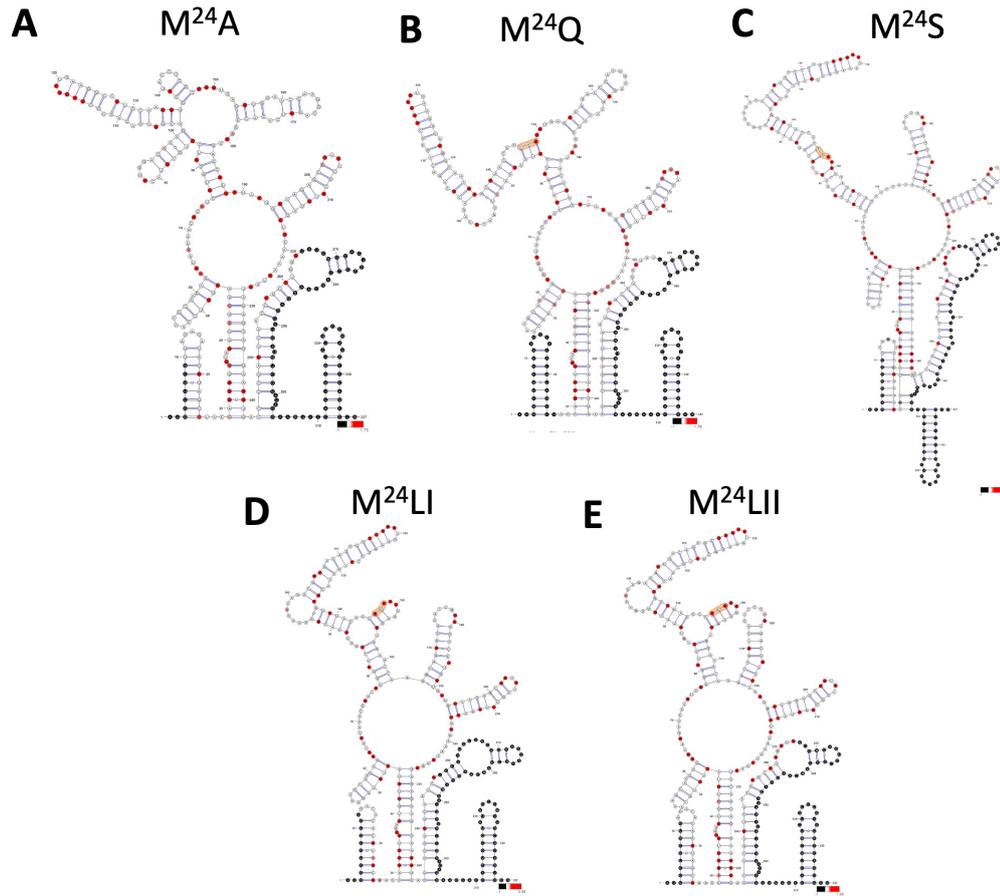


Figure 7.10: SHAPE map of CHIKV M^{24} mutant panel

NMIA reactivity for individual nucleotides, as analysed by SHAPE reactions and capillary electrophoresis. Data was analysed using QuSHAPE, negative values were set to 0 and a maximum of 3 was given to nucleotides that were 'super-reactive'. Predicted structure of the WT CHIKV 5' region when SHAPE reactivity data was used to guide the prediction of RNA secondary structure using the Vienna RNAprobing software. The structure was visualised using VARNA. NMIA reactivities are overlaid and represented on a colour scale of white (0 - 0.3), a gradient from white to red (0.3 - 0.7) and red (> 0.7). Black denotes nucleotides with no data. **A** $M^{24}A$. **B** $M^{24}Q$. **C** $M^{24}S$. **D** $M^{24}LI$. **E** $M^{24}LII$.

Table 7.1: Mutations detected across the CHIKV genome following passaging of M²⁴ in Huh7 cells

The Huh7 monolayer infected with M²⁴A during passaging was harvested with TRI-reagent and cellular RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P8 - P10 is shown. The percentage of the population containing the denoted nucleotide is presented in the table for each passage, alongside the dominant residue encoded at the corresponding amino acid residue. Read depth for the nucleotide is stated.

| Nucleotide | 146-148 | nsP1 24 | | | | |
|------------|---------|---------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | GCG | GUG | C - 87% | U - 13% | A | 223x |
| Passage 9 | GCG | GUG | C - 54% | U - 43% | A | 329x |
| Passage 10 | GCG | GUG | C - 26% | U - 71% | V | 374x |

| Nucleotide | 425 | nsP1 117 | | | | |
|------------|-----|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | GGG | AGG | G - 48% | A - 50% | R | 289x |
| Passage 9 | GGG | AGG | G - 36% | A - 63% | R | 299x |
| Passage 10 | GGG | AGG | G - 47% | A - 51% | R | 425x |

| Nucleotide | 764 | nsP1 230 | | | | |
|------------|-----|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | GGG | AGG | G - 11% | A - 89% | R | 205x |
| Passage 9 | GGG | AGG | G - 13% | A - 87% | R | 165x |
| Passage 10 | GGG | AGG | G - 8% | A - 92% | R | 88X |

| Nucleotide | 3130 | nsP2 483 | | | | |
|------------|------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | UUG | UUA | G - 70% | A - 30% | L | 204x |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 4382 | nsP3 103 - AUD domain | | | | |
|------------|------|-----------------------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | AGU | GGU | A - 75% | G - 25% | S | 393x |
| Passage 9 | AGU | GGU | A - 75% | G - 24% | S | 388x |
| Passage 10 | - | - | - | - | - | - |

CHAPTER 7. APPENDIX

| Nucleotide | 4424 | nsP3 117 - AUD domain | | | | |
|------------|------|-----------------------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | GGG | AGG | G - 15% | A - 85% | R | 386x |
| Passage 9 | GGG | AGG | G - 12% | A - 87% | R | 380x |
| Passage 10 | GGG | AGG | G - 7% | A - 93% | R | 382x |

| Nucleotide | 5351 | nsP3 426 | | | | |
|------------|------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | AGC | GGC | A - 70% | G - 30% | S | 565x |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 7695 | ORF-2 | | | | |
|------------|------|-------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | CAG | CAA | G - 73% | A - 26% | Q | 370x |
| Passage 10 | CAG | CAA | G - 73% | A - 26% | Q | 384x |

| Nucleotide | 9376 | ORF-2 | | | | |
|------------|------|-------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | GGG | AGG | G - 67% | A - 30% | G | 172x |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 11368 | 3' UTR | | | | |
|------------|-------|--------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 8 | G | A | G - 64% | A - 33% | n/a | 240x |
| Passage 9 | G | A | G - 69% | A - 29% | n/a | 178x |
| Passage 10 | G | A | G - 75% | A - 23% | n/a | 195x |

Table 7.2: Mutations detected across the CHIKV genome following passaging of M²⁴ alternately in Huh7 and C6/36 cells

The Huh7 and C6/36 monolayers infected with M²⁴A during passaging was harvested with TRI-reagent and cellular RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P1 - P10 is shown. The percentage of the population containing the denoted nucleotide is presented in the table for each passage, alongside the dominant residue encoded at the corresponding amino acid residue. Read depth for the nucleotide is stated.

| Nucleotide | 147 | | nsP1 24 | | | |
|------------|-----|-----|----------------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | GCG | GUG | C - 50% | U - 49% | A | 304x |
| Passage 3 | GCG | GUG | C - 5% | U - 95% | V | 361x |
| Passage 4 | GCG | GUG | C - 5% | U - 92% | V | 372x |
| Passage 5 | GCG | GUG | C - 9% | U - 87% | V | 337x |
| Passage 6 | GCG | GUG | C - 2% | U - 95% | V | 317x |
| Passage 7 | GCG | GUG | C - 17% | U - 77% | V | 146x |
| Passage 8 | GCG | GUG | C - 9% | U - 82% | V | 57x |
| Passage 9 | GCG | GUG | 6% = A, 2% = G | U - 90% | V | 125x |
| Passage 10 | GCG | GUG | 1% = A, 1% = G | U - 98% | V | 342x |

| Nucleotide | 2276-2278 | | nsP2 199 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | CUA | CCA | U - 71% | C - 28% | L | 258x |
| Passage 6 | CUA | CCA | U - 63% | C - 35% | L | 260x |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | CUA | CCA | U - 59% | C - 40% | L | 81x |

CHAPTER 7. APPENDIX

| Nucleotide | 3128-3130 | | nsP2 483 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | UUG | UUA | G - 71% | A - 29% | L | 218x |
| Passage 3 | UUG | UUA | G - 74% | A - 26% | L | 237x |
| Passage 4 | UUG | UUA | G - 76% | A - 24% | L | 229x |
| Passage 5 | UUG | UUA | G - 80% | A - 20% | L | 93x |
| Passage 6 | - | - | - | - | - | - |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 3665-3667 | | nsP2 662 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | GCG | GCA | G - 65% | A - 33% | A | 162x |
| Passage 5 | - | - | - | - | - | - |
| Passage 6 | - | - | - | - | - | - |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 4382-4384 | | nsP3 103 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | AGU | GGU | A - 76% | G - 24% | S | 388x |
| Passage 3 | AGU | GGU | A - 75% | G - 24% | S | 389x |
| Passage 4 | AGU | GGU | A - 78% | G - 21% | S | 393x |
| Passage 5 | AGU | GGU | A - 78% | G - 21% | S | 398x |
| Passage 6 | AGU | GGU | A - 90% | G - 9% | S | 225x |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | - | - | - | - | - | - |
| Passage 10 | - | - | - | - | - | - |

