Defining RNA packaging signals for virus assembly

Emma Louise Wroblewski

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

The University of Leeds

School of Molecular and Cellular Biology

September 2018

The candidate confirms that the work submitted is her 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.

Chapter 4: Defining the STNV coat protein recognition motif.

Publication: Rewriting nature's assembly manual for a ssRNA virus. Nikesh Patel, Emma Wroblewski, German Leonov, Simon E. V. Phillips, Roman Tuma, Reidun Twarock, and Peter G. Stockley. PNAS November 14, 2017. 114 (46) 12255-12260. https://doi.org/10.1073/pnas.1706951114

Nikesh Patel, Emma Wroblewski and German Leonov performed research.

Nikesh Patel: "Roles of electrostatics and PS co-operativity in VLP assembly" Figures 3 and 4, supplementary Figures 2-5, "Assembly of nonviral substrates" supplementary Figures 6-8, "Transfer of critical assembly features to genomic scale RNA's" Figure 5 and supplementary Figure 9.

Emma Wroblewski: "Sequence specific recognition of individual PS sites" Figure 2 and supplementary Figure 1.

German Leonov: Designed STNV-1 ΔAXXA construct.

Nikesh Patel, Emma Wroblewski, Simon E.V. Phillips, Roman Tuma, Reidun Twarock and Peter G. Stockley wrote the paper.

Acknowledgements

“I have been bent and broken, but - I hope - into a better shape”

Charles Dickens, Great Expectations.

There are no better words I think, to sum up my PhD. It has been exhilarating and frustrating. I have learnt a lot about myself and what I am capable of, and hopefully come out of the other side a better researcher. Firstly, I would like to thank my supervisor, Peter Stockley, for his support and motivation. Thank you for always pushing me to achieve my best and for guiding me through this journey. I’d like to thank Nikesh Patel, for teaching me the joys of STNV, how to use an AKTA without breaking it, and the frustrating magic of protein production and Simon White, for teaching me everything I know about RNA purification and poster making on adobe illustrator. Thanks for putting up with me and all my questions/translation queries. Thanks to Becca Chandler-Bostock and Amanda Richardson for keeping me sane, seeing me through some tough times and being all-round great friends, I could not have done this without you. Huge thanks go to Amy Barker for being such a great example and fountain of wisdom during my PhD, and buying me sweets during my write up. You all truly are #POSTDOCGOALS. Thanks also to my amazing Masters student Chloe Fisher for all her hard work.

On a more personal level I’d like to thank my family, especially my parents, for teaching me I that could achieve anything with hard work, and my grandad (B.M.W) for constant inspiration with his example of life-long learning. Thanks to the Brendan Rogers fan club for sharing the ups and downs of the PhD

experience, from bake off to board games and everything in between it was a blast. Thanks also to friends outside the lab for listening to me talk about my work and nodding and smiling supportively.

Lastly, I'd like to thank Alex, this honestly would have been impossible without you. Thank you for your unwavering support, for showing me how strong I can be and believing in me when I didn't believe in myself.

Abstract

Assembly mechanisms of single-stranded, positive-sense RNA viruses have been thought to be dominated by protein-protein interactions. However, a protein-controlled mechanism does not account for observed packaging selectivity seen *in vivo.* Previous investigations in our laboratory identified multiple areas of viral genomes with affinity for cognate coat protein (CP), termed Packaging Signals (PSs). This recently discovered assembly mechanism depends on multiple sequence-specific CP-RNA interactions, at sites distributed throughout the genome, leading to highly co-operative assembly along a defined pathway. Using a library of PS sequence variants, the key recognition features of the RNA-protein interaction has been investigated for a model RNA virus, Satellite Tobacco Necrosis Virus (STNV). Results suggest that AXXA is the optimal recognition motif, with the 3´ adenosine in the loop being more important for assembly than its 5´ equivalent. The collective action of multiple STNV PSs was investigated using sequence variants of both the 5´ 127 nt genomic fragment and full-length templates. RNAs containing PSs are preferentially packaged over those lacking PSs highlighting their important role(s) in assembly. Although the 5´ 127 nt of STNV encompasses the highest affinity PS, it is unable to restore packaging of a PS deficient construct to wild type levels, confirming that although this region may act as a nucleation site, other PSs are required for assembly. Identification of PS motifs is a key requirement in such studies requiring access to viral capsids/core proteins (Cp). To explore PS-mediated assembly in alphaviruses, bacterial and insect cell protein expression systems were established (3 Cps), purification strategies developed (2 Cps), and preliminary self-assembly studies carried out (2 Cps). The role(s) of PSs in assembly have been shown to be a novel antiviral target which would be useful in treatment of alphavirus infections given no drugs or vaccines are currently available. This project is now set-up for further studies to begin.

[This page has been left intentionally blank]

Table of Contents

Word count: 59,055

List of Figures

List of Tables

Abbreviations

Å	angstrom (10^{-10} m)
aa	amino acid
AcMNPV	autographa californica multicapsid nucleopolyhedrovirus
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
APS	ammonium persulphate
AF	Alexa Fluor
ARD	arginine-rich domain
ART	antiretroviral therapy
ATP	adenosine triphosphate
AUD	Australian dollar
BHK	baby hamster kidney
BMV	brome mosaic virus
bp	base pair
BPMV	bean pod mottle virus
BVD	bovine viral diarrhoea/mucosal disease
cccDNA	covalently closed circular DNA
CD	circular dichroism
CDMS	charge detection mass spectrometry
CHIKV	chikungunya virus
CLP	core like particle
CNS	central nervous system
CP	capsid protein
Cp	core protein

CPMV	cowpea mosaic virus
CPV	cytopathic vacuole
cryoEM	cryo-electron microscopy
CTAB	cetrimonium bromide
CTP	cytidine triphosphate
DEFRA	Department for Environment, Food and Rural Affairs
DEPC	diethyl pyrocarbonate
DIP	defective interfering particle
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EB	elution buffer
ECDC	European Centre for Disease Prevention and Control
EDTA	ethyldiaminetetraacetic acid
EEE	eastern equine encephalitis
EEEV	eastern equine encephalitis virus
EM	electron microscope/microscopy
ER	endoplasmic reticulum
FMDV	foot and mouth disease virus
FPLC	fast protein liquid chromatography
GTP	guanosine triphosphate
HBV	hepatitis b virus
HCL	hydrochloride
HF	high fidelity
HIV	human immunodeficiency virus

HPeV	human parecho virus
HPV	human papilloma virus
ICTV	International Committee on the Taxonomy of Viruses
IDT	Integrated DNA Technologies
INF	interferon
IPTG	isopropyl β-D-1-thiogalactopyranoside
IPV	inactivated polio vaccine
LC-MS	liquid chromatography mass spectrometry
LDF	linear discriminative function
LED	light emitting diode
LB	Luria-Bertani medium
LS	light scattering
MATLAB	matrix laboratory
MOI	multiplicity of infection
MOPS	3-[N-morpholino]propanesulphonic acid
mRNA	messenger RNA
MS	mass spectrometry
MWCO	molecular weight cut off
NC	nucleocapsid
NCS	non-crystallographic symmetry
NCSA	non-crystallographic symmetry averaging
NEB	New England Biolabs
NHS	N-hydroxysuccinimide
nm	nanometres
nM	nanomolar

NNDSS	National Notifiable Disease Surveillance System
nt	nucleotide
OPV	oral polio vaccine
ORF	open reading frame
PAHO	Pan American Health Organization
PAGE	polyacrylamide gel electrophoresis
PaV	pariacoto virus
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PH	polyhedrin
pI	isoelectric point
PIB	polyhedral inclusion bodies
PMSF	phenylmethane sulfonyl fluoride
PS	packaging signal
Q-FF	Q-Sepharose fast flow
R_h	hydrodynamic radius
RNA	ribonucleic acid
RRV	Ross River virus
S	Svedberg
SARS	severe acute respiratory syndrome
SB	solubilisation buffer
SBMV	southern bean mosaic virus
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography

SEC-MALS	size exclusion chromatography multi angle light scattering
SELEX	systematic evolution of ligands by exponential enrichment
SEM	Standard error of the mean
SFV	Semliki Forest virus
SINV	Sindbis virus
SL	stem loop
smFCS	single molecule fluorescence correlation spectroscopy
SOC	super optimal broth with catabolite repression
SP-FF	SP-Sepharose fast flow
ssRNA	single stranded RNA
ss(+)RNA	single stranded positive sense RNA
STMV	satellite tobacco mosaic virus
STNV	satellite tobacco necrosis virus
svAUC	sedimentation velocity analytical ultracentrifugation
T	triangulation number
TB	terrific broth
TBE	tris-borate EDTA
TBS	tris-buffered saline
TBS-T	tris-buffered saline-tween 20
TBSV	tomato bushy stunt virus
TCV	turnip crinkle virus
TEM	transmission election microscopy
TEMED	N,N,N',N'-tetramethylethylenediamine
TESSy	The European Surveillance System
TFA	trifluoroacetic acid

TMV	tobacco mosaic virus
TNV	tobacco necrosis virus
tr	translational operator
TuYV	turnip yellow virus
UK	United Kingdom
USA	United States of America
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet
VDPV	vaccine derived poliovirus
VEE	Venezuelan equine encephalitis
VEEV	Venezuelan equine encephalitis virus
VLP	virus like particle
WEE	Western equine encephalitis
WEEV	Western equine encephalitis virus
WHO	World Health Organization
WT	wild type
w/v	weight/volume
w/w	weight/weight
α2M	α,2-macroglobulin

[This page has been left intentionally blank]

Chapter 1: Introduction

[This page has been left intentionally blank]

1.1 General introduction to Virology

1.1.1 Why is virology important?

According to the World Health Organisation (WHO) global health estimates, 5,491,338 people died in 2016 of infectious and parasitic diseases, which accounts for just under 10% of all deaths recorded in that year (World Health Organization, 2016). Of the >1,400 species of human pathogen, comprising viruses (including prions), bacteria, fungi, protozoa and helminths (Woolhouse and Gowtage-Sequeria, 2005), 1,340,615 of these deaths were attributed to just nine viruses. These nine viruses represent some of the best-known human pathogens. The three with the highest mortality being Hepatitis B Virus (HBV), measles and HIV (World Health Organization, 2016). At the end of 2017 there were an estimated 36.9 million people living with HIV worldwide, 1.8 million of these became infected in that year alone. The virus has claimed over 35 million lives so far, however antiretroviral therapy (ART) is estimated to have have saved 11.4 million lives in the past decade (World Health Organization, 2018).

There are an estimated 219 human viruses recognised by the International Committee on the Taxonomy of Viruses (ICTV), found in over 23 viral families (Woolhouse *et al*., 2012). However, they are not distributed equally. Over 50% of human pathogens are found in just four viral families *Togaviridae, Flaviviridae, Bunyaviridae* and *Reoviridae.* Additionally, pathogens described as ‘emerging’ or ‘re-emerging’ are predominantly viruses, with RNA viruses alone accounting for 37% of this group (Woolhouse and Gowtage-Sequeria, 2005).