CHAPTER 7. APPENDIX

| Nucleotide | 5351-5353 | nsP3 426 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | AGC | GGC | A - 68% | G - 31% | S | 182x |
| Passage 3 | AGC | GGC | A - 71% | G - 29% | S | 255x |
| Passage 4 | AGC | GGC | A - 76% | G - 24% | S | 242x |
| Passage 5 | AGC | GGC | A - 70% | G - 30% | S | 176x |
| Passage 6 | AGC | GGC | A - 73% | G - 26% | S | 134x |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | AGC | GGC | A - 70% | G - 30% | S | 180x |
| Passage 10 | AGC | GGC | A - 74% | G - 25% | S | 171x |

| Nucleotide | 9089 | | | | | |
|------------|------|-----|---------|---------|-------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | - | - | - | - | - | - |
| Passage 6 | - | - | - | - | - | - |
| Passage 7 | - | - | - | - | - | - |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | CAA | CAG | A - 38% | G - 61% | Q (Silent mutant) | 333x |
| Passage 10 | CAA | CAG | A - 42% | G - 57% | Q (Silent mutant) | 336x |

| Nucleotide | 10774-10776 | 1070 ORF-2 | | | | |
|------------|-------------|------------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | - | - | - | - | - | - |
| Passage 6 | AUA | ACA | U - 73% | C - 27% | I | 297x |
| Passage 7 | AUA | ACA | U - 69% | C - 29% | I | 35x |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | AUA | ACA | U - 70% | C - 30% | I | 53x |
| Passage 10 | - | - | - | - | - | - |

| Nucleotide | 11368 | | 3' UTR | | | |
|------------|-------|----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | G | A | G - 68% | A - 29% | - | 241x |
| Passage 3 | G | A | G - 64% | A - 34% | - | 235x |
| Passage 4 | G | A | G - 65% | A - 33% | - | 241x |
| Passage 5 | G | A | G - 64% | A - 32% | - | 219x |
| Passage 6 | G | A | G - 64% | A - 33% | - | 206x |
| Passage 7 | G | A | G - 82% | A - 17% | - | 138x |
| Passage 8 | - | - | - | - | - | - |
| Passage 9 | G | A | G - 62% | A - 34% | - | 112x |
| Passage 10 | G | A | G - 60% | A - 38% | - | 239x |

Table 7.3: Mutations detected across the CHIKV genome following passaging of M²⁴ in C6/36 cells

The C6/36 monolayer infected with M²⁴A during passaging was harvested with TRI-reagent and cellular RNA was extracted and purified for analysis of mutations across the CHIKV genome by NGS utilising the Oxford Nanopore Technology platform. Data generated from P1 - P10 is shown. The percentage of the population containing the denoted nucleotide is presented in the table for each passage, alongside the dominant residue encoded at the corresponding amino acid residue. Read depth for the nucleotide is stated.

| Nucleotide | 77-79 | | nsP1 1 | | | |
|------------|-------|-----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | AUG | GUG | A - 99% | G - 1% | M | 396x |
| Passage 2 | AUG | GUG | A - 86% | G - 13% | M | 386x |
| Passage 3 | AUG | GUG | A - 39% | G - 60% | V | 387x |
| Passage 4 | AUG | GUG | A - 19% | G - 81% | V | 376x |
| Passage 5 | AUG | GUG | A - 21% | G - 79% | V | 381x |
| Passage 6 | AUG | GUG | A - 23% | G - 77% | V | 383x |
| Passage 7 | AUG | GUG | A - 27% | G - 73% | V | 377x |

| Nucleotide | 146-148 | | nsP1 24 | | | |
|------------|---------|-----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | GCG | GUG | C - 87% | U - 13% | A | 274x |
| Passage 2 | GCG | GUG | C - 89% | U - 9% | A | 282x |
| Passage 3 | GCG | GUG | C - 89% | U - 10% | A | 283x |
| Passage 4 | GCG | GUG | C - 87% | U - 12% | A | 271x |
| Passage 5 | GCG | GUG | C - 79% | U - 20% | A | 281x |
| Passage 6 | GCG | GUG | C - 83% | U - 16% | A | 277x |
| Passage 7 | GCG | GUG | C - 88% | U - 10% | A | 233x |

CHAPTER 7. APPENDIX

| Nucleotide | 1547-1549 | nsP1 491 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | CGG | CAG | G - 66% | A - 28% | R | 178x |
| Passage 5 | CGG | CAG | G - 71% | A - 29% | R | 49x |
| Passage 6 | CGG | CAG | G - 74% | A - 22% | R | 245x |
| Passage 7 | CGG | CAG | G - 78% | A - 20% | R | 54x |

| Nucleotide | 2276-2278 | nsP2 199 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | CUA | CCA | U - 61% | C - 38% | L | 135X |
| Passage 2 | CUA | CCA | U - 70% | C - 30% | L | 30x |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | CUA | CCA | U - 76% | C - 23% | L | 388x |
| Passage 5 | CUA | CCA | U - 74% | C - 25% | L | 387x |
| Passage 6 | CUA | CCA | U - 73% | C - 25% | L | 391x |
| Passage 7 | CUA | CCA | U - 63% | C - 34% | L | 82x |

| Nucleotide | 2810-2812 | nsP2 377 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | - | - | - | - | - | - |
| Passage 6 | AUG | ACG | U - 76% | C - 22% | M | 366x |
| Passage 7 | AUG | ACG | U - 80% | C - 18% | M | 333x |

| Nucleotide | 3128-3130 | nsP2 483 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | UUG | UUA | G - 68% | A - 31% | L | 215x |
| Passage 4 | UUG | UUA | G - 72% | A - 25% | L | 217x |
| Passage 5 | UUG | UUA | G - 77% | A - 23% | L | 213x |
| Passage 6 | UUG | UUA | G - 75% | A - 24% | L | 218x |
| Passage 7 | UUG | UUA | G - 86% | A - 12% | L | 97x |

CHAPTER 7. APPENDIX

| Nucleotide | 4328-4330 | | nsP3 85 | | | |
|------------|-----------|-----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | UUG | CUG | U - 74% | C - 19% | L | 128x |
| Passage 4 | UUG | CUG | U - 75% | C - 9% | L | 528x |
| | UUG | GUG | U - 75% | G - 15% | L | |
| Passage 5 | UUG | CUG | U - 70% | C - 11% | L | 509x |
| | UUG | GUG | U - 70% | G - 17% | L | |
| Passage 6 | UUG | CUG | U - 71% | C - 11% | L | 523x |
| | UUG | GUG | U - 71% | G - 18% | L | |
| Passage 7 | UUG | CUG | U - 75% | C - 10% | L | 522x |
| | UUG | GUG | U - 75% | G - 15% | L | |

| Nucleotide | 4334-4336 | | nsP3 87 | | | |
|------------|-----------|-----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | GCU | GCC | U - 71% | C - 28% | A | 103x |
| Passage 3 | GCU | GCC | U - 84% | C - 15% | A | 124x |
| Passage 4 | GCU | GCC | U - 86% | C - 13% | A | 385x |
| Passage 5 | GCU | GCC | U - 85% | C - 15% | A | 391x |
| Passage 6 | GCU | GCC | U - 86% | C - 16% | A | 391x |
| Passage 7 | GCU | GCC | U - 85% | C - 14% | A | 386x |

| Nucleotide | 4382-4384 | | nsP3 103 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | AGU | GGU | A - 74% | G - 26% | S | 390x |
| Passage 5 | AGU | GGU | A - 77% | G - 23% | S | 393x |
| Passage 6 | AGU | GGU | A - 80% | G - 20% | S | 396x |
| Passage 7 | AGU | GGU | A - 75% | G - 25% | S | 390x |

| Nucleotide | 5351-5353 | | nsP3 426 | | | |
|------------|-----------|-----|----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | AGC | GGC | A - 65% | G - 35% | S | 164x |
| Passage 2 | AGC | GGC | A - 75% | G - 23% | S | 57x |
| Passage 3 | AGC | GGC | A - 58% | G - 42% | S | 86x |
| Passage 4 | AGC | GGC | A - 75% | G - 24% | S | 566x |
| Passage 5 | AGC | GGC | A - 79% | G - 20% | S | 570x |
| Passage 6 | AGC | GGC | A - 79% | G - 20% | S | 564x |
| Passage 7 | AGC | GGC | A - 79% | G - 21% | S | 563x |

CHAPTER 7. APPENDIX

| Nucleotide | 6326-6328 | nsP4 221 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | GCA | GCG | A - 72% | G - 28% | A | 208x |
| Passage 5 | GCA | GCG | A - 69% | G - 31% | A | 242x |
| Passage 6 | GCA | GCG | A - 75% | G - 25% | A | 226x |
| Passage 7 | GCA | GCG | A - 68% | G - 32% | A | 222x |

| Nucleotide | 6737-6739 | nsP4 | | | | |
|------------|-----------|------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | - | - | - | - | - | - |
| Passage 6 | GCC | GCU | C - 75% | U - 23% | A | 389x |
| Passage 7 | GCC | GCU | C - 76% | U - 22% | A | 386x |

| Nucleotide | 7693-7695 | ORF-2 43 | | | | |
|------------|-----------|----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | CAG | CAA | G - 73% | A - 26% | Q | 373x |
| Passage 5 | CAG | CAA | G - 73% | A - 26% | Q | 370x |
| Passage 6 | CAG | CAA | G - 68% | A - 29% | Q | 371x |
| Passage 7 | CAG | CAA | G - 82% | A - 17% | Q | 239x |

| Nucleotide | 9064-9066 | ORF-2 500 | | | | |
|------------|-----------|-----------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | ACC | GCC | A - 67% | G - 32% | T | 356x |
| Passage 6 | ACC | GCC | A - 53% | G - 41% | T | 361x |
| Passage 7 | ACC | GCC | A - 40% | G - 54% | A | 347x |

CHAPTER 7. APPENDIX

| Nucleotide | 9376-9378 | | ORF-2 604 | | | |
|------------|-----------|-----|-----------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | GGG | AGG | G - 70% | A - 26% | G | 223x |
| Passage 5 | GGG | AGG | G - 64% | A - 33% | G | 221x |
| Passage 6 | GGG | AGG | G - 69% | A - 28% | G | 226x |
| Passage 7 | GGG | AGG | G - 67% | A - 30% | G | 230x |

| Nucleotide | 10669-10671 | | ORF-2 1035 | | | |
|------------|-------------|-----|------------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | GUG | GCG | U - 95% | C - 4% | V | 388x |
| Passage 2 | GUG | GCG | U - 92% | C - 4% | V | 226x |
| Passage 3 | GUG | GCG | U - 86% | C - 11% | V | 109x |
| Passage 4 | GUG | GCG | U - 63% | C - 36% | V | 72x |
| Passage 5 | GUG | GCG | U - 52% | C - 47% | V | 91x |
| Passage 6 | GUG | GCG | U - 41% | C - 57% | A | 173x |
| Passage 7 | GUG | GCG | U - 24% | C - 71% | A | 376x |

| Nucleotide | 11215-11217 | | ORF-2 1217 | | | |
|------------|-------------|-----|------------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | UCA | UCG | A - 71% | G - 13% | S | 239x |
| | UCA | UCU | A - 71% | U - 12% | S | |
| | UCA | UCC | A - 71% | C - 4% | S | |
| Passage 5 | UCA | UCG | A - 75% | G - 10% | S | 163x |
| | UCA | UCU | A - 75% | U - 11% | S | |
| | UCA | UCC | A - 75% | C - 4% | S | |
| Passage 6 | UCA | UCG | A - 77% | G - 9% | S | 248x |
| | UCA | UCU | A - 77% | U - 10% | S | |
| | UCA | UCC | A - 77% | C - 4% | S | |
| Passage 7 | - | - | - | - | - | - |

| Nucleotide | 11368 | | 3' UTR | | | |
|------------|-------|----|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | G | A | G - 69% | A - 29% | - | 231x |
| Passage 2 | G | A | G - 65% | A - 32% | - | 214x |
| Passage 3 | G | A | G - 58% | A - 41% | - | 198x |
| Passage 4 | G | A | G - 56% | A - 42% | - | 170x |
| Passage 5 | G | A | G - 65% | A - 34% | - | 144x |
| Passage 6 | G | A | G - 63% | A - 34% | - | 104x |
| Passage 7 | G | A | G - 63% | A - 34% | - | 148x |