Although the burden of viruses is typically measured in human morbidity and mortality, viruses infect all kingdoms of life (Prangishvili *et al*., 2006; Clokie *et al*., 2014; Wang, 1991; Van Etten, 2011; Hollings and Stone, 1971; Woolhouse *et al*., 2012) and have significant economic burdens outside of human health. There are an estimated 1,200 species of viruses which infect plants (King *et al*., 2012), and the global cost of crop losses due to viral infections has been estimated at over $30 billion per year. Many of these crops are food staples including potatoes (Potato Leafroll Polerovirus), and cereals (Barley Yellow Dwarf Luteovirus), which cost £30-50 million and £10 million per year in the UK respectively (Nicaise, 2014). Additionally, viruses of livestock also carry a high cost. The Department for Environment, Food and Rural Affairs (DEFRA) estimates that swine flu (H1N1) and bovine viral diarrhoea/mucosal disease (BVD) have associated costs of £8 million and £40 million per year (Department for Environment, 2002). A notable example which illustrates the economic impact of animal virus infections is the 2001 Foot and Mouth Disease Virus (FMDV) epidemic. The outbreak cost the public sector over £3 billion, including compensation for farmers, and the private sector over £5 billion (National Audit Office, 2002).

1.1.2 What is a virus?

The simplest virions are assembled from identical, or a limited number of, protein subunit complexes which encompass and protect one or more nucleic acids. These protein subunits can be arranged around the genome in two ways, as shown in Figure 1.1. Proteins arranged in offset discs form a helical tube, or proteins can form icosahedral surface lattices to surround the

DNA/RNA (Crick and Watson, 1956). Additionally, both helical and icosahedral virus particles can be surrounded by a lipid envelope, derived from the host cell membrane, containing viral glycoproteins. Regardless of structural variations, the primary role of virions/capsids remains the same, to deliver the viral genome to a host cell (Prasad and Schmid, 2012) and protect it from chemical or physical damage (Cann, 2016).

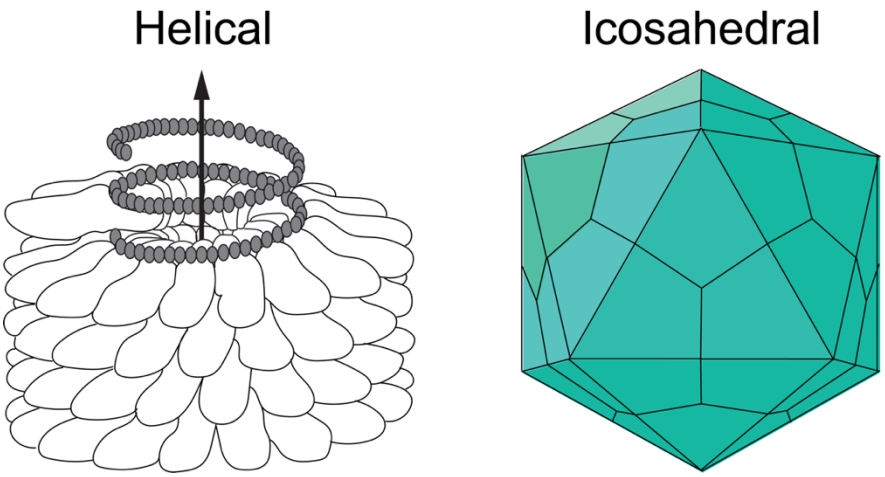


Figure 1.1: Viral capsid structure, helical and icosahedral. Geometric models of helical and icosahedral viral capsid structures. On the left, helical virus which consists of multiple protein subunits forming a tube-like structure around the viral RNA. Icosahedral capsids, associate with 20-sided symmetry to contain the viral genome. Adapted from Caspar and Klug 1962.

1.1.3 Viral Classification

Viruses can be grouped by a number of metrics including virion morphology, method of replication, host range, tropism, genomic arrangement and genome chemistry (DNA or RNA). In 1971, David Baltimore proposed a method of viral classification centred around the synthesis of messenger RNA (mRNA), a task all viruses must complete for the translation of proteins (Baltimore, 1971). Originally containing 6 groups, later revised to 7, viruses are classified by virtue of their genome chemistry and mRNA synthesis (Figure 1.2).

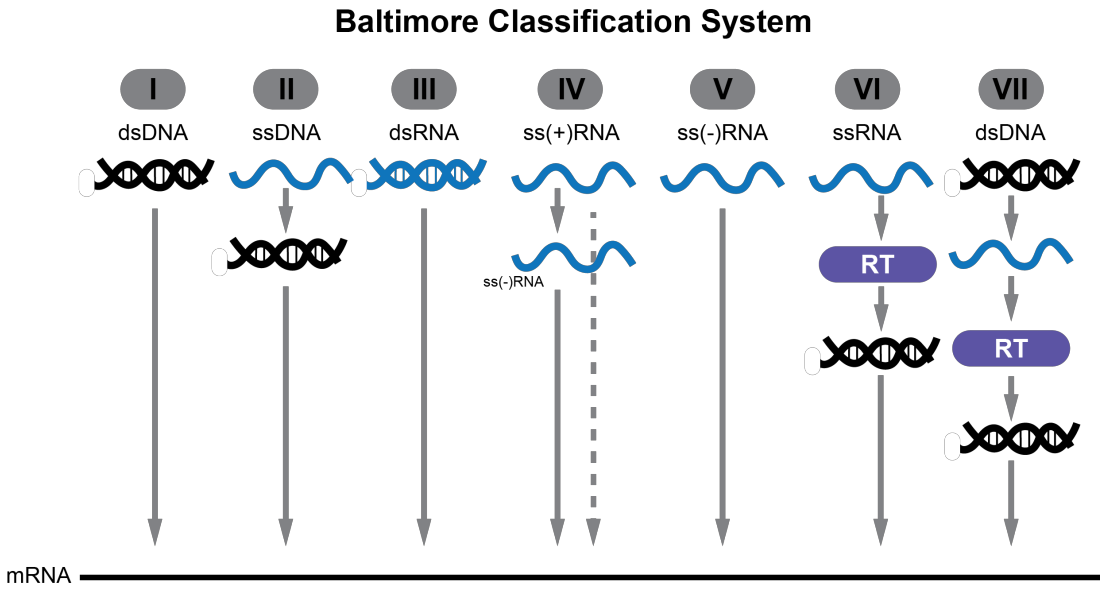


Figure 1.2: Baltimore classification of viruses. Illustration depicting the Baltimore classification system where viruses are ordered into 7 categories based on mechanism of RNA synthesis and genome chemistry (DNA or RNA).

1.1.4 Group IV viruses

The process by which group IV viruses, single-stranded positive-sense RNA (ss(+)RNA) viruses replicate is the simplest of all viral replication mechanisms. In this instance, viral genetic material can be used as mRNA and directly translated into protein (Sun *et al*., 2010). The size of RNA genomes is constrained by the nature of RNA in comparison to DNA, i.e. it is more easily hydrolysed, especially when single stranded (Voet and Voet, 2010). This, coupled with mutation rates due to lack of proofreading encourages small genomes among RNA viruses (Cann, 2012). Ss(+)RNA viruses are the largest viral group and contain some of the most well-known human and animal pathogens (Francki *et al*., 1991). Polioviruses, rhinoviruses, and coronaviruses, including severe acute respiratory syndrome (SARS), are all examples of ss(+)RNA viruses that act as human pathogens (Haaheim *et al*., 2002).

1.2 Principles of Viral Structure

Despite differences in architecture, all viruses share a universal trait in the construction of a protein capsid to contain and protect their genetic material. In 1950, Dorothy Hodgkin observed cubic symmetry in tomato bushy stunt virus (TBSV) and turnip yellow virus (TuYV) crystals. She postulated that the roughly spherical particles were formed from "12*n* identical submolecules" and suggested a potential cubo-octahedral shape which would seem spherical (Hodgkin, 1950). Crick and Watson then took these observations and hypothesised that all small virus capsids were formed from a regular arrangement of multiple identical protein subunits. Helical viruses such as tobacco mosaic virus (TMV) have a screw axis of symmetry where all the protein subunits contact neighbouring subunits at the same points i.e. all subunits are equivalent (Crick and Watson, 1956).

In the same 1956 volume of Nature, Caspar presented evidence of cubic symmetry in TBSV particles. X-ray crystallography indicated that the particles had 532 point-group symmetry, meaning that they have 5-fold, 2-fold and 3-fold symmetry axes. He furthered this by suggesting that TBSV was constructed of a multiple of 12, and likely a multiple of 60, subunits (Caspar, 1956). Given Caspar's crystallographic description, Crick and Watson were able to assign the possible platonic solids capable of forming particles with 532 point-groups. Of the five platonic solids, only the dodecahedron and the icosahedron have 5-, 2- and 3-fold symmetry axes and additionally, 60 asymmetric units (Crick and Watson, 1956). Since these initial investigations,

technological advances have allowed for the study of capsid structure in finer detail which has validated the initial findings of icosahedral viral capsid organisation (Prasad and Schmid, 2012).

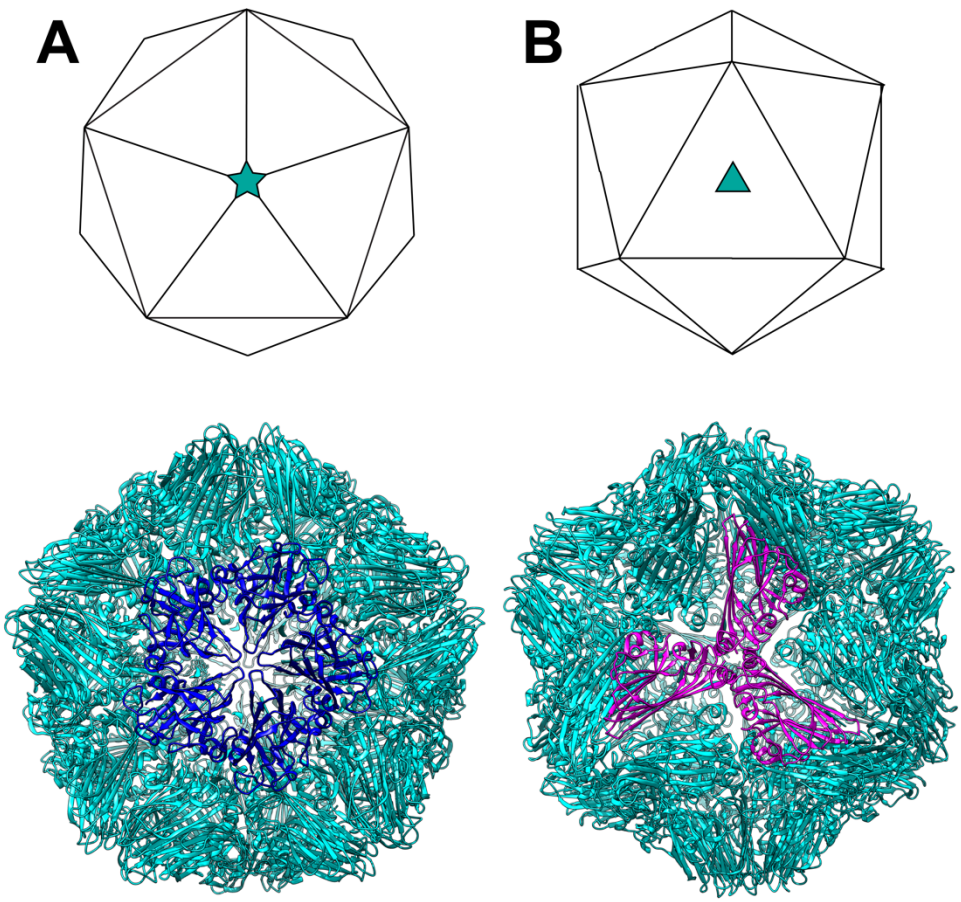


Figure 1.3: 5-fold and 3-fold symmetry axes of an icosahedron. (A) Schematic representation of 5-fold rotational symmetry axis of an icosahedron with 5-fold symmetry axis of satellite tobacco necrosis virus (STNV) (PDBID 4V4M) below (Blue). **(B)** Schematic representation of 3-fold symmetry axis of an icosahedron, with 3-fold symmetry axis of STNV (PDBID 4V4M) below (Magenta).