CHAPTER 7. APPENDIX

| Nucleotide | 11396 | 3' UTR | | | | |
|------------|-------|--------|---------|---------|------------------|------------|
| | RO | AO | RO % | AO% | Dominant Residue | Read Depth |
| Passage 1 | - | - | - | - | - | - |
| Passage 2 | - | - | - | - | - | - |
| Passage 3 | - | - | - | - | - | - |
| Passage 4 | - | - | - | - | - | - |
| Passage 5 | G | A | G - 82% | A - 16% | - | 134x |
| Passage 6 | - | - | - | - | - | - |
| Passage 7 | - | - | - | - | - | - |

References

- (1) Nowee, G., Bakker, J. W., Geertsema, C., Ros, V. I., Göertz, G. P., Fros, J. J., and Pijlman, G. P. (2021). A Tale of 20 Alphaviruses; Inter-species Diversity and Conserved Interactions Between Viral Non-structural Protein 3 and Stress Granule Proteins. *Frontiers in Cell and Developmental Biology* 9, 125.
- (2) Phane Villoing, S., Bé Arzotti, M., Chilmonczyk, S., Castric, J., and Bré Mont, M. Rainbow Trout Sleeping Disease Virus Is an Atypical Alphavirus, tech. rep. 1, 2000, pp 173–183.
- (3) Linn, M. L., Gardner, J., Warrilow, D., Darnell, G. A., McMahon, C. R., Field, I., Hyatt, A. D., Slade, R. W., and Suhrbier, A. (2001). Arbovirus of Marine Mammals: a New Alphavirus Isolated from the Elephant Seal Louse, *Lepidophthirus macrorhini*. *Journal of Virology* 75, 4103.
- (4) Nasar, F., Palacios, G., Gorchakov, R. V., Guzman, H., Travassos Da Rosa, A. P., Savji, N., Popov, V. L., Sherman, M. B., Lipkin, W. I., Tesh, R. B., and Weaver, S. C. (2012). Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. *Proceedings of the National Academy of Sciences of the United States of America* 109, 14622–14627.
- (5) Altinli, M., Leggewie, M., Badusche, M., Gyanwali, R., Scherer, C., Schulze, J., Sreenu, V. B., Fegebank, M., Zibrat, B., Fuss, J., Junglen, S., and Schnettler, E. (2022). Antiviral RNAi Response against the Insect-Specific Agua Salud Alphavirus. *mSphere* 7.
- (6) Weaver, S. C., Kang, W., Shirako, Y., Tillman Ru"menapf, †., Ru"menapf, R., Strauss, E. G., and Strauss, J. H. (1997). Recombinational History and Molecular Evolution of Western Equine Encephalomyelitis Complex Alphaviruses. *JOURNAL OF VIROLOGY* 71, 613–623.

REFERENCES

- (7) COLVILLE, J. L., and BERRYHILL, D. L. (2007). EASTERN EQUINE ENCEPHALITIS. *Handbook of Zoonoses*, 73–76.
- (8) Banda, C., and Samanta, D., *Eastern Equine Encephalitis*; StatPearls Publishing: 2022.
- (9) Morens, D. M., Folkers, G. K., and Fauci, A. S. (2019). Eastern Equine Encephalitis Virus - Another Emergent Alphavirus in the United States. *New England Journal of Medicine* 381, 1985–1992.
- (10) Tesh, R. B., Watts, D. M., Russell, K. L., Damodaran, C., Calampa, C., Cabezas, C., Ramirez, G., Vasquez, B., Hayes, C. G., Rossi, C. A., Powers, A. M., Hice, C. L., Chandler, L. J., Cropp, B. C., Karabatsos, N., Roehrig, J. T., and Gubler, D. J. (1999). Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 28, 67–73.
- (11) Rezza, G., Chen, R., and Weaver, S. C. (2017). O’nyong-nyong fever: a neglected mosquito-borne viral disease. *Pathogens and Global Health* 111, 271–275.
- (12) Powers, A. M., Brault, A. C., Shirako, Y., Strauss, E. G., Kang, W., Strauss, J. H., and Weaver, S. C. (2001). Evolutionary Relationships and Systematics of the Alphaviruses. *Journal of Virology* 75, 10118.
- (13) Robinson, M. C. (1955). An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 49, 28–32.
- (14) Jain, M., Rai, S., and Chakravarti, A. (2008). Chikungunya: A review. *Tropical Doctor* 38, 70–72.
- (15) Caglioti, C., Lalle, E., Castilletti, C., Carletti, F., Capobianchi, M. R., and Bordi, L. Chikungunya virus infection: An overview, 2013.
- (16) PIALOUX, G., GAÜZÈRE, B.-A., JAURÉGUIBERRY, S., and STROBEL, M. (2007). Chikungunya, an epidemic arbovirosis. *The Lancet Infectious Diseases* 7, 319–327.

REFERENCES

- (17) Josseran, L., Paquet, C., Zehgnoun, A., Caillere, N., Le Tertre, A., Solet, J. L., and Ledrans, M. (2006). Chikungunya Disease Outbreak, Reunion Island. *Emerging Infectious Diseases* 12, 1994.
- (18) Mavalankar, D., Shastri, P., Bandyopadhyay, T., Parmar, J., and Ramani, K. V. (2008). Increased Mortality Rate Associated with Chikungunya Epidemic, Ahmedabad, India. *Emerging Infectious Diseases* 14, 412–415.
- (19) Yergolkar, P. N., Tandale, B. V., Arankalle, V. A., Sathe, P. S., B, S. A., Gandhe, S. S., Gokhle, M. D., Jacob, G. P., Hundekar, S. L., and Mishra, A. C. Chikungunya Outbreaks Caused by African Genotype, India, tech. rep., 2006.
- (20) Wimalasiri-Yapa, B. M., Stassen, L., Huang, X., Hafner, L. M., Hu, W., Devine, G. J., Yakob, L., Jansen, C. C., Faddy, H. M., Viennet, E., and Frentiu, F. D. (2019). Chikungunya virus in Asia – Pacific: a systematic review. *Emerging Microbes & Infections* 8, 70.
- (21) Liu, L. B., Li, M., Gao, N., Shen, J. Y., Sheng, Z. Y., Fan, D. Y., Zhou, H. N., Yin, X. X., Mao, J. R., Jiang, J. Y., Wang, P. G., and An, J. (2022). Epidemiological and clinical characteristics of the chikungunya outbreak in Ruili City, Yunnan Province, China. *Journal of Medical Virology* 94, 499–506.
- (22) de Lamballerie, X., Leroy, E., Charrel Rémi N., Tsetsarkin Konstantin, Higgs Stephen, and Gould Ernest A (2008). Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? *Virology Journal* 5, 33.
- (23) Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E., and Higgs, S. (2007). A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. *PLoS Pathog* 12, DOI: 10.1371/journal.ppat.0030201.
- (24) Rezza, G., Nicoletti, L., Angelini, R., Romi, R., Finarelli, A., Panning, M., Cordioli, P., Fortuna, C., Boros, S., Magurano, F., Silvi, G., Angelini, P., Dottori, M., Ciufolini, M., Majori, G., and Cassone, A. (2007). Infection with chikungunya virus in Italy: an outbreak in a temperate region. *The Lancet* 370, 1840–1846.

REFERENCES

- (25) Grandadam, M., Caro, V., Plumet, S., Thiberge, J. M., Souarès, Y., Failloux, A. B., Tolou, H. J., Budelot, M., Cosserat, D., Leparc-Goffart, I., and Desprès, P. (2011). Chikungunya virus, Southeastern France. *Emerging Infectious Diseases* 17, 910–913.
- (26) Silva, L. A., and Dermondy T S (2017). The Journal of Clinical Investigation Ecology and epidemiology. *The Journal of Clinical Investigation* 127, DOI: 10.1172/JCI84417.
- (27) An, W., Ge, N., Cao, Y., Sun, J., and Kin, X. (2017). Recent progress on chikungunya virus research. *Virologica Sinica* 32, 441–453.
- (28) Cassadou, S., Boucau, S., Petit-Sinturel, M., Huc, P., Leparc-Goffart, I., and Ledrans, M. (2014). Emergence of chikungunya fever on the French side of Saint Martin island, October to December 2013. *Eurosurveillance* 19, 20752.
- (29) Mowatt, L., and Jackson, S. T. (2014). Chikungunya in the Caribbean: An Epidemic in the Making. *Infectious Diseases and Therapy* 3, 63–68.
- (30) Wahid, B., Ali, A., Rafique, S., and Idrees, M. (2017). Global expansion of chikungunya virus: mapping the 64-year history. DOI: 10.1016/j.ijid.2017.03.006.
- (31) CDC Geographic Distribution — Chikungunya virus — CDC, 2018.
- (32) Weaver, S. C., Winegar, R., Manger, I. D., and Forrester, N. L. (2012). Alphaviruses: Population genetics and determinants of emergence. *Antiviral Research* 94, 242–257.
- (33) Schuffenecker, I., Itean, I., Michault, A., Verine Murri, S., Frangeul, L., Vaney, M.-C., Lavenir, R., Pardigon, N., Reynes, J.-M., Pettinelli, F., Biscornet, L., Diancourt, L., Phanie Michel, S., Phane Duquerroy, S., Guigon, G., Frenkiel, M.-P., Bré hin, A.-C., Ge Cubito, N., Desprès, P., Kunst, F., Lix Rey, F. A., Zeller, H., and Brisse, S. (2006). Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. *PLoS Medicine* — *www* 3, 1058.

REFERENCES

- (34) Diallo, M., Thonnon, J., Traore-Lamizana, M., and Fontenille, D. (1999). Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *The American Journal of Tropical Medicine and Hygiene* 60, 281–286.
- (35) Althouse, B. M., Guerbois, M., Cummings, D. A., Diop, O. M., Faye, O., Faye, A., Diallo, D., Sadio, B. D., Sow, A., Faye, O., Sall, A. A., Diallo, M., Benefit, B., Simons, E., Watts, D. M., Weaver, S. C., and Hanley, K. A. (2018). Role of monkeys in the sylvatic cycle of chikungunya virus in Senegal. *Nature Communications* 9, DOI: 10.1038/s41467-018-03332-7.
- (36) Chevillon, C., Briant, L., Renaud, F., and Devaux, C. (2008). The Chikungunya threat: an ecological and evolutionary perspective. *Trends in Microbiology* 16, 80–88.
- (37) Moreira Soto, A., de Oliveira Carneiro, I., Fischer, C., Feldmann, M., Kümmerer, B. M., Santos Silva, N., Góes Santos, U., Frederico de Carvalho Dominguez Souza, B., de Azevedo Liborio, F., Mafra Valença-Montenegro, M., de Oliveira Laroque, P., Rosa da Fontoura, F., Vinicius Dantas Oliveira, A., Drosten, C., de Lamballerie, X., Roberto Franke, C., and Felix Drexler, J. (2018). Limited Evidence for Infection of Urban and Peri-urban Nonhuman Primates with Zika and Chikungunya Viruses in Brazil. DOI: 10.1128/mSphere.
- (38) Yusof, M. A., Ahmad, N. W., and Vythilingam, I. The first isolation of Chikungunya virus from non-human primates in Malaysia Interactions of Co-infection of *Brugia pahangi* and *Plasmodium berghei* ANKA in Gerbil (*Meriones unguiculatus*) View project Vectors of *Plasmodium knowlesi* View project, tech. rep., 2009.
- (39) Sam, I. C., Chua, C. L., Rovie-Ryan, J. J., Fu, J. Y., Tong, C., Sitam, F. T., and Chan, Y. F. (2015). Chikungunya Virus in Macaques, Malaysia. *Emerging Infectious Diseases* 21, 1683.
- (40) Nakgoi, K., Nitatpattana, N., Wajjwalku, W., Pongsopawijit, P., Kaewchot, S., Yoksan, S., Siripolwat, V., Souris, M., and Gonzalez, J. P. (2014). Dengue, Japanese encephalitis and Chikungunya virus antibody preva-

REFERENCES

- lence among captive monkey (*Macaca nemestrina*) colonies of Northern Thailand. *American Journal of Primatology* 76, 97–102.
- (41) Valentine, M. J., Murdock, C. C., and Kelly, P. J. Sylvatic cycles of arboviruses in non-human primates, 2019.
- (42) Diallo, D., Sall, A. A., Buenemann, M., Chen, R., and Faye, O. (2012). Landscape Ecology of Sylvatic Chikungunya Virus and Mosquito Vectors in Southeastern Senegal. *PLoS Negl Trop Dis* 6, 1649.
- (43) Labadie, K., Larcher, T., Joubert, C., Mannioui, A., Delache, B., Brochard, P., Guigand, L., Dubreil, L., Lebon, P., Verrier, B., De Lamballerie, X., Suhrbier, A., Cherel, Y., Grand, R. L., and Roques, P. (2010). Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. *The Journal of clinical investigation* 120, 894–906.
- (44) Das, T., Jaffar-Bandjee, M. C., Hoarau, J. J., Krejbich Trotot, P., Denizot, M., Lee-Pat-Yuen, G., Sahoo, R., Guiraud, P., Ramful, D., Robin, S., Alessandri, J. L., Gauzere, B. A., and Gasque, P. (2010). Chikungunya fever: CNS infection and pathologies of a re-emerging arbovirus. *Progress in Neurobiology* 91, 121–129.
- (45) Panning, M., Grywna, K., Van Esbroeck, M., Emmerich, P., and Drosten, C. (2008). Chikungunya Fever in Travelers Returning to Europe from the Indian Ocean Region, 2006 - Volume 14, Number 3—March 2008 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases* 14, 416–422.
- (46) Panning, M., Grywna, K., Van Esbroeck, M., Emmerich, P., and Drosten, C. (2008). Chikungunya Fever in Travelers Returning to Europe from the Indian Ocean Region, 2006. *Emerging Infectious Diseases* 14, 416–422.
- (47) Ozden, S., Huerre, M., Riviere, J.-P., Coffey, L. L., Afonso, P. V., Mouly, V., De Monredon, J., Roger, J.-C., Amrani, M. E., Yvin, J.-L., Jaffar, M.-C., Frenkiel, M.-P., Sourisseau, M., Schwartz, O., Butler-Browne, G., Desprè, P., Gessain, A., and Ceccaldi, P.-E. (2007). Human Muscle Satellite Cells as Targets of Chikungunya Virus Infection. *PLoS ONE*, DOI: 10.1371/journal.pone.0000527.

REFERENCES

- (48) Chua, H. H., Abdul Rashid, K., Law, W. C., Hamizah, A., Chem, Y. K., Khairul, A. H., and Chua, K. B. (2010). A Fatal Case of Chikungunya Virus Infection with Liver Involvement. *Med J Malaysia* 65.
- (49) Hawman, D. W., Stoermer, K. A., Montgomery, S. A., Pal, P., Oko, L., Diamond, M. S., and Morrison, T. E. (2013). Chronic Joint Disease Caused by Persistent Chikungunya Virus Infection Is Controlled by the Adaptive Immune Response. *Journal of Virology* 87, 13878–13888.
- (50) Cunha, R. V., and Trinta, K. S. Chikungunya virus: Clinical aspects and treatment, 2017.
- (51) Sam, I. C., Kümmerer, B. M., Chan, Y. F., Roques, P., Drosten, C., and AbuBakar, S. (2015). Updates on Chikungunya Epidemiology, Clinical Disease, and Diagnostics. <https://home.liebertpub.com/vbz> 15, 223–230.
- (52) Simon, F., Javelle, E., Cabie, A., Bouquillard, E., Troisgros, O., Gentile, G., Leparc-Goffart, I., Hoen, B., Gandjbakhch, F., Rene-Corail, P., Franco, J. M., Caumes, E., Combe, B., Poiraudreau, S., Gane-Troplent, F., Djossou, F., Schaerverbeke, T., Criquet-Hayot, A., Carrere, P., Malvy, D., Gaillard, P., and Wendling, D. (2015). French guidelines for the management of chikungunya (acute and persistent presentations). November 2014. *Medecine et Maladies Infectieuses* 45, 243–263.
- (53) Schilte, C., Staikovsky, F., Couderc, T., Madec, Y., Carpentier, F., Kassab, S., Albert, M. L., Lecuit, M., and Michault, A. (2013). Chikungunya Virus-associated Long-term Arthralgia: A 36-month Prospective Longitudinal Study. *PLoS Neglected Tropical Diseases* 7, DOI: 10.1371/journal.pntd.0002137.
- (54) Wong, H. V., Chan, F., Sam, I.-C., Yusof, W., Sulaiman, W., and Vythilingam, I. (2016). Chikungunya Virus Infection of Aedes Mosquitoes. *Methods in Molecular Biology* 1426, DOI: 10.1007/978-1-4939-3618-2{_}11.
- (55) Sanchez-Vargas, I., Travanty, E. A., Keene, K. M., Franz, A. W., Beaty, B. J., Blair, C. D., and Olson, K. E. (2004). RNA interference, arthropod-borne viruses, and mosquitoes. *Virus Research* 102, 65–74.

REFERENCES

- (56) Keene, K. M., Foy, B. D., Sanchez-Vargas, I., Beaty, B. J., Blair, C. D., and Olson, K. E. (2004). RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 101, 17240.
- (57) Forrester, N. L., Coffey, L. L., and Weaver, S. C. (2014). Arboviral Bottlenecks and Challenges to Maintaining Diversity and Fitness during Mosquito Transmission. *Viruses* 6, 3991–4004.
- (58) Chen, G. L., Coates, E. E., Plummer, S. H., Carter, C. A., Berkowitz, N., Conan-Cibotti, M., Cox, J. H., Beck, A., O'Callahan, M., Andrews, C., Gordon, I. J., Larkin, B., Lampley, R., Kaltovich, F., Gall, J., Carlton, K., Mendy, J., Haney, D., May, J., Bray, A., Bailer, R. T., Dowd, K. A., Brockett, B., Gordon, D., Koup, R. A., Schwartz, R., Mascola, J. R., Graham, B. S., Pierson, T. C., Donastorg, Y., Rosario, N., Pape, J. W., Hoen, B., Cabié, A., Diaz, C., and Ledgerwood, J. E. (2020). Effect of a Chikungunya Virus–Like Particle Vaccine on Safety and Tolerability Outcomes: A Randomized Clinical Trial. *JAMA* 323, 1369–1377.
- (59) Valneva Valneva Successfully Completes Pivotal Phase 3 Trial of Single-Shot Chikungunya Vaccine Candidate, 2022.
- (60) Button, J. M., Qazi, S. A., Wang, J. C. Y., and Mukhopadhyay, S. Revisiting an old friend: new findings in alphavirus structure and assembly, 2020.
- (61) Owen, K. E., and Kuhn, R. J. (1997). Alphavirus Budding Is Dependent on the Interaction between the Nucleocapsid and Hydrophobic Amino Acids on the Cytoplasmic Domain of the E2 Envelope Glycoprotein. *Virology* 230, 187–196.
- (62) Zhao, H., Lindqvist, B., Garoff, H., Von Bonsdorff, C. H., and Liljeström, P. (1994). A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *The EMBO Journal* 13, 4204–4211.