1.2.1 Building larger virus capsids

For helical viruses, larger capsids can be constructed by the repetition of subunits along the direction of the screw axis (Caspar and Klug, 1962). However, this is not the case for icosahedral viruses; larger viral capsids are not formed using larger coat proteins. The inherent symmetry of viruses is a

result of genetic economy. The information required to code for the structural elements of the virus must not go beyond the coding capacity of the genome. Consequently, there is often only one coat protein gene, so capsids are composed of chemically identical repeating subunits (Crick and Watson, 1956). Due to the size limit imposed by the internal volume of the capsid, viruses must efficiently use their genome (i.e. use overlapping reading frames) to maximise coding capacity. Production of larger coat proteins would require more genome space be given over to the coat protein gene. Each amino acid is encoded by 3 bases with an approximate molecular weight of 1000 Da. Given the average weight of an amino acid is 150 Da, a nucleic acid sequence can only encode a protein <15% of its own weight (Cann, 2016). The goal of a larger capsid is the ability to package a larger genome and production of larger coat proteins does not increase net genome space.

1.2.2 Quasi-equivalence and triangulation

When an asymmetric object is placed on one surface of an icosahedron, governed by the rules of symmetry and allowing for rotation, 59 others are created as seen in Figure 1.4. The smallest icosahedral viruses are built from 60 of such subunits, with all 20 faces being occupied by 3 equivalent proteins. However, it is not possible for an icosahedron to accommodate more than 60 identical subunits placed equivalently. But as Caspar and Klug noted “molecular structures are not built to conform to exact mathematical concepts but, rather, to satisfy the condition that the system be in a minimum energy configuration” (Caspar and Klug, 1962).

Inspired by Fuller's geodesic domes Caspar and Klug investigated triangulation of spheres as a possible solution. If each of the 20 faces is subdivided into four equal triangles and a subunit placed at each vertex 240 subunits can now be accommodated. As seen in Figure 1.4B these subunits form pentamers around the original 5-fold axes and hexamers at the new vertices. As not all the subunits are in the same environment making identical contacts they are not equivalent, but with minimal digression from equivalence the arrangement results in a minimum energy structure termed quasi-equivalence (Caspar and Klug, 1962).

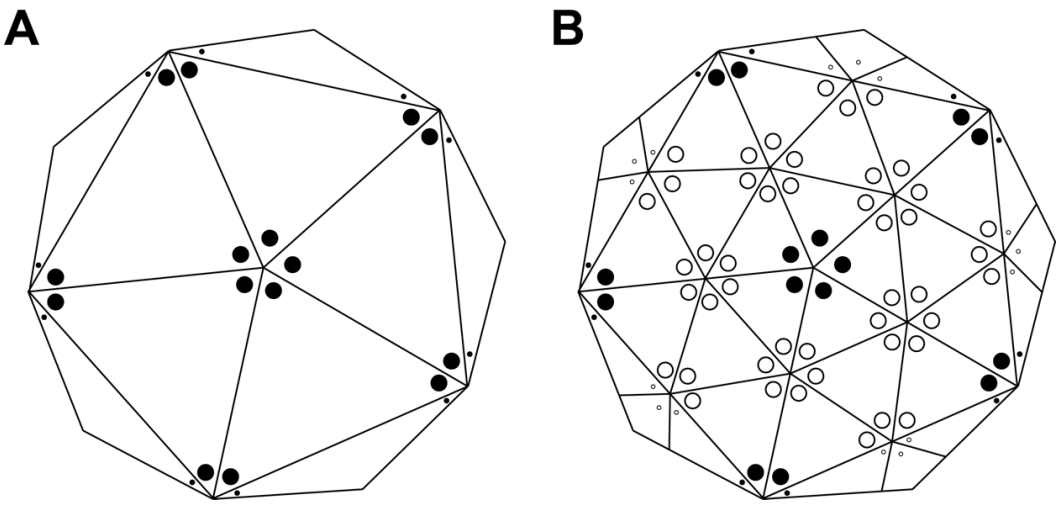


Figure 1.4: Arrangement of subunits on an icosahedron. (A) 60 subunits arranged on an icosahedron, as all the subunits have the same arrangements of neighbours they are in equivalent environments. **(B)** Arrangement of subunits where *n*=4. By dividing each face into 4 equilateral triangles with a subunit at each vertex 240 can be accommodated. Unlike in A, the subunits in B are arranged in both pentamers and hexamers and therefore they are quasi-equivalent.

Triangulation is governed by the equation $T=H^2+HK+K^2$, where T= triangulation number and H and K are positive integers or 0. How triangulation allows for capsids composed of more than 60 subunits can be demonstrated using a hexagonal lattice with H and K axes crossing at the 60˚ angle, as shown in Figure 1.5. Any point on the lattice is chosen and considered the location of a 5-fold icosahedral vertex, designated 0,0. Using the aforementioned equation it is possible to plot the location of the next 5-fold icosahedral vertex. A T=1 capsid will have a H or K value of 1, the other value being 0. Moving graphically with this vector in the direction of H or K (for T=1; H=1 and K=0) this point on the graph represents the next 5-fold vertex. An equilateral triangle, where the length of each side is equal to the distance between H, K and 0,0 equals one of the faces of an icosahedron with each corner representing a 5-fold vertex (shown in Figure 1.5 as a Black triangle). This is true of all icosahedrons, a T=3 capsid will have H and K values of 1, so one is moved on both H and K axes giving the face of a T=3 capsid (Red triangle).

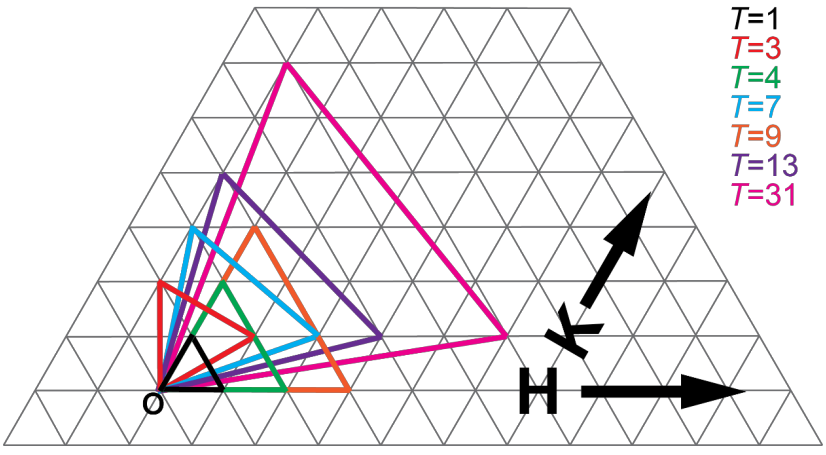


Figure 1.5: Triangulation number. Hexagonal net for the explaination of triangulation. The face of an icosahedron is given by an equilateral triangle where the length of each side is equal to the distance between the origin (0,0) and H, K. Facet of a T=1 icosahedron (Black) has the unit vector 1,0. The faces of sucessively larger icosahedrons are indicated in the legend.

1.3 Solving Viral Structures

Although identified in the 19th century as filterable agents (Ivanowski, 1892; Beijerinck, 1891; Lustig and Levine, 1992) investigations of viral structure did not begin until 1935 with Stanley's description of crystalline material from purified TMV (Stanley, 1935). In 1941, this was taken further by Bernal and Fankuchen, who, using X-ray diffraction were able to estimate the size and shape of TMV and TBSV (Bernal and Fankuchen, 1941a; Bernal and Fankuchen, 1941b). As discussed previously, X-ray diffraction studies allowed for greater inference on the structure of virus particles (Hodgkin, 1950; Crick and Watson, 1956; Caspar and Klug, 1962) however, at this point the first protein structures had not yet been determined (myoglobin 1958 (Kendrew *et al*., 1958) and hemoglobin 1960 (Perutz *et al*., 1960)). It was not until 1975 that a near atomic resolution structure of TMV was published (Holmes *et al*., 1975) and only in 1978 (TBSV) (Harrison *et al*., 1978) and 1980 (southern bean mosaic virus, SBMV) (Abad-Zapatero *et al*., 1980) that the first near atomic resolution structures of icosahedral viruses were released. By the end of the 1980s another first was achieved, in the visualisation of RNA within a virus capsid (Chen *et al*., 1989).

Development of the electron microscope (EM) began in the early 1930s (Mulvey, 2002) and in 1939 EM was first used to visualise a virus *(*Kausche *et al*., 1939; Goldsmith and Miller, 2009) progressing to the first 3D reconstructions in 1970 (Crowther *et al*., 1970). With the advent of cryo-electron microscopy (cryoEM) and flash freezing at liquid nitrogen

temperatures (Adrian *et al*., 1984; Dubochet *et al*., 1988), resolution to rival crystallography is often achieved (Cheng, 2015).

Due to the repetitive nature of symmetrical molecules resolution can be increased by taking advantage of redundancy. When there are multiple, similar subunits in an asymmetric unit they are often related by non-crystallographic symmetry (NCS). By rotation, these subunits can be superimposed onto each other and new structural information can be obtained which can improve accuracy, this is non-crystallographic symmetry averaging (NCSA). For example, an icosahedron has 12 5-fold axes and NCSA can be used to average across all these axes improving resolution (Figure 1.6). Additionally, if a virus forms the whole crystallographic asymmetric unit, 60-fold redundancy is seen (Rossmann and Blow, 1962; Blow, 2006).