REFERENCES

- (63) Zhang, Y.-N., Deng, C.-L., Li, J.-Q., Li, N., Zhang, Q.-Y., Ye, H.-Q., Yuan, Z.-M., and Zhang, B. (2019). Infectious Chikungunya Virus (CHIKV) with a Complete Capsid Deletion: a New Approach for a CHIKV Vaccine. *Journal of Virology* 93, DOI: 10.1128/JVI.00504-19/FORMAT/EPUB.
- (64) Rozanov, M. N., Koonin, E. V., and Gorbalenya, A. E. (1992). Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *The Journal of general virology* 73 (Pt 8), 2129–2134.
- (65) Martin, J. L., and McMillan, F. M. (2002). SAM (dependent) I AM: The S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology* 12, 783–793.
- (66) Ahola, T., Lampio, A., Auvinen, P., Kä, L., and Riä Inen, . Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity, tech. rep. 11, 1999, pp 3164–3172.
- (67) Laakkonen, P., Ahola, T., and Kääriäinen, L. (1996). The effects of palmitoylation on membrane association of Semliki forest virus RNA capping enzyme. *The Journal of biological chemistry* 271, 28567–28571.
- (68) Lampio, A., Kilpeläinen, I., Pesonen, S., Karhi, K., Auvinen, P., Somerharju, P., and Kääriäinen, L. (2000). Membrane binding mechanism of an RNA virus-capping enzyme. *Journal of Biological Chemistry* 275, 37853–37859.
- (69) Spuul, P., Salonen, A., Merits, A., Jokitalo, E., Kääriäinen, L., and Ahola, T. (2007). Role of the amphipathic peptide of Semliki forest virus replicase protein nsP1 in membrane association and virus replication. *Journal of virology* 81, 872–83.
- (70) Birdwell, C. R., Strauss, E. G., and Strauss, J. H. (1973). Replication of Sindbis virus: III. An electron microscopic study of virus maturation using the surface replica technique. *Virology* 56, 429–438.
- (71) Martinez, M. G., Snapp, E.-L., Perumal, G. S., Macaluso, F. P., and Kielian, M. (2014). Imaging the alphavirus exit pathway. *Journal of virology* 88, 6922–6933.

REFERENCES

- (72) Zusinaite, E., Tints, K., Kiiver, K., Spuul, P., Karo-Astover, L., Merits, A., and Sarand, I. (2007). Mutations at the palmitoylation site of non-structural protein nsP1 of Semliki Forest virus attenuate virus replication and cause accumulation of compensatory mutations. *Journal of General Virology* 88, 1977–1985.
- (73) Tan, Y. B., Chmielewski, D., Law, M. C. Y., Zhang, K., He, Y., Chen, M., Jin, J., and Luo, D. (2022). Molecular architecture of the Chikungunya virus replication complex. *Science Advances* 8, 2536.
- (74) Rupp, J. C., Sokoloski, K. J., Gebhart, N. N., and Hardy, R. W. (2015). Alphavirus RNA synthesis and non-structural protein functions. *Journal of General Virology* 96, 2483–2500.
- (75) Gomatos, P. J., Kaariainen, L., Keranen, S., Ranki, M., and Sawicki, D. L. (1980). Semliki Forest Virus Replication Complex Capable of Synthesizing 42S and 26S Nascent RNA Chains. *Journal of General Virology* 49, 61–69.
- (76) Gomez De Cedrón, M., Ehsani, N., Mikkola, M. L., García, J. A., and Kääriäinen, L. (1999). RNA helicase activity of Semliki Forest virus replicase protein NSP2. *FEBS letters* 448, 19–22.
- (77) Rikkinen, M., Peranen, J., and Kaariainen, L. (1994). ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. *Journal of Virology* 68, 5804–5810.
- (78) Das, P. K., Merits, A., and Lulla, A. (2014). Functional Cross-talk between Distant Domains of Chikungunya Virus Non-structural Protein 2 Is Decisive for Its RNA-modulating Activity. DOI: 10.1074/jbc.M113.503433.
- (79) Hardy, W. R., and Strauss, J. H. (1989). Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *Journal of virology* 63, 4653–4664.
- (80) Hahn, Y. S., Strauss, E. G., and Strauss, J. H. (1989). Mapping of RNA- temperature-sensitive mutants of Sindbis virus: assignment of

REFERENCES

- complementation groups A, B, and G to nonstructural proteins. *Journal of virology* 63, 3142–3150.
- (81) Strauss, J. H., and Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews* 58, 491.
- (82) Vasiljeva, L., Merits, A., Golubtsov, A., Sizemskaja, V., Kääriäinen, L., and Ahola, T. (2003). Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *The Journal of biological chemistry* 278, 41636–41645.
- (83) Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A., and Marcotrigiano, J. (2012). Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 109, 16534–16539.
- (84) Suopanki, J., Sawicki, D. L., Sawicki, S. G., and Ka\$ A\$ Ria\$ Inen, L. (1998). Printed in Great Britain Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. *Journal of General Virology* 79, 309–319.
- (85) Sawicki, D. L., and Sawicki, S. G. (1993). A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis. *Journal of Virology* 67, 3605–3610.
- (86) Fros, J. J., Domeradzka, N. E., Baggen, J., Geertsema, C., Flipse, J., Vlak, J. M., and Pijlman, G. P. (2012). Chikungunya Virus nsP3 Blocks Stress Granule Assembly by Recruitment of G3BP into Cytoplasmic Foci. *Journal of Virology* 86, 10873–10879.
- (87) Fros, J. J., Geertsema, C., Zouache, K., Baggen, J., Domeradzka, N., Van Leeuwen, D. M., Flipse, J., Vlak, J. M., Failloux, A. B., and Pijlman, G. P. (2015). Mosquito Rasputin interacts with chikungunya virus nsP3 and determines the infection rate in *Aedes albopictus*. *Parasites and Vectors* 8, 1–15.
- (88) Scholte, F. E. M., Tas, A., Albulescu, I. C., Žusinaite, E., Merits, A., Snijder, E. J., and van Hemert, M. J. (2015). Stress Granule Components

REFERENCES

- G3BP1 and G3BP2 Play a Proviral Role Early in Chikungunya Virus Replication. *Journal of Virology* 89, 4457–4469.
- (89) Rathore, A. P., Haystead, T., Das, P. K., Merits, A., Ng, M. L., and Vasudevan, S. G. (2014). Chikungunya virus nsP3 & nsP4 interacts with HSP-90 to promote virus replication: HSP-90 inhibitors reduce CHIKV infection and inflammation in vivo. *Antiviral Research* 103, 7–16.
- (90) Gao, Y., Goonawardane, N., Ward, J., Tuplin, A., and Harris, M. (2019). Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during Chikungunya virus genome replication and transcription. *PLOS Pathogens* 15, ed. by Kuhn, R. J., e1007239.
- (91) Krieg, S., Pott, F., Ecke, L., Verheirstraeten, M., Bütepage, M., Lip-pok, B., Goffinet, C., Lüscher, B., and Verheugd, P. (2020). Mono-ADP-ribosylation by ARTD10 restricts Chikungunya virus replication by interfering with the proteolytic activity of nsP2. *bioRxiv*, 2020.01.07.896977.
- (92) Lulla, A., Lulla, V., and Merits, A. (2012). Macromolecular Assembly-Driven Processing of the 2/3 Cleavage Site in the Alphavirus Replicase Polyprotein. *Journal of Virology* 86, 553–565.
- (93) Rubach, J. K., Wasik, B. R., Rupp, J. C., Kuhn, R. J., Hardy, R. W., and Smith, J. L. (2009). Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* 384, 201–208.
- (94) Li, M.-L., Stollar, V., and Joklik, W. K. (2004). Identification of the amino acid sequence in Sindbis virus nsP4 that binds to the promoter for the synthesis of the subgenomic RNA. *PNAS* 101, 9429–9434.
- (95) Rupp, J. C., Jundt, N., and Hardy, R. W. (2011). Requirement for the Amino-Terminal Domain of Sindbis Virus nsP4 during Virus Infection. *Journal of Virology* 85, 3449–3460.
- (96) Shirako, Y., and Strauss, J. H. 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, tech. rep. 3, 1994, pp 1874–1885.

REFERENCES

- (97) HolmesID, A. C., Basore, K., Fremont, D. H., DiamondID, M. S., and Bursky, J. M. (2020). A molecular understanding of alphavirus entry. DOI: 10.1371/journal.ppat.1008876.
- (98) Zhang, R., Kim, A. S., Fox, J. M., Nair, S., Basore, K., Klimstra, W. B., Rimkunas, R., Fong, H., Lin, H., Poddar, S., Crowe, J. E., Doranz, B. J., Fremont, D. H., and Diamond, M. S. (2018). Mxra8 is a receptor for multiple arthritogenic alphaviruses. DOI: 10.1038/s41586-018-0121-3.
- (99) Basore, K., Kim, A. S., Nelson, C. A., Zhang, R., Smith, B. K., Uranga, C., Vang, L., Cheng, M., Gross, M. L., Smith, J., Diamond, M. S., and Fremont, D. H. (2019). Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. *Cell* 177, 1725–1737.
- (100) Song, H., Zhao, Z., Chai, Y., Jin, X., Li, C., Yuan, F., Liu, S., Gao, Z., Wang, H., Song, J., Vazquez, L., Zhang, Y., Tan, S., Morel, C. M., Yan, J., Shi, Y., Qi, J., Gao, F., and Gao, G. F. (2019). Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. *Cell* 177, 1714–1724.
- (101) Wintachai, P., Wikan, N., Kuadkitkan, A., Jaimipuk, T., Ubol, S., Pulmanausahakul, R., Auewarakul, P., Kasinrerak, W., Weng, W. Y., Panyasrivanit, M., Paemanee, A., Kittisenachai, S., Roytrakul, S., and Smith, D. R. (2012). Identification of prohibitin as a Chikungunya virus receptor protein. *Journal of medical virology* 84, 1757–1770.
- (102) De Caluwé, L., Coppens, S., Vereecken, K., Daled, S., Dhaenens, M., Ostade, X. V., Deforce, D., Ariën, K. K., and Bartholomeeusen, K. (2021). The CD147 Protein Complex Is Involved in Entry of Chikungunya Virus and Related Alphaviruses in Human Cells. DOI: 10.3389/fmicb.2021.615165.
- (103) Fongsaran, C., Jirakanwisal, K., Kuadkitkan, A., Wikan, N., Wintachai, P., Thepparit, C., Ubol, S., Phaonakrop, N., Roytrakul, S., and Smith, D. R. Involvement of ATP synthase b subunit in chikungunya virus entry into insect cells. DOI: 10.1007/s00705-014-2210-4.
- (104) Kielian, M., Chanel-Vos, C., and Liao, M. Alphavirus entry and membrane fusion, 2010.