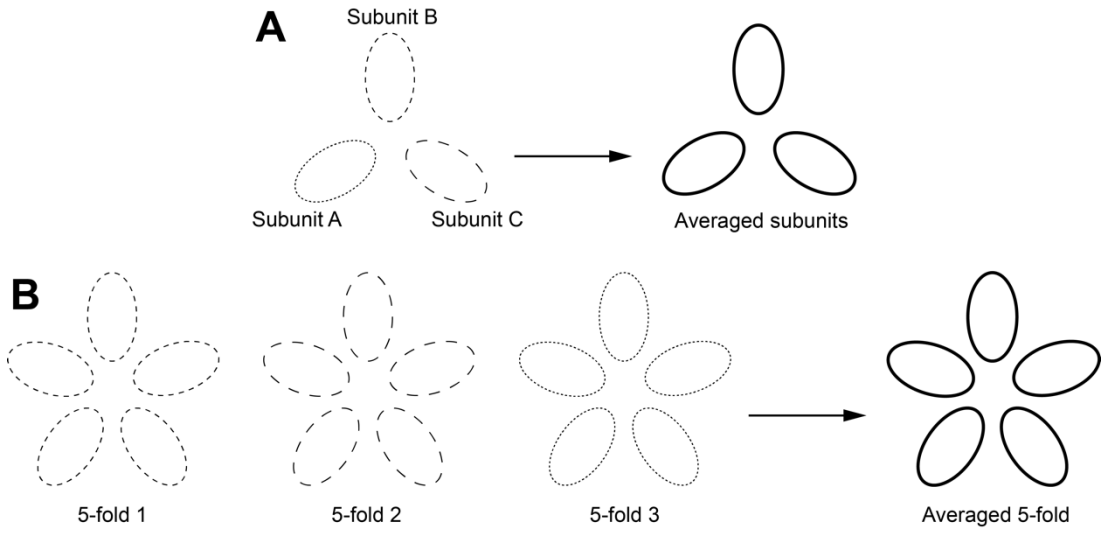


Figure 1.6: Non-crystallographic symmetry averaging. When there is more than one similar subunit in an asymmetric unit, this gives rise to non-crystallographic symmetry. By superimposing subunits and averaging the accuracy of structural results can be improved. **(A)** Subunits A, B and C in an asymmetric unit are superimposed and averaged to get improved structures. **(B)** 5-fold axes 1, 2 and 3 are superimposed and averaged to improve the accuracy of a 5-fold axes structure.

Even with cryo-EM structures symmetry averaging does occur. However as with x-ray crystallography, this can lead to oversimplification of capsid structures. An icosahedron is symmetric, but perfect symmetry may not be ideal for biological activity. Some viruses have asymmetric features such as portals (McElwee *et al*., 2018), receptor binding sites (Hafenstein *et al*., 2007) or maturation proteins (Dai *et al*., 2017; Gorzelnik *et al*., 2016). A recent asymmetric reconstruction of bacteriophage MS2 (Dai *et al*., 2017) allows for the visualisation of a large unique surface feature, the maturation protein (A protein) which occupies the position of a CC dimer, but also the encapsidated RNA. Previous symmetry averaged structures impose a dimer in the place of the maturation protein (Valegård *et al*., 1990) generating structural artefacts, so while averaging improves resolution, it comes at a cost.

1.4 Visualising viral genomes *in situ*

In 1989 X-ray crystallography revealed density inside bean pod mottle virus (BPMV) capsids not connected to the polypeptide chain. Seven ribonucleotides were modelled into the density which was strongest at the 5' end, in close proximity to the coat proteins, and weakest at the 3' towards the centre of the particle. Since this structure was refined using NCSA, the presence of this density implies that RNA is present at most of the 60 positions and accounts for ~20% of the packaged RNA (Chen *et al*., 1989). More recent analysis of the density suggests that the nucleotide sequence is non-random and defines the principal sequence as A-Pu-Py-Py-Py-X, where A denotes adenosine, Pu denotes a purine, Py denotes a pyrimidine and X denotes any

nucleotide. Due to the NCSA, the presence of a consensus suggests that this recurring sequence is distributed throughout the genome and could act as a signal for packaging and assembly (Lin *et al*., 2003). Bioinformatic analysis has identified 200 sites that fit this consensus in BPMV RNA2 (Schneemann, 2006). When a further 5 nucleotides were modelled into the density, trefoils formed around the 3-fold axes (Figure 1.7) (Lin *et al*., 2003).

In 1993, a structure of the plant satellite virus satellite tobacco mosaic virus (STMV) showed 44% of the viral genome in 30, 7 bp, double-stranded helical segments associated with coat protein dimers located at the 2-fold axes (Figure 1.7) (Larson *et al*., 1993). Refinement indicated the helices comprised of at least 9 base pairs (Larson *et al*., 1998). Similarly, in STNV, neutron scattering showed RNA at low resolution, elucidated as double-helices along the icosahedral edges (Bentley *et al*., 1987), which is similar to that seen in STMV. RNA density has also been described for the insect virus, pariacoto virus (PaV), from the *Nodaviridae* family. The X-ray crystal structure shows 30, 24 nt long, RNA duplexes inside the virus capsid forming a dodecahedral cage accounting for 35% of the viral genome (Figure 1.7) (Tang *et al*., 2001). This dodecahedral arrangement is also seen in cryo-EM structures of the comovirus cowpea mosaic virus (CPMV) (Hesketh *et al*., 2015). However, these structures were all averaged and therefore may not reflect true genome organisation.

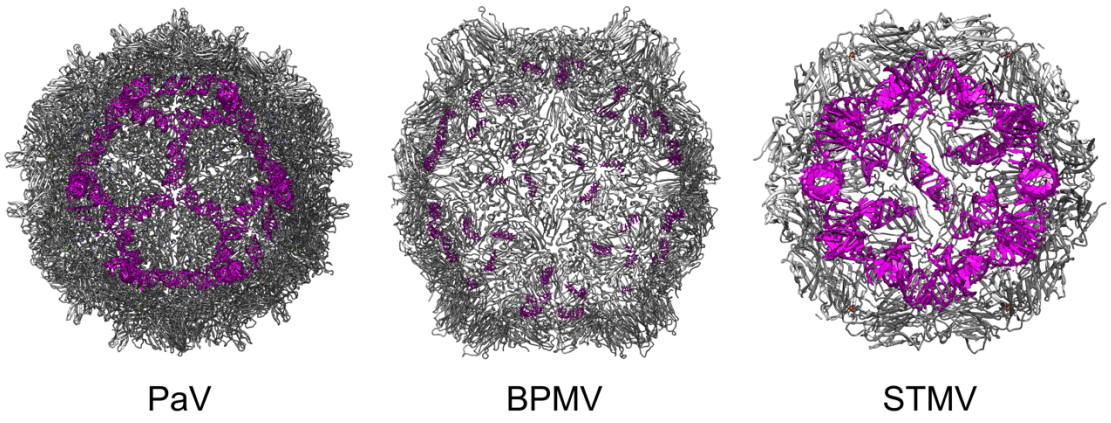


Figure 1.7: Visualisation of RNA inside viral capsids. Cut-away view of PaV, BPMV, and STMV with internal RNA. Models were accessed from the protein data bank at 1F8V, 1PGL and 1A34 respectively and created using ChimeraX (Goddard *et al*., 2018).

More recent asymmetric structures of the phage MS2, indicate that the genetic material contained within the viral capsid is highly ordered, with stem loop structures contacting the inside of the capsid shell. The two models, from 2016 and 2017 assigned 95 and ~80% of the genome respectively and indicate that the genomic density is not shared equally across the capsid but rather in a non-uniform distribution with the majority being around the maturation protein (Figure 1.8A). These observations support the theory that in MS2 assembly stem loops (SLs) interact with CP dimers inducing a compaction of RNA prior to complete capsid assembly (Dai *et al*., 2017; Koning *et al*., 2016). Unlike with previous symmetry averaged RNA densities, the RNA forms defined structures which does not conform to icosahedral symmetry and extends beyond the peripheral contacts with the protein shell to the interior of the capsid.

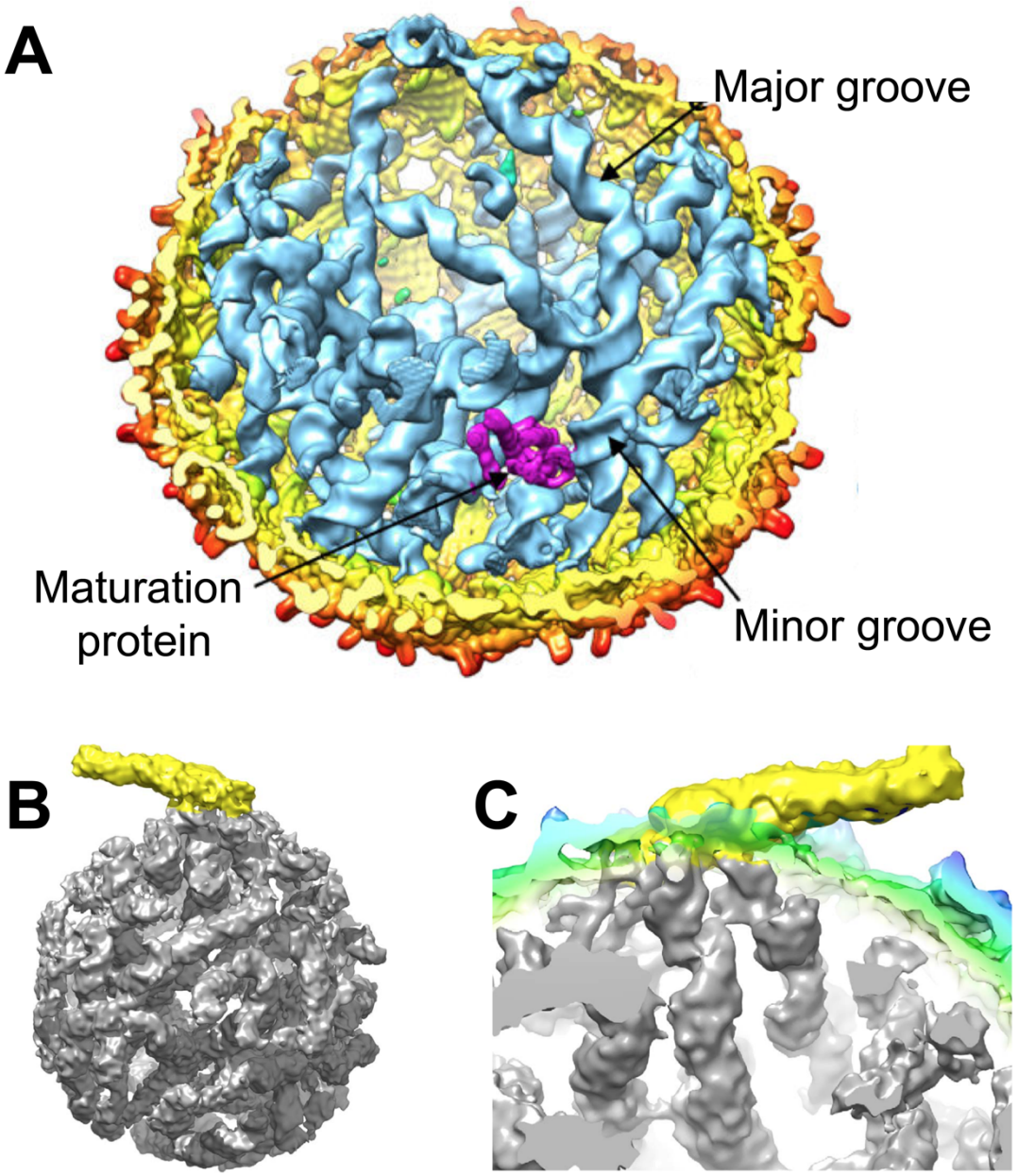


Figure 1.8: Asymmetric reconstructions of MS2 showing ordered RNA. (A) Cut away of MS2 capsid showing genome density and maturation protein taken from Dai *et al*., 2017. **(B)** Structured genome in contact with the maturation protein (A protein) of MS2. **(C)** Genome in context of the protein shell and the maturation protein. B and C taken from Koning *et al*., 2016.

1.5 Satellite Tobacco Necrosis Virus

Satellite tobacco necrosis virus (STNV) is one of the smallest viruses known, with a mean diameter of 170 Å (~17 nm). As a satellite virus, STNV requires co-infection of its helper tobacco necrosis virus (TNV), which provides replication machinery *in trans* (Unge *et al*., 1980). There are three distinct STNV strains (STNV-1, -2 and -C) all of which are serologically unrelated to the helper virus and have little sequence similarity to each other excepting a region of 100-150 nucleotides at the 3′ end of the genome thought to be the binding site of TNV replicase (Rees *et al*., 1970). Replication of STNV-1 and -2 is assisted by TNV strain A (TNV-A) whereas STNV-C undergoes replication with the assistance of TNV strain D (TNV-D). The strains also differ in the size of their 5′ untranslated regions (UTRs). Both STNV-1 and -2 have 5’ UTRs of approximately 30 nt, whereas the 5′ UTR of STNV-C is 101 nt long (Bringloe *et al*., 1998).

The STNV virion is icosahedral in structure, composed of 60 identical protein subunits in a *T*=1 capsid around a single RNA molecule (Figure 1.9A) (Liljas *et al*., 1982). This RNA is 1239 (Ysebaert *et al*., 1980), 1245 (Danthinne *et al*., 1991) or 1221 (Bringloe *et al*., 1998) nucleotides long respectively for strains -1, -2 and -C, and codes for the coat protein, with a molecular mass of 21.54 kDa for STNV-1 (Liljas *et al*., 1982). Complete nucleotide sequences of all three strains have been determined (Bringloe *et al*., 1998) as have the coat protein sequences (Jones and Liljas, 1984).

The coat protein of STNV, like most viral coat proteins adopts one of a limited number of folded conformations. There are eight identified plant satellite viruses, all of which have ss(+)RNA genomes, *T*=1 capsids and protein monomers predicted to be jelly-roll β-barrels (Krupovic *et al*., 2016). The STNV monomer adopts a jelly-roll β-barrel fold, formed from two, four-stranded β-sheets (Liljas *et al*., 1982) with an extended N-terminal arm, another common feature of viral coat proteins (Figure 1.9B). These arms, often unresolved in crystal structures, are thought to be important in the assembly mechanisms of spherical plant viruses (Sorger *et al*., 1986; Annamalai *et al*., 2005) where it has been suggested that these arms interact with the viral genetic material and stabilise CP subunits (Patel *et al*., 2015; Ford, 2012; Ford *et al*., 2013). In the crystal structure the N-terminal arm of STNV, residues 12-24, forms an α-helix whereas residues 1-11 are disordered (Jones and Liljas, 1984).

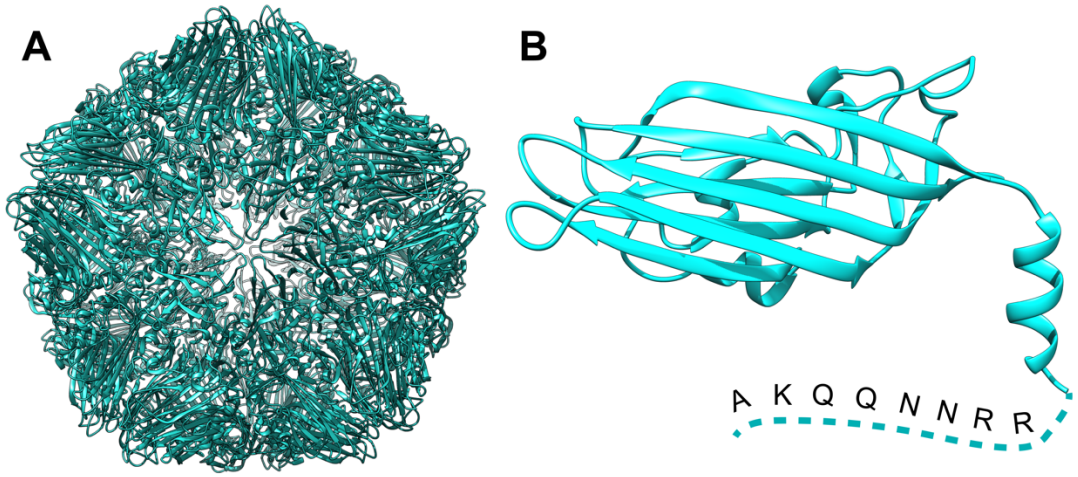


Figure 1.9: Structures of STNV and coat protein. (A) Structure of STNV down a 5-fold axis from (PDBID 4V4M). **(B)** STNV coat protein subunit showing jelly roll structure formed from two, four-stranded β-sheets.

Recombinant STNV CP was first expressed in 1984, in *E.coli*. In these initial investigations, however, CP was largely insoluble, prone to proteolysis and assembly of virus like particles (VLPs) was not observed (van Emmelo *et al*., 1984). In 2011, Lane *et al* created a codon optimised STNV-1 CP gene for expression in *E.coli* which did assemble into *T*=1 VLPs that resembled wild type particles (Lane *et al*., 2011). These *E.coli* derived VLPs have been characterised and used in investigations probing ss(+)RNA viral assembly mechanisms and RNA:protein interactions (Ford, 2012; Borodavka *et al*., 2012; Borodavka *et al*., 2013; Patel *et al*., 2015; Stockley *et al*., 2013).

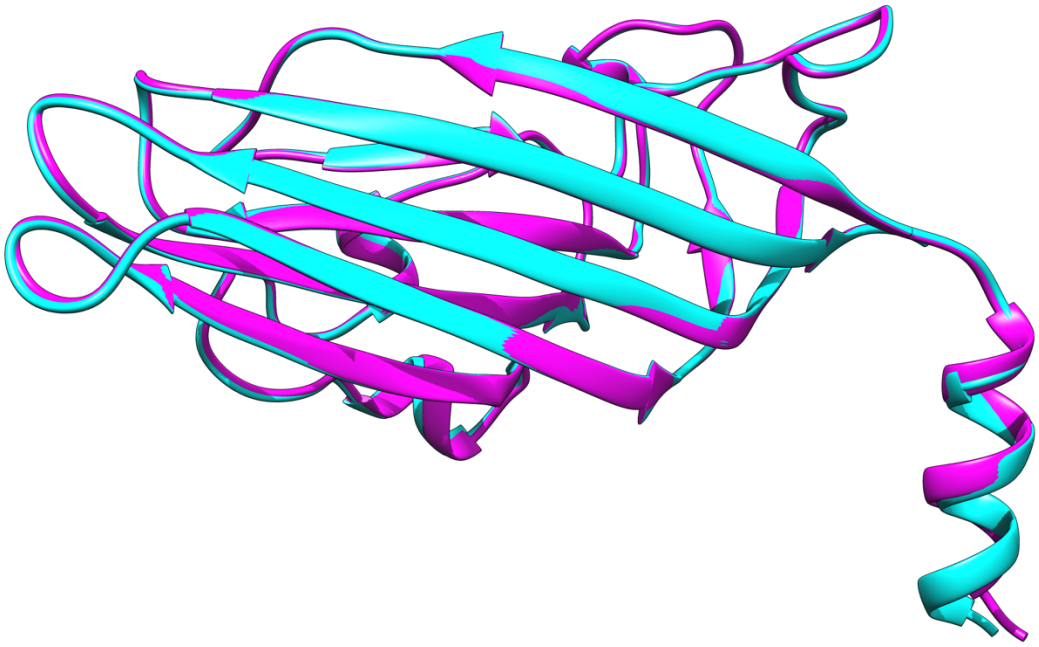


Figure 1.10: Comparison of wild type and recombinant STNV coat protein structures. Recombinant and wild type STNV coat protein structures from PDB, 3SRV (Magenta) and 2BUK (Cyan) respectively, aligned using ChimeraX (Goddard *et al*., 2018). Structures of the two proteins, as noted in Lane *et al*, do not differ significantly (Lane *et al*., 2011).

1.6 Alphaviruses

Alphaviruses are enveloped, spherical, ss(+)RNA viruses which include zoonotic mosquito borne pathogens responsible for disease in both humans and animals. Alphaviruses are globally dispersed and the causative agents of both large human epidemics and localised epizootics and outbreaks. The genus *Alphavirus*, from the family *Togaviridae*, contains over 40 members spread by blood-sucking arthropods which typically cause rash, fever and joint pain (Jose *et al*., 2009). Alphaviruses can be subdivided into seven serocomplexes (Figure 1.11) 4 of which; Eastern Equine Encephalitis (EEE), Western Equine Encephalitis (WEE), Venezuelan Equine Encephalitis (VEE) and Semliki Forest (SF), contain the majority of the medically relevant viruses (Powers and Roehrig, 2010). The most familiar members of this family are spread by mosquitos, Semliki Forest virus spreads by *Aedes* sp. in Asia and Africa; Ross River virus by *Culex* and *Aedes* sp. in Australia; Sindbis virus is transmitted by *Culex* sp. in Europe, Africa, Australia and Asia (Korsman *et al*., 2012); Western equine encephalitis virus by *Culex* sp. in North and South America (Bergren *et al*., 2014) and perhaps the most well-known alphavirus chikungunya, spreads by *Culex* and *Aedes* sp. in Africa, Asia, the Americas, Europe and Australia (Korsman *et al*., 2012; European Centre for Disease Prevention and Control, 2018).