REFERENCES

- (105) Lee, R. C. H., Hapuarachchi, H. C., Chen, K. C., Hussain, K. M., Chen, H., Low, S. L., Ng, L. C., Lin, R., Ng, M. M. L., and Chu, J. J. H. (2013). Mosquito Cellular Factors and Functions in Mediating the Infectious entry of Chikungunya Virus. *PLoS Neglected Tropical Diseases* 7, DOI: 10.1371/journal.pntd.0002050.
- (106) Ooi, Y. S., Stiles, K. M., Liu, C. Y., Taylor, G. M., and Kielian, M. (2013). Genome-Wide RNAi Screen Identifies Novel Host Proteins Required for Alphavirus Entry. *PLOS Pathogens* 9, e1003835.
- (107) Hoornweg, T. E., van Duijl-Richter, M. K. S., Ayala Nuñez, N. V., Albuлесcu, I. C., van Hemert, M. J., and Smit, J. M. (2016). Dynamics of Chikungunya Virus Cell Entry Unraveled by Single-Virus Tracking in Living Cells. *Journal of Virology* 90, 4745–4756.
- (108) Bernard, E., Solignat, M., Gay, B., Chazal, N., Higgs, S., Devaux, C., and Briant, L. (2010). Endocytosis of Chikungunya Virus into Mammalian Cells: Role of Clathrin and Early Endosomal Compartments. DOI: 10.1371/journal.pone.0011479.
- (109) Ulmanen, I., Soderlund, H., and Kaariainen, L. (1976). Semliki Forest virus capsid protein associates with the 60S ribosomal subunit in infected cells. *Journal of Virology* 20, 203–210.
- (110) Singh, I., and Helenius, A. (1992). Role of ribosomes in Semliki Forest virus nucleocapsid uncoating. *Journal of Virology* 66, 7049–7058.
- (111) Sokoloski, K. J., Nease, L. M., May, N. A., Gebhart, N. N., Jones, C. E., Morrison, T. E., and Hardy, R. W. (2017). Identification of Interactions between Sindbis Virus Capsid Protein and Cytoplasmic vRNA as Novel Virulence Determinants. *PLoS Pathogens* 13, DOI: 10.1371/journal.ppat.1006473.
- (112) Jones, J. E., Long, K. M., Whitmore, A. C., Sanders, W., Thurlow, L. R., Brown, J. A., Morrison, C. R., Vincent, H., Peck, K. M., Browning, C., Moorman, N., Lim, J. K., Heise, M. T., and Dermody, T. S. (2017). Disruption of the Opal Stop Codon Attenuates Chikungunya Virus-Induced Arthritis and Pathology. DOI: 10.1128/mBio.

REFERENCES

- (113) Jones, R., Bragagnolo, G., Arranz, R., and Reguera, J. (2020). Capping pores of alphavirus nsP1 gate membranous viral replication factories. *Nature* 2020 589:7843 589, 615–619.
- (114) Bartholomeeusen, K., Utt, A., Coppens, S., Rausalu, K., Vereecken, K., Ariën, K. K., and Merits, A. (2018). A Chikungunya Virus trans-Replicase System Reveals the Importance of Delayed Nonstructural Polyprotein Processing for Efficient Replication Complex Formation in Mosquito Cells. *Journal of Virology* 92, DOI: 10.1128/JVI.00152-18/ASSET/9E742AF0-ACE3-40BB-B239-80E86B196FD6/ASSETS/GRAPHIC/ZJV0141836670007.JPEG.
- (115) Lemm, J. A., Rümenapf, T., Strauss, E. G., Strauss, J. H., and Rice, C. M. (1994). Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *The EMBO journal* 13, 2925–34.
- (116) Thomas, S., Rai, J., John, L., Günther, S., Drosten, C., Pützer, B. M., and Schaefer, S. (2010). Functional dissection of the alphavirus capsid protease: Sequence requirements for activity. *Virology Journal* 7, 1–7.
- (117) Metz, S. W., Geertsema, C., Martina, B. E., Andrade, P., Heldens, J. G., Van Oers, M. M., Vlak, J. M., and Pijlman, G. P. (2011). Functional processing and secretion of Chikungunya virus E1 and E2 glycoproteins in insect cells. *Virology Journal* 8, DOI: 10.1186/1743-422X-8-353.
- (118) Singh, A., Kumar, A., Uversky, V. N., and Giri, R. (2018). Understanding the interactability of chikungunya virus proteins: Via molecular recognition feature analysis. *RSC Advances* 8, 27293–27303.
- (119) Nelson, S., Hernandez, R., Ferreira, D., and Brown, D. T. (2005). In vivo processing and isolation of furin protease-sensitive alphavirus glycoproteins: A new technique for producing mutations in virus assembly. *Virology* 332, 629–639.
- (120) Jose, J., Taylor, A. B., and Kuhn, R. J. (2017). Spatial and Temporal Analysis of Alphavirus Replication and Assembly in Mammalian and Mosquito Cells. *mBio* 8, DOI: 10.1128/MBIO.02294-16.

REFERENCES

- (121) Simizu, B., Yamamoto, K., Hashimoto, K., and Ogata, T. (1984). Structural proteins of Chikungunya virus. *Journal of Virology* 51, 254–258.
- (122) Lee, C. Y., Kam, Y. W., Fric, J., Malleret, B., Koh, E. G., Prakash, C., Huang, W., Lee, W. W., Lin, C., Lin, R. T., Renia, L., Wang, C. I., Ng, L. F., and Warter, L. (2011). Chikungunya Virus Neutralization Antigens and Direct Cell-to-Cell Transmission Are Revealed by Human Antibody-Escape Mutants. *PLOS Pathogens* 7, e1002390.
- (123) Varani, G., and McClain, W. H. (2000). The G·U wobble base pair: A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Reports* 1, 18.
- (124) Svoboda, P., and Di Cara, A. Hairpin RNA: A secondary structure of primary importance, 2006.
- (125) Peselis, A., and Serganov, A. Structure and function of pseudoknots involved in gene expression control, 2014.
- (126) Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., Segal, E., and Contributions, A. M. (2010). Genome-wide Measurement of RNA Secondary Structure in Yeast helped with sequencing; M. *Nature* 467, DOI: 10.1038/nature09322.
- (127) Barnwal, R. P., Yang, F., and Varani, G. Applications of NMR to structure determination of RNAs large and small. DOI: 10.1016/j.abb.2017.06.003.
- (128) Xu, B., Zhu, Y., Cao, C., Chen, H., Jin, Q., Li, G., Ma, J., Yang, S. L., Zhao, J., Zhu, J., Ding, Y., Fang, X., Jin, Y., Kwok, C. K., Ren, A., Wan, Y., Wang, Z., Xue, Y., Zhang, H., Zhang, Q. C., and Zhou, Y. Recent advances in RNA structurome, 2022.
- (129) Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.-C., Schmid, M. F., Das, R., and Chiu, W. (2019). Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature Communications* 10, DOI: 10.1038/s41467-019-13494-7.

REFERENCES

- (130) Huo, X., Han, S., Wu, G., Latchoumanin, O., Zhou, G., Hebbard, L., George, J., and Qiao, L. Dysregulated long noncoding RNAs (lncRNAs) in hepatocellular carcinoma: Implications for tumorigenesis, disease progression, and liver cancer stem cells, 2017.
- (131) Mirsafian, H., Manda, S. S., Mitchell, C. J., Sreenivasamurthy, S., Ripen, A. M., Mohamad, S. B., Merican, A. F., and Pandey, A. (2016). Long non-coding RNA expression in primary human monocytes. *Genomics* 108, 37–45.
- (132) Ponjavic, J., Ponting, C. P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Research* 17, DOI: 10.1101/gr.6036807.
- (133) Novikova, I. V., Hennelly, S. P., and Sanbonmatsu, K. Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acid Research* 40, DOI: 10.1093/nar/gks071.
- (134) Zhao, X., Patton, J. R., Ghosh, S. K., Fischel-Ghodsian, N., Shen, L., and Spanjaard, R. A. (2007). Pus3p-and Pus1p-Dependent Pseudouridylation of Steroid Receptor RNA Activator Controls a Functional Switch that Regulates Nuclear Receptor Signaling. DOI: 10.1210/me.2006-0414.
- (135) Kurisu, T., Tanaka, T., Ishii, J., Matsumura, K., Sugimura, K., Nakatani, T., and Kawashima, H. (2006). Expression and function of human steroid receptor RNA activator in prostate cancer cells: role of endogenous hSRA protein in androgen receptor-mediated transcription. *Prostate Cancer and Prostatic Diseases* 9, 173–178.
- (136) Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97, 17–27.
- (137) Van Der Velden, A. W., and Thomas, A. A. (1999). The role of the 5 untranslated region of an mRNA in translation regulation during development. *The International Journal of Biochemistry & Cell Biology* 31, 87–106.

REFERENCES

- (138) Jansen, R. P. (2001). mRNA localization: message on the move. *Nature reviews. Molecular cell biology* 2, 247–256.
- (139) Haimov, O., Sinvani, H., and Dikstein, R. (2015). Cap-dependent, scanning-free translation initiation mechanisms. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1849, 1313–1318.
- (140) Muckenthaler, M. U., Rivella, S., Hentze, M. W., and Galy, B. (2017). A Red Carpet for Iron Metabolism. *Cell* 168, 344–361.
- (141) Hentze, M. W., Caughman, S. W., Casey, J. L., Kodier, D. M., Rouault, T. A., Harford, J. B., and Klausner, R. D. (1988). A model for the structure and functions of iron-responsive elements. *Gene* 72, 201–208.
- (142) Peixeiro, I., Silva, A. L., and Romão, L. (2011). Control of human β -globin mRNA stability and its impact on beta-thalassemia phenotype. *Haematologica* 96, DOI: 10.3324/haematol.2010.039206.
- (143) Leppek, K., Das, R., and Barna, M. (2017). Secondary structures Tertiary structures Interactions that orient RNA double helices into specific three-dimensional arrangements, often involving non-Watson-Crick base pairs. Functional 5 UTR mRNA structures in eukaryotic translation regulation and how to find them. DOI: 10.1038/nrm.2017.103.
- (144) Carrillo, C., Tulman, E. R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G. F., and Rock, D. L. (2005). Comparative Genomics of Foot-and-Mouth Disease Virus. *Journal of Virology* 79, 6487–6504.
- (145) López-Manríquez, E., Vashist, S., Ureña, L., Goodfellow, I., Chavez, P., Mora-Heredia, J. E., Cancio-Lonches, C., Garrido, E., and Gutiérrez-Escolano, A. L. (2013). Norovirus genome circularization and efficient replication are facilitated by binding of PCBP2 and hnRNP A1. *Journal of virology* 87, 11371–11387.
- (146) Harrich, D., Mavankal, G., Mette-Snider, A., and Gaynor, R. B. (1995). Human immunodeficiency virus type 1 TAR element revertant viruses define RNA structures required for efficient viral gene expression and replication. *Journal of Virology* 69, 4906–4913.