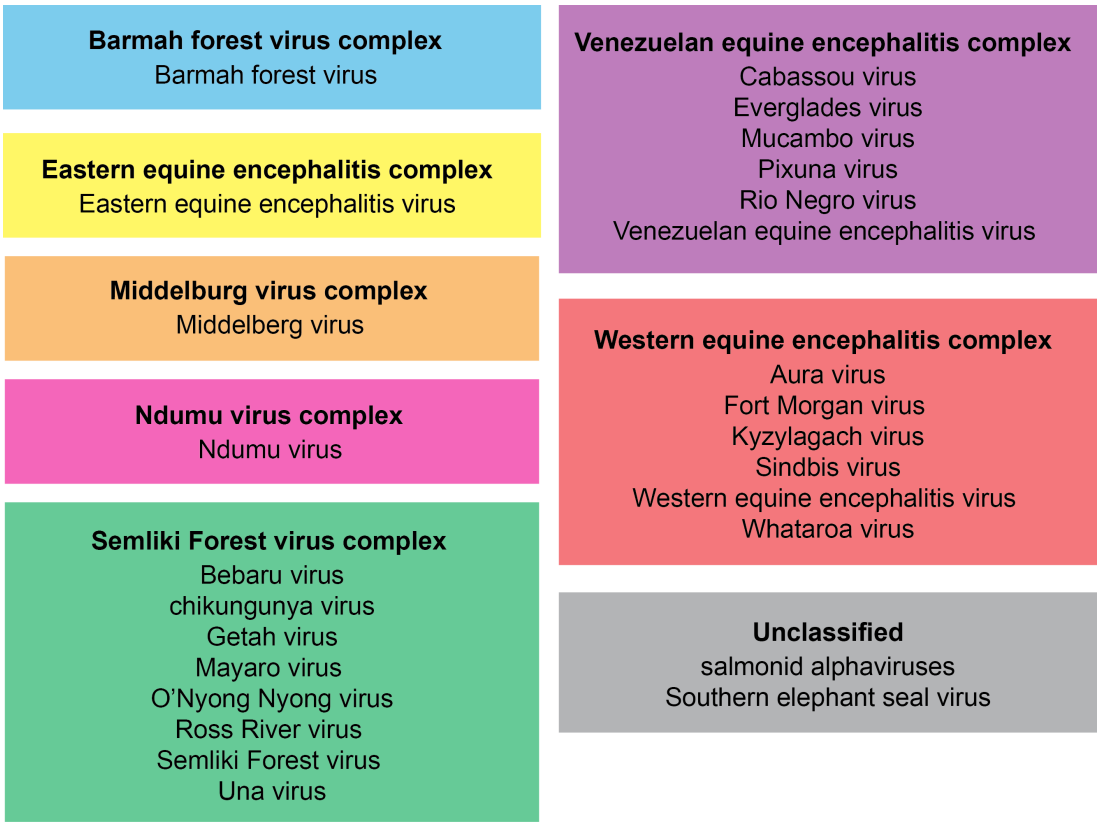


Figure 1.11: Alphavirus serocomplexes. Members of the *Alphavirus* genus divided into antigenic serocomplexes including unclassified members. Four of these serocomplexes, Eastern equine encephalitis, Western equine encephalitis, Venezuelan equine encephalitis and Semliki Forest, contain the majority of the medically relevant viruses.

1.6.1 Alphavirus genome organisation and structure

All seven *Alphavirus* serocomplexes share similar genomic features, having a 11-12 kb ss(+)RNA genome, capped at the 5′ end and with a 3′ polyA tail (Figure 1.12A). This 49S RNA contains two open reading frames (ORFs) coding for the non-structural and structural genes, respectively (Wilks, 2016). Upon entry into the cell, the 49S RNA is translated and produces a polyprotein (p270) coding for the non-structural proteins numbered 1-4 (nsP1-4). Alphaviruses, like other RNA viruses, modify intracellular membranes in the construction of cytopathic vacuoles (CPVs) or replication factories. Type1 CPVs, thought to be the site of alphavirus RNA synthesis are formed from

endosomal membranes and termed spherules. The non-structural proteins form a replication complex facilitating the production of RNA which in turn promotes the formation of these spherules (Strauss and Strauss, 1994).

Genomic RNA is replicated via the production of a negative sense intermediate strand. A separate structural polyprotein, p130, coding for the structural proteins (Cp-E3-E2-6K-E1) is translated from the 26S subgenomic RNA in the late stage of infection after which the core protein (Cp) is proteolytically cleaved from p130 and the remaining structural proteins are trafficked to the endoplasmic reticulum (ER). The Cp gene is positioned at the N-terminus of the structural polyprotein and typically has a mass of ~30 kDa. The alphavirus Cp consists of 2 domains, the C-terminus which contains a chymotrypsin-like serine protease which facilitates cleavage from p130 and forms the pentamers and hexamers from which the nucleocapsid is built, and the N-terminus, which is highly positively charged (Mendes and Kuhn, 2018).

Alphavirus particles are 60-70 nm in diameter and consist of a nucleocapsid (NC), surrounded by a host derived lipid envelope, studded with virally encoded E1 and E2 proteins. The nucleocapsid is formed from 240 copies of the Cp and displays *T*=4 icosahedral architecture (Figure 1.12B). The E1 and E2 proteins are arranged as trimers of heterodimers, located at 80 positions across the lipid envelope (Knipe and Howley, 2013). Various studies have investigated the architecture and assembly mechanisms of alphavirus particles due to their importance as a human pathogen, resulting in multiple structures from both cryo-EM and X-ray crystallography (Wiley and Bonsdorff,

1978; Cheng *et al*., 1995; Mancini *et al*., 2000; Zhang *et al*., 2002; Zhang *et al*., 2005; Mukhopadhyay *et al*., 2006; Zhang *et al*., 2011; Kostyuchenko *et al*., 2011).

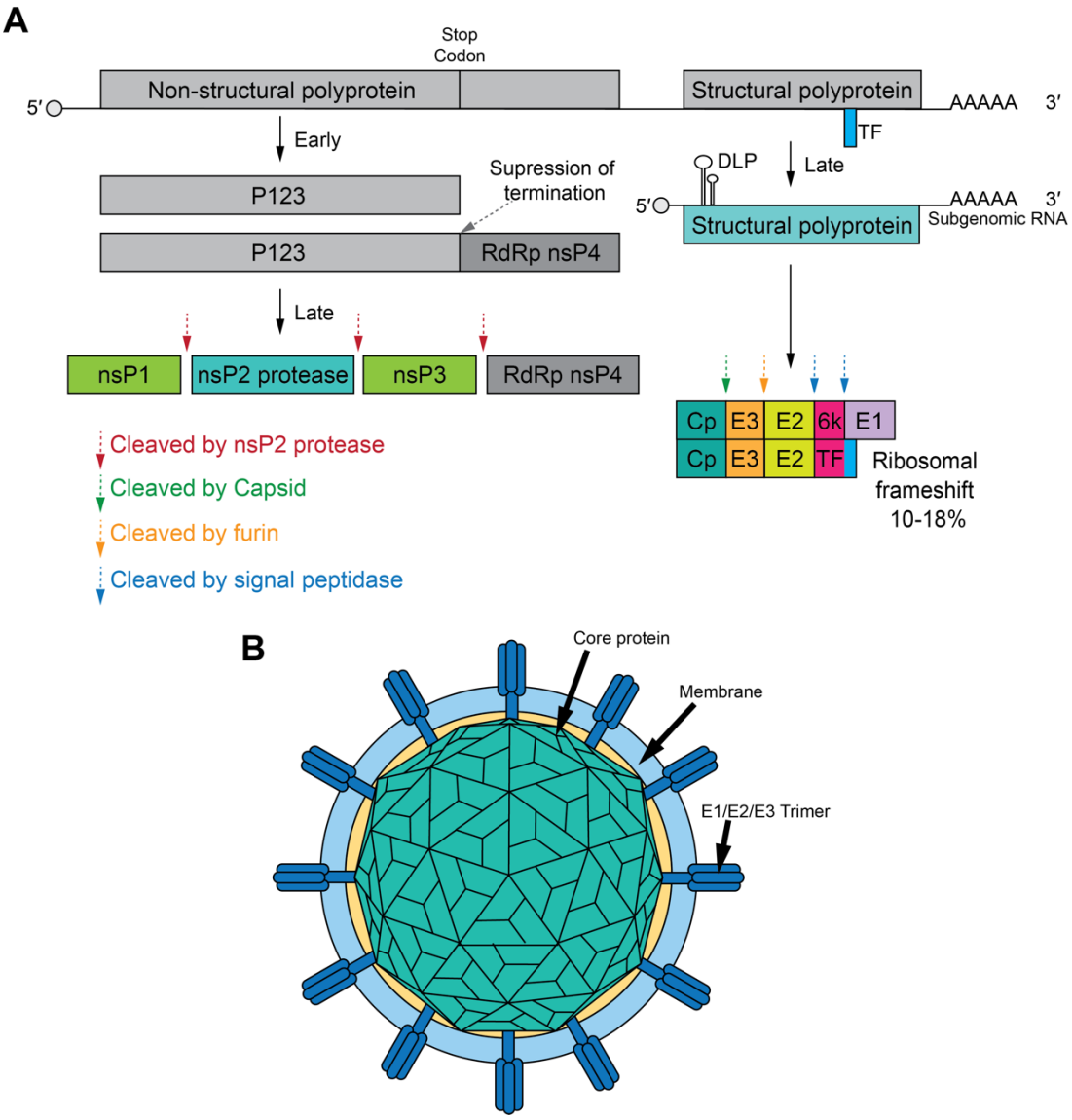


Figure 1.12: Genome organisation and structure of an alphavirus. (A) Organisation of the Alphavirus genome showing coding regions for both the non-structural polyprotein and structural polyprotein. These polyproteins are in turn cleaved to form structural, C – E1, and non-structural proteins nsP1-4. **(B)** External view of the virion structure of an Alphavirus, with the nucleocapsid showing *T*=4 architecture.

1.6.2 Semliki Forest Virus Complex

Semliki Forest virus (SFV), is a commonly viral model, particularly in Europe, as it grows to high titre in both chick embryo and baby hamster kidney (BHK) cells. Human infection is relatively common in Africa where the virus is widely distributed (Briolant *et al*., 2004). SFV was the causative agent of an outbreak of febrile illness among Europeans from October-December 1987 in the Central African Republic (Mathiot *et al*., 1990). One fatal, human case of SFV was found in the literature in an exposed laboratory worker, which resulted in development of meningioencephalomyelitis (Willems *et al*., 1979).

Due to its neurotropism SFV has become of interest as a vector for use in cancer virotherapy, particularly that of the central nervous system (CNS). An oncolytic SFV vector derived from the avirulent strain A7(74) has been shown to eliminate orthotopic human glioma xenografts in nude mice (Ylösmäki *et al*., 2013). Additionally, mice immunised with SFV particles expressing P1A, a tumour rejection antigen, were protected from tumour development and vectors expressing human papilloma virus (HPV) 16 E6-E7 fusion protein had prophylactic and therapeutic responses. SFV vectors elicit a therapeutic effect at a lower dose in comparison to other viral vectors used for cancer immunotherapy investigations. An adenovirus vector expressing HPV 16 E6-E7 is needed in a 100-1000-fold higher dose than the SFV equivalent to achieve the same effect. SFV vectors have been used in at least 12 cancer immunotherapy investigations, the vast majority of which resulted in tumour regression (Lundstrom, 2017).