REFERENCES

- (147) Von Der Haar, T., Gross, J. D., Wagner, G., and McCarthy, J. E. (2004). The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nature Structural and Molecular Biology* 11, 503–511.
- (148) Flores, R., Romero-López, C., Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., and Embarek, A. M. (2018). Insights into Structural and Mechanistic Features of Viral IRES Elements. 8, 2629.
- (149) Hinnebusch, A. G. (2014). The Scanning Mechanism of Eukaryotic Translation Initiation. DOI: 10.1146/annurev-biochem-060713-035802.
- (150) Yang, Y., and Wang, Z. (2019). IRES-mediated cap-independent translation, a path leading to hidden proteome. *Journal of Molecular Cell Biology* 911, 911–919.
- (151) Komar, A. A., and Hatzoglou, M. (2011). Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell cycle (Georgetown, Tex.)* 10, 229–240.
- (152) Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G., and Diaz-Toledano, R. (2015). Picornavirus IRES elements: RNA structure and host protein interactions. *Virus Research* 206, 62–73.
- (153) Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334.
- (154) Heng, X., Kharytonchyk, S., Garcia, E. L., Lu, K., Divakaruni, S. S., Lacotti, C., Edme, K., Telesnitsky, A., and Summers, M. F. (2012). Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *Journal of molecular biology* 417, 224–239.
- (155) Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosis, A., Florwick, A., Santos, J., Bolden, N. C., Mccowin, S., Case, D. A., Johnson, B., Salemi, M., Telesnitsky, A., and Summers, M. F. (2015). Structure of the HIV-1 RNA Packaging Signal HHS Public Access. *Science* 348, 917–921.

REFERENCES

- (156) Nikolaitchik, O. A., Somoulay, X., Rawson, J. M. O., Yoo, J. A., Pathak, V. K., and Hu, W.-S. (2020). Unpaired Guanosines in the 5' Untranslated Region of HIV-1 RNA Act Synergistically To Mediate Genome Packaging. *Journal of Virology* *94*, DOI: 10.1128/JVI.00439-20.
- (157) Ingemarsdotter, C. K., Zeng, J., Long, Z., Lever, A. M. L., and Kenyon, J. C. (2018). An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *15*, 25.
- (158) Malnou, C. E., Pöyry, T. A. A., Jackson, R. J., and Kean, K. M. (2002). Poliovirus Internal Ribosome Entry Segment Structure Alterations That Specifically Affect Function in Neuronal Cells: Molecular Genetic Analysis. *Journal of Virology* *76*, 10617–10626.
- (159) Olson, E. D., Cantara, W. A., and Musier-Forsyth, K. (2015). New Structure Sheds Light on Selective HIV-1 Genomic RNA Packaging. *Viruses* *2015*, Vol. 7, Pages 4826-4835 7, 4826–4835.
- (160) Sanford, T. J., Mears, H. V., Fajardo, T., Locker, N., and Sweeney, T. R. (2019). Circularization of flavivirus genomic RNA inhibits de novo translation initiation. *Nucleic Acids Research* *47*, 9789–9802.
- (161) Filomatori, C. V., Lodeiro, M. F., Alvarez, D. E., Samsa, M. M., Pietrasanta, L., and Gamarnik, A. V. (2006). A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes and Development* *20*, 2238–2249.
- (162) Dong, H., Zhang, B., and Shi, P. Y. (2008). Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology* *381*, 123–135.
- (163) Liu, Z.-Y., Li, X.-F., Jiang, T., Deng, Y.-Q., Ye, Q., Zhao, H., Yu, J.-Y., and Qin, C.-F. Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. DOI: 10.7554/eLife.17636.001.

REFERENCES

- (164) Villordo, S. M., Alvarez, D. E., and Gamarnik, A. V. (2010). A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA* 16, 2325.
- (165) Kim, D. Y., Firth, A. E., Atasheva, S., Frolova, E. I., and Frolov, I. (2011). Conservation of a Packaging Signal and the Viral Genome RNA Packaging Mechanism in Alphavirus Evolution. *Journal of Virology* 85, 8022–8036.
- (166) Kendra, J. A., Advani, V. M., Chen, B., Briggs, J. W., Zhu, J., Bress, H. J., Pathy, S. M., and Dinman, J. D. (2018). Functional and structural characterization of the chikungunya virus translational recoding signals. *Journal of Biological Chemistry* 293, 17536–17545.
- (167) Sanz, M. Á., Castelló, A., Ventoso, I., Berlanga, J. J., and Carrasco, L. (2009). Dual Mechanism for the Translation of Subgenomic mRNA from Sindbis Virus in Infected and Uninfected Cells. *PLOS ONE* 4, e4772.
- (168) Hardy, R. W. (2006). The role of the 3' terminus of the Sindbis virus genome in minus-strand initiation site selection. *Virology* 345, 520–531.
- (169) Fayzulin, R., and Frolov, I. (2004). Changes of the Secondary Structure of the 5 End of the Sindbis Virus Genome Inhibit Virus Growth in Mosquito Cells and Lead to Accumulation of Adaptive Mutations. *Journal of Virology* 78, 4953–4964.
- (170) Hardy, R. W., and Rice, C. M. (2005). Requirements at the 3' end of the sindbis virus genome for efficient synthesis of minus-strand RNA. *Journal of virology* 79, 4630–4639.
- (171) Kulasegaran-Shylini, R., Thiviyanathan, V., Gorenstein, D. G., and Frolov, I. (2009). 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* 387, 211–221.
- (172) Hyde, J. L., Gardner, C. L., Kimura, T., White, J. P., Liu, G., Trobaugh, D. W., Huang, C., Tonelli, M., Paessler, S., Takeda, K., Klimstra, W. B., Amarasinghe, G. K., and Diamond, M. S. (2014). A viral RNA structural element alters host recognition of nonself RNA. *Science* 343, 783–787.

REFERENCES

- (173) Kendall, C., Khalid, H., Uller, M. M. ., Banda, D. H., Kohl, A., Merits, A., Stonehouse, N. J., and Tuplin, A. (2019). Structural and phenotypic analysis of Chikungunya virus RNA replication elements. *Nucleic Acids Research* *47*, 9296–9312.
- (174) Madden, E. A., Plante, K. S., Morrison, C. R., Kutchko, K. M., Sanders, W., Long, K. M., Taft-Benz, S., Cruz Cisneros, M. C., White, A. M., Sarkar, S., Reynolds, G., Vincent, H. A., Laederach, A., Moorman, N. J., and Heise, M. T. (2020). Using SHAPE-MaP To Model RNA Secondary Structure and Identify 3UTR Variation in Chikungunya Virus. *Journal of Virology* *94*, 701–721.
- (175) Nickens, D. G., and Hardy, R. W. (2008). Structural and functional analyses of stem-loop 1 of the Sindbis virus genome. *Virology* *370*, 158–172.
- (176) Frolov, I., Hardy, R., and Rice, C. M. (2001). Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis.
- (177) Niesters, H. G., and Strauss, J. H. (1990). Defined mutations in the 5' nontranslated sequence of Sindbis virus RNA. *Journal of virology* *64*, 4162–4168.
- (178) Ou, J. H., Strauss, E. G., and Strauss, J. H. (1981). Comparative studies of the 3-terminal sequences of several alphavirus RNAs. *Virology* *109*, 281–289.
- (179) Johnson, B., VanBlargan, L. A., Xu, W., White, J. P., Shan, C., Shi, P. Y., Zhang, R., Adhikari, J., Gross, M. L., Leung, D. W., Diamond, M. S., and Amarasinghe, G. K. (2018). Binding of IFIT1 by IFIT3 modulates protein stability and RNA binding specificity. *Immunity* *48*, 487.
- (180) Reynaud, J. M., Kim, D. Y., Atasheva, S., Rasaloukaya, A., White, J. P., Diamond, M. S., Weaver, S. C., Frolova, E. I., and Frolov, I. (2015). IFIT1 Differentially Interferes with Translation and Replication of Alphavirus Genomes and Promotes Induction of Type I Interferon. *PLoS Pathogens* *11*, DOI: 10.1371/journal.ppat.1004863.

REFERENCES

- (181) Gorchakov, R., Hardy, R., Rice, C. M., and Frolov, I. (2004). Selection of Functional 5 cis-Acting Elements Promoting Efficient Sindbis Virus Genome Replication. *JOURNAL OF VIROLOGY* 78, 61–75.
- (182) Frey, T. K., Gard, D. L., and Strauss, J. H. (1979). Biophysical studies on circle formation by Sindbis virus 49 S RNA. *Journal of Molecular Biology* 132, 1–18.
- (183) Kulasegaran-Shylini, R., Atasheva, S., Gorenstein, D. G., and Frolov, I. (2009). Structural and Functional Elements of the Promoter Encoded by the 5 Untranslated Region of the Venezuelan Equine Encephalitis Virus Genome. *Journal of Virology* 83, 8327.
- (184) Michel, G., Petrakova, O., Atasheva, S., and Frolov, I. (2007). Adaptation of Venezuelan equine encephalitis virus lacking 51-nt conserved sequence element to replication in mammalian and mosquito cells. *Virology* 362, 475–487.
- (185) Hyde, J. L., Chen, R., Trobaugh, D. W., Diamond, M. S., Weaver, S. C., Klimstra, W. B., and Wilusz, J. (2015). The 5 and 3 ends of alphavirus RNAs-Non-coding is not non-functional. *Virus Research* 206, 99–107.
- (186) Kuhn, R. J., Hong, Z., and Strauss, J. H. (1990). Mutagenesis of the 3' Nontranslated Region of Sindbis Virus RNA. *JOURNAL OF VIROLOGY* 64, 1465–1476.
- (187) Chen, R., Wang, E., Tsetsarkin, K. A., and Weaver, S. C. (2013). Chikungunya Virus 3 Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces. *PLoS Pathogens* 9, DOI: 10.1371/journal.ppat.1003591.
- (188) Kuhn, R. J., Griffin, D. E., Zhang, H., Niesters, H. G., and Strauss, J. H. (1992). Attenuation of Sindbis virus neurovirulence by using defined mutations in nontranslated regions of the genome RNA. *Journal of virology* 66, 7121–7127.
- (189) Hyde, J. L., Chen, R., Trobaugh, D. W., Diamond, M. S., Weaver, S. C., Klimstra, W. B., and Wilusz, J. (2015). The 5' and 3' ends of alphavirus