Ross River virus (RRV) causes ross river fever, a notifiable disease in Australia since 1993, where it is not only the most common alphavirus, but the most common mosquito borne notifiable disease. Between 1993 and 2017 there were on average 4,712 cases of RRV per year peaking at 9,544 cases in 2015. Figure 1.13 shows reports for all mosquito borne notifiable disease in Australia from 1991-2017 (inclusive) as percentages of total reported cases. As seen in the graph, RRV (light green) accounts for the vast majority of cases from 1993 to present averaging 64.87% of all cases and never dropping below 39% (Communicable Diseases Branch, Department of Health, Government of the Commonwealth of Australia, 2018).

Despite the prevalence of this disease, there is no therapy beyond treatment of symptoms and prevention consists of mosquito avoidance. However, vector control programs are estimated to cost over $20 million (AUD) per year and appear ineffectual in the reduction of RRV incidence. In addition to the cost of control measures, there is significant financial burden associated with the disease (Wressnigg *et al*., 2015). Combined healthcare costs and productivity loss has been estimated between $1,018 - $1,180 (AUD) per patient and the cost for Australia as a whole was estimated at $4.3 – $4.9 million (AUD) in 2007 (Yu *et a*l., 2014).

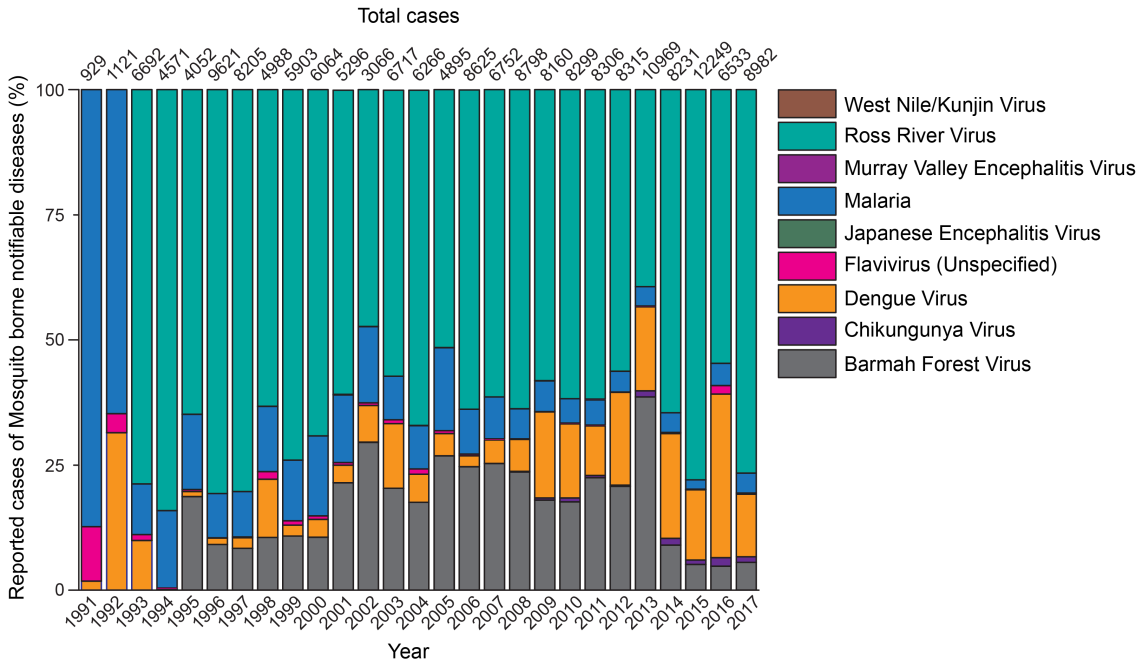


Figure 1.13: Notifiable mosquito borne disease in Australia. Percentage stacked bar chart of notifiable diseases in Australia with a mosquito vector. Values have been normalised to percentage of total cases. Data was accessed from the National Notifiable Disease Surveillance System (NNDSS), (Communicable Diseases Branch, Department of Health, Government of the Commonwealth of Australia, 2018).

Chikungunya (CHIKV) is a re-emerging arbovirus first isolated in Tanzania during an epidemic in 1952-53 (Weaver and Lecuit, 2015). The fatality rate is approximately 1 in 1000 where the most at-risk populations are the elderly, those with underlying conditions and neonates infected at birth. CNS involvement is seen in neonates infected peripartum including haemorrhage and cerebral oedema culminating in long term impairment (Knipe and Howley, 2013).

While typically seen in the context of travellers from regions of endemic infection, since 2007 there has been autochthonous transmission of the virus in Europe. In 2007 there was an outbreak within rural villages in Emilia Romagna, Italy which infected ≈330 people between July and September.

Since this initial outbreak, autochthonous CHIKV has been contained within Italy and France with further outbreaks occurring in 2010, 2014 (France) and 2017 (France and Italy) (European Centre for Disease Prevention and Control, 2017).

While European outbreaks so far have affected a small number of individuals, the spread of CHIKV to the Americas displays a different pattern. At the end of 2013 CHIKV was identified in the Caribbean Islands, and by June 2014 there were 264,444 reported cases. By August 2015, autochthonous transmission was detected in 33 territories and countries in the Americas (Pineda *et al*., 2016). The 2017 CHIKV dataset from the Pan American Health Organization (PAHO) shows that while many countries and territories within the Americas have shown some reduction in cases since the explosive introduction of the virus, the number of reported cases is significantly larger than those after introduction to Europe. The number of reported cases in Brazil in 2017 was 195,962 which is a reduction of almost 65% on the previous year (Pan American Health Organization, 2017). However, this level of infection comes with a significant economic burden not only to the country as a loss of productivity but the out of pocket expenses in some cases are double average monthly incomes for an entire family (Pineda *et al*., 2016). Figure 1.14 shows graphical maps of reported cases of CHIKV in both Europe and the Americas for the first and latest complete data set. Across the data set there is increased range of reported cases within Europe, mostly from imported infections due to travel in endemic areas, with total reported cases remaining below 1500. Within the Americas however, total reported cases for 2017 was 207,914.

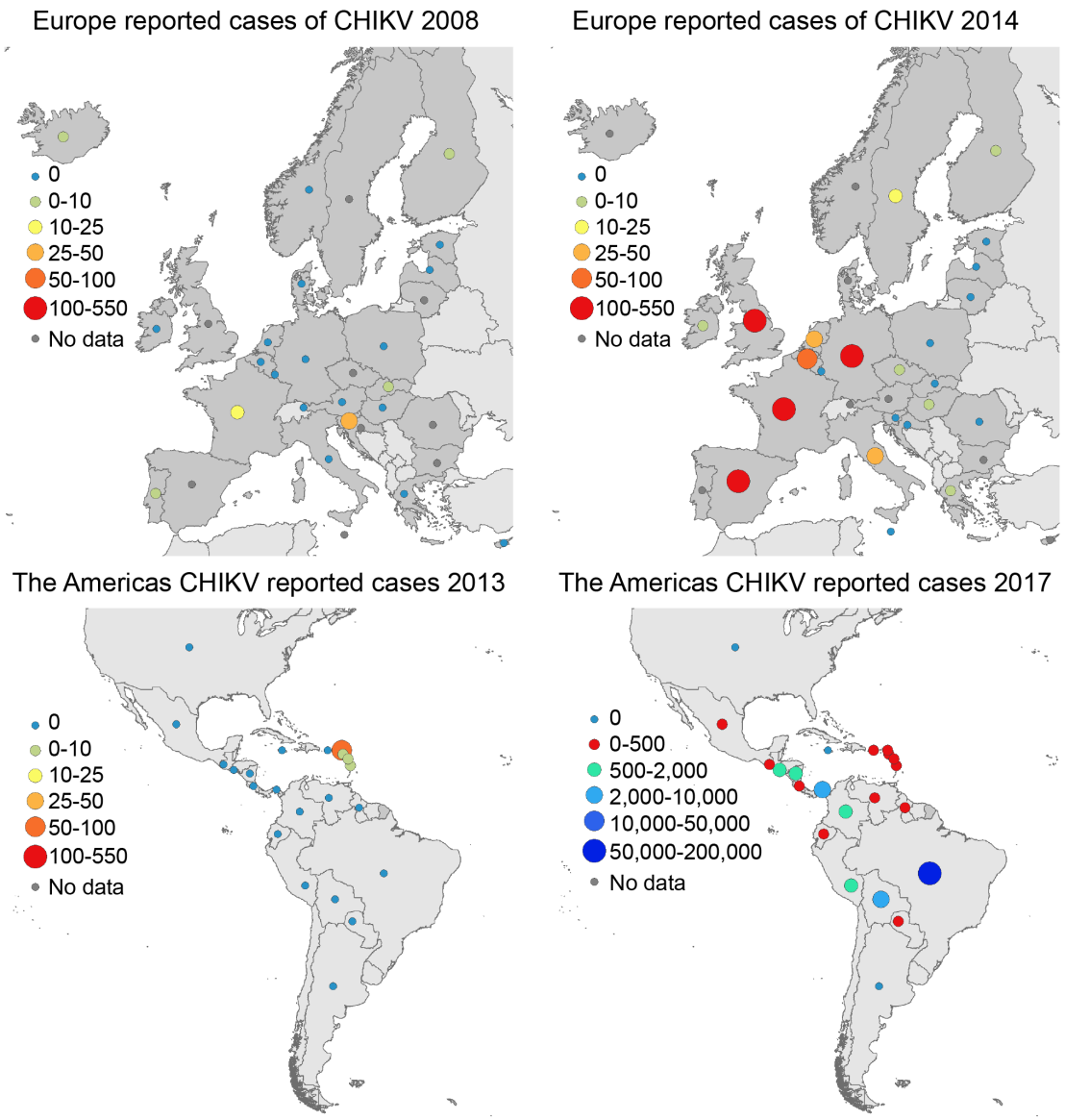


Figure 1.14: Reported cases of chikungunya in Europe and the Americas. Top panel: Maps showing reported cases of CHIKV in Europe for both 2008 and 2014. **Bottom Panel:** Maps showing reported cases of CHIKV in the Americas in both 2013 and 2017. Data publicly accessible and obtained from the ECDC and PAHO. Mapped using ECDC map maker EMMa available through The European Surveillance System (TESSy).

1.6.3 Western Equine Encephalitis Complex

Sindbis virus (SINV), originally isolated from *Culex* mosquitos in Egypt in 1953, is endemic in Europe, Africa, the Middle East, Asia, India, Australia and the Philippines (Knipe and Howley, 2013). Human infections are typically subclinical however, in certain areas SINV infection causes significant arthritic illness (Strauss and Strauss, 2008). In Sweden, disease caused by SINV is known as Ockelbo and was first isolated in 1982 (Shirako *et al*., 1991). This strain of SINV is also responsible for Karelian Fever in Russia and Pogosta disease in Finland (Laine *et al*., 2004) where infections correlate with mosquito activity (Adouchief *et al*., 2016). SINV, like SFV, has historically been used as a model for assembly studies of enveloped viruses (Jose *et al*., 2009) and neuropathogenesis in mice (Atkins, 2013).