REFERENCES

- RNAs-non-coding is not non-functional HHS Public Access. *Virus Res* 206, 99–107.
- (190) Sokoloski, K. J., Dickson, A. M., Chaskey, E. L., Garneau, N. L., Wilusz, C. J., and Wilusz, J. (2010). Sindbis virus usurps the cellular HuR protein to stabilize its transcripts and promote productive infections in mammalian and mosquito cells. *Cell host & microbe* 8, 196–207.
- (191) Pardigon, N., and Strauss, J. H. (1996). Mosquito homolog of the La autoantigen binds to Sindbis virus RNA. *Journal of Virology* 70, 1173.
- (192) Dickson, A. M., Anderson, J. R., Barnhart, M. D., Sokoloski, K. J., Oko, L., Opyrchal, M., Galanis, E., Wilusz, C. J., Morrison, T. E., and Wilusz, J. (2012). Dephosphorylation of HuR protein during alphavirus infection is associated with HuR relocalization to the cytoplasm. *The Journal of biological chemistry* 287, 36229–36238.
- (193) Garneau, N. L., Sokoloski, K. J., Opyrchal, M., Neff, C. P., Wilusz, C. J., and Wilusz, J. (2008). The 3 Untranslated Region of Sindbis Virus Represses Deadenylation of Viral Transcripts in Mosquito and Mammalian Cells. *Journal of Virology* 82, 880.
- (194) Strauss, E. G., Rice, C. M., and Strauss, J. H. (1983). Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. *Proceedings of the National Academy of Sciences of the United States of America* 80, 5271–5275.
- (195) Firth, A. E., Wills, N. M., Gesteland, R. F., and Atkins, J. F. (2011). Stimulation of stop codon readthrough: frequent presence of an extended 3 RNA structural element. *Nucleic Acids Research* 39, 6679–6691.
- (196) Firth, A. E., Chung, B. Y., Fleton, M. N., and Atkins, J. F. (2008). Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virology Journal* 5, 108.
- (197) Kendra, J. A., Fuente, C. d. l., Brahms, A., Woodson, C., Bell, T. M., Chen, B., Khan, Y. A., Jacobs, J. L., Kehn-Hall, K., and Dinman, J. D. (2017). Ablation of Programmed 1 Ribosomal Frameshifting in Venezuelan

REFERENCES

- Equine Encephalitis Virus Results in Attenuated Neuropathogenicity. *Journal of Virology* 91, DOI: 10.1128/JVI.01766-16.
- (198) Ventoso, I., Sanz, M. A., Molina, S., Berlanga, J. J., Carrasco, L., and Esteban, M. (2006). Translational resistance of late alphavirus mRNA to eIF2alpha phosphorylation: a strategy to overcome the antiviral effect of protein kinase PKR. *Genes & development* 20, 87–100.
- (199) Castelló, A., Sanz, M. Á., Molina, S., and Carrasco, L. (2006). Translation of Sindbis Virus 26 S mRNA Does Not Require Intact Eukariotic Initiation Factor 4G. *Journal of Molecular Biology* 355, 942–956.
- (200) Sanz, M. A., González Almela, E., and Carrasco, L. (2017). Translation of Sindbis Subgenomic mRNA is Independent of eIF2, eIF2A and eIF2D. *Scientific Reports* 2017 7:1 7, 1–14.
- (201) Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., De Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., and Loman, N. J. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* 12, 1261–1266.
- (202) Ferguson, J. M., Gamaarachchi, H., Nguyen, T., Gollon, A., Tong, S., Aquilina-Reid, C., Bowen-James, R., and Deveson, I. W. InterARTIC: an interactive web application for whole-genome nanopore sequencing analysis of SARS-CoV-2 and other viruses. DOI: 10.1093/bioinformatics/btab846.
- (203) Rana, J., Rajasekharan, S., Gulati, S., Dudha, N., Gupta, A., Chaudhary, V. K., and Gupta, S. (2014). Network mapping among the functional domains of Chikungunya virus nonstructural proteins. *Proteins: Structure, Function, and Bioinformatics* 82, 2403–2411.
- (204) Tomar, S., Narwal, M., Harms, E., Smith, J. L., and Kuhn, R. J. (2011). Heterologous production, purification and characterization of enzymati-

REFERENCES

- cally active Sindbis virus nonstructural protein nsP1. *Protein Expression and Purification* 79, 277–284.
- (205) Zhang, K., Law, Y. S., Law, M. C. Y., Tan, Y. B., Wirawan, M., and Luo, D. (2021). Structural insights into viral RNA capping and plasma membrane targeting by Chikungunya virus nonstructural protein 1. *Cell Host and Microbe* 29, 757–764.
- (206) Li, C., Guillén, J., Rabah, N., Blanjoie, A., Debart, F., Vasseur, J.-J., Canard, B., Decroly, E., and Coutard, B. (2015). mRNA Capping by Venezuelan Equine Encephalitis Virus nsP1: Functional Characterization and Implications for Antiviral Research. *Journal of virology* 89, 8292–303.
- (207) Laakkonen, P., Auvinen, P., Kujala, P., and Kääriäinen, L. (1998). Alphavirus replicase protein NSP1 induces filopodia and rearrangement of actin filaments. *Journal of virology* 72, 10265–10269.
- (208) Ahola, T., Kujala, P., Tuittila, M., Blom, T., Laakkonen, P., Hinkkanen, A., and Auvinen, P. (2000). Effects of palmitoylation of replicase protein nsP1 on alphavirus infection. *Journal of virology* 74, 6725–6733.
- (209) Shirako, Y., Strauss, E. G., and Strauss, J. H. (2000). 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* 276, 148–160.
- (210) Fata, C. L., Sawicki, S. G., and Sawicki, D. L. (2002). Modification of Asn374 of nsP1 Suppresses a Sindbis Virus nsP4 Minus-Strand Polymerase Mutant. *Journal of Virology* 76, 8641.
- (211) Wang, Y. F., Sawicki, S. G., and Sawicki, D. L. (1991). Sindbis virus nsP1 functions in negative-strand RNA synthesis. *Journal of Virology* 65, 985.
- (212) Roberts, G. C., Zothner, C., Remenyi, R., Merits, A., Stonehouse, N. J., and Harris, M. (2017). Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research. *Scientific Reports* 7, DOI: 10.1038/s41598-017-15269-w.

REFERENCES

- (213) Müller, M., Slivinski, N., Todd, E. J., Khalid, H., Li, R., Karwatka, M., Merits, A., Mankouri, J., and Tuplin, A. (2019). Chikungunya virus requires cellular chloride channels for efficient genome replication. *PLoS Neglected Tropical Diseases* 13, DOI: 10.1371/journal.pntd.0007703.
- (214) Utt, A., Rausalu, K., Jakobson, M., Männik, A., Alphey, L., Frangkoudis, R., Merits, A., and Dermody, T. S. (2019). Design and Use of Chikungunya Virus Replication Templates Utilizing Mammalian and Mosquito RNA Polymerase I-Mediated Transcription. *Journal of Virology* 93, 794–813.
- (215) Lello, L. S., Utt, A., Bartholomeeusen, K., Wangid, S., Rausaluid, K., Kendallid, C., Coppens, S., Frangkoudis, R., Tuplinid, A., Alpheyid, L., Arië N, K. K., and Meritsid, A. (2020). Cross-utilisation of template RNAs by alphavirus replicases. DOI: 10.1371/journal.ppat.1008825.
- (216) Kearse, M. G., and Wilusz, J. E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes & development* 31, 1717–1731.
- (217) Kutchko, K. M., Madden, E. A., Morrison, C., Plante, K. S., Sanders, W., Vincent, H. A., Cisneros, M. C. C., Long, K. M., Moorman, N. J., Heise, M. T., and Laederach, A. (2018). Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Research* 46, 3657–3670.
- (218) Karabiber, F., Mcginnis, J. L., Favorov, O. V., and Weeks, K. M. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. DOI: 10.1261/rna.036327.112.
- (219) Arteché-López, A., Ávila-Fernández, A., Romero, R., Riveiro-Álvarez, R., López-Martínez, M. A., Giménez-Pardo, A., Vélez-Monsalve, C., Gallego-Merlo, J., García-Vara, I., Almoguera, B., Bustamante-Aragonés, A., Blanco-Kelly, F., Tahsin-Swafiri, S., Rodríguez-Pinilla, E., Minguez, P., Lorda, I., Trujillo-Tiebas, M. J., and Ayuso, C. (2021). Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Scientific Reports* 2021 11:1 11, 1–7.

REFERENCES

- (220) Vicens, Q., and Kieft, J. S. (2022). Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences of the United States of America* 119, DOI: 10.1073/PNAS.2112677119/-/DCSUPPLEMENTAL.
- (221) Asano, K. (2014). Review Article: Why is start codon selection so precise in eukaryotes? *Translation* 2, e28387.
- (222) Takahashi, K., Maruyama, M., Tokuzawa, Y., Murakami, M., Oda, Y., Yoshikane, N., Makabe, K. W., Ichisaka, T., and Yamanaka, S. (2005). Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). *Genomics* 85, 360–371.
- (223) Wei, Y., Thyparambil, A. A., and Latour, R. A. Protein Helical Structure Determination Using CD Spectroscopy for Solutions with Strong Background Absorbance from 190-230 nm. DOI: 10.1016/j.bbapap.2014.10.001.
- (224) Khoshouei, M., Radjainia, M., Baumeister, W., and Danev, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nature Communications* 8, DOI: 10.1038/ncomms16099.
- (225) Ahola, T., and Ahlquist, P. (1991). Deletion analysis of brome mosaic virus 2a protein: effects on RNA replication and systemic spread. *Journal of virology* 65, 2807–15.
- (226) Niesters, H. G., and Strauss, J. H. (1990). Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. *Journal of virology* 64, 1639–1647.
- (227) Mustoe, A. M., Brooks, C. L., and Al-Hashimi, H. M. (2014). Hierarchy of RNA functional dynamics. *Annual review of biochemistry* 83, 441–466.
- (228) Dethoff, E. A., Chugh, J., Mustoe, A. M., and Al-Hashimi, H. M. (2012). Functional complexity and regulation through RNA dynamics. DOI: 10.1038/nature10885.
- (229) Cruz, J. A., and Westhof, E. (2009). The Dynamic Landscapes of RNA Architecture. *Cell* 136, 604–609.