Western equine encephalitis virus (WEEV) is a mosquito borne pathogen found in both North and South America and the causative agent of Western equine encephalitis (WEE). Infections in both humans and horses can be fatal with survivors typically suffering permanent neurological complications. In the United States of America (USA) WEEV has been known to cause large outbreaks in the human population extending into Canada throughout the 1930s-1950s. However, incidence has since decreased, in the 1970s there were 209 cases, in the following decade 87 cases were reported, followed by only 4 in the 1990s, with no cases reported since 1998 (Bergren *et al*., 2014).

However, WEEV remains a notifiable disease in the United States and a category B pathogen due to its potential use as a biological weapon. The

closely related Venezuelan equine encephalitis virus (VEEV) was tested as a bioweapon in the 1950s and 60s by the USA and was potentially weaponised by the former Soviet Union (Spurgers and Glass, 2011). WEEV was also researched by the USA as a part of their chemical and biological weapons program which ended in 1969 (James Martin Center for Nonproliferation Studies, 2018). One of the reasons WEEV is a favourable candidate as a biological weapon is its infectivity as an aerosol. There have been three laboratory accidents in which staff were exposed to the virus and two fatalities from probable aerosol exposure. Intranasal inoculation of the virus causes lethal encephalitis in Rhesus macaques as does aerosol exposure in Cynomolgus macaques, and in both cases, virus was not detected in the blood but only in the tissues of the central nervous system (Reed *et al*., 2005).

1.7 RNA:Protein interactions

RNA has multiple functions within a cell and interactions between proteins and RNA are essential for a number of biological processes. RNA is a component of spliceosomes (Lührmann *et al*., 1990) and ribosomes (Littlefield *et al*., 1955) and can act as a catalyst (Scott and Klug, 1996; Scott, 1998; Tarasow and Eaton, 1998). RNA also has multiple functions within a virus. When a ss(+)RNA virus enters a cell there are a number of tasks it must achieve. It must disassemble, replicate and translate its genome and build protective capsids. Many of these functions are mutually exclusive and it is thought that viruses switch between these processes by conformational changes (Simon and Gehrke, 2009). RNA can present many different types of secondary and

tertiary structures including helices, pseudoknots, bulges and stem loops. It is at these structural elements and regions of non-Watson-Crick base pairing, where proteins and RNA tend to interact (Nagai, 1996).

A well-characterised example of a RNA:Protein interaction is that of the MS2 translational repressor (TR) complex (Talbot *et al*., 1990). MS2 has a 3,569 nt ss(+)RNA genome and was the first to be sequenced completely (Fiers *et al*., 1976). Protein dimers, in two distinct conformations, A/B and C/C form the *T*=3 capsid (Valegård *et al*., 1994), along with one copy of the maturation protein (Dent *et al*., 2013). TR is a stem loop, 19 nt in length which contains the replicase gene initiation codon and can trigger capsid assembly at low concentrations (Valegård *et al*., 1994). During infection, as CP concentration increases, TR binds sequence-specifically to a CP dimer, which downregulates the viral replicase and therefore replication (Witherell *et al*., 1991). TR binding to CP dimers mediates conformer switching from symmetric C/C to asymmetric A/B dimers, both of which are necessary components of the T=3 capsid (Stockley *et al*., 2007; Basnak *et al*., 2010). This protein:RNA interaction is highly sequence-specific as TR has been shown to discriminate between closely related mutant proteins (Fouts *et al*., 1997; Rowsell *et al*., 1998; Hirao *et al*., 1998).

1.8 RNA Virus Assembly

The historical view of ssRNA virus assembly concludes that genome packaging is driven by electrostatic interactions, with the RNA playing no active role (Bancroft *et al*., 1969; Garmann *et al*., 2014; van der Schoot and Bruinsma, 2005; Parent *et al*., 2005). However, the electrostatic only mechanism does not explain how, in an environment that is occupied by multiple competing RNAs, viruses preferentially package their own genomes. This pathway has led to the idea that only nascent RNA is encased, or that viral RNA and coat proteins are segregated from cellular components in specialised compartments (Dykeman, Stockley and Twarock, 2013b). The strategy of self-assembly is the simplest of the proposed mechanisms as it does not involve assistance from nucleic acids, scaffold proteins or other macromolecules, it is the spontaneous association of protein subunits into a closed capsid structure. Self-assembly and assembly with non-cognate nucleic acids leads to the conclusion that production of a protein shell is an intrinsic characteristic of viral coat proteins just as folding is dictated by amino acid sequence. Indeed, *in vitro* assembly investigations in the literature often show outcomes not observed in nature, with coat proteins assembling VLPs around anionic polymers (Bancroft *et al*., 1967; Garmann *et al*., 2015; Hu *et al*., 2008), nanoparticle templates (Chen *et al*., 2006) and host RNAs (Routh, Domitrovic and Johnson, 2012a; Routh, Domitrovic and Johnson, 2012b).

Protein-centric models are applicable to the assembly of double stranded (ds) DNA viruses where the genome is packaged into a pre-formed container using an adenosine triphosphate (ATP) driven motor. Due to the nature of ds nucleic

acids, high charge density and relative stiffness, encapsidation of genetic material is not linked with the formation of a capsid without the aid of auxiliary proteins e.g. nucleic acid folding proteins. However, infectious ssRNA virus particles can be produced *in vitro* from viral proteins and RNA, as seen with TMV in 1955 by Fraenkel-Conrat and Williams (Fraenkel-Conrat and Williams, 1955) indicating that these components alone are sufficient for complete viral assembly.

The specificity of viral genome packaging, especially in a cellular environment, suggests that there is an element of selectivity in this process. One investigation found that Flock House virus, a prototypic nodavirus, is ~99% selective for cognate RNA by next generation sequencing (NGS) analysis, with the other 1% being host mRNA (Routh, Domitrovic and Johnson, 2012a). These observations coupled with structures of virions where segments of the genome seem to interact with the capsid, as discussed in section 1.4, and single molecule assembly investigations, support the idea that sequence specific elements with affinity for CP (packaging signals) dispersed throughout the viral genome direct the assembly process of some ssRNA viruses (Perlmutter and Hagan, 2015).

1.9 Packaging signals

Packaging signals (PSs) are short, conserved sequences within the viral genome that interact with coat proteins during the assembly process. PSs have been identified in several viral genomes including human viruses such as Hepatitis B virus (HBV) (Patel *et al*, 2017) and human parecho virus (HPeV) (Shakeel *et al*., 2017), bacteriophages such as MS2 (Rolfsson *et al*., 2016), and plant viruses such as STNV (Bunka *et al*., 2011; Patel *et al*., 2015). While historically thought of as purely electrostatic, packaging signals highlight the interdependent nature of the relationship between RNA and CP in the assembly process. *In vitro* reassembly reactions during assembly investigations have typically been carried out at high protein concentration (≥10 μM) and often result in atypical outcomes i.e. reassembly around non-cognate RNAs. Such concentrations may have masked the actions of packaging signals, whereas recent advances in single molecule techniques allows assays to be carried out at nanomolar concentrations (Stockley *et al.*, 2013). Indeed, when the assembly of MS2 genomic RNA is contrasted with that of a non-cognate viral RNA control (STNV), MS2 CP causes cognate RNA to collapse followed by an increase in hydrodynamic radius until the size expected for a complete MS2 particle. (Figure 1.15). This is not seen with the control RNA, which remains the same size.

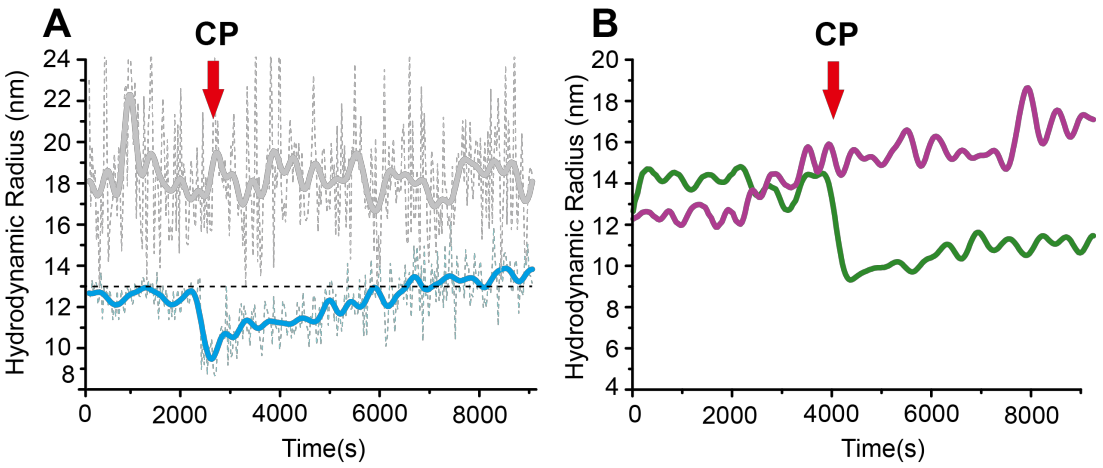


Figure 1.15: RNA collapse during viral assembly. (A) Single molecule traces for genomic MS2 RNA (blue) and non-viral RNA of the same length (grey). Upon the addition of MS2 CP (red arrow) hydrodynamic radius of viral RNA decreases and subsequently increases to the size expected for an MS2 particle (black dashed line). This behaviour is not seen in the non-viral control RNA suggesting that this behaviour is due to sequence specific recognition as opposed to electrostatics. **(B)** Single molecule traces for a 3′ fragment of MS2 RNA (green) and a similarly sized non-viral control RNA (purple). As in A, addition of MS2 CP causes a collapse in the viral RNA which is not seen in the control. Adapted from Borodavka *et al*., 2012.

1.9.1 Identification of packaging signals for STNV

For the model system and plant virus, STNV, SELEX (Systematic Evolution of Ligands by EXponential enrichment) was used to determine the favoured RNA motif involved in CP-RNA interactions. First described in the 1990s, SELEX is a technique that screens an oligonucleotide pool for specific targets (Chai *et al*., 2011). ′

Unlike MS2, STNV has no obvious capsomere beyond the monomer. In 2011 Bunka *et al* devised a protocol that took advantage of the lysine residues on the exterior faces of the capsid subunits (e.g. Lys27) to perform SELEX on STNV VLPs (Figure 1.16). These residues were biotinylated, after which the VLPs were disassembled. Biotinylated CP monomer was immobilised via incubation with streptavidin-coated Dynabeads which were subsequently washed to remove unbound material. Disassembled VLPs (including RNA

encapsidated in *E.coli*) were then added in a substoichiometric ratio i.e. not enough to form complete capsids around all immobilised subunits. The resulting immobilised, quasi- formed VLPs should in theory have exposed RNA binding sites that would ordinarily be concealed in fully formed entities and represent a range of assembly intermediates. Using complete capsids to eliminate sequences aimed at exterior faces, a library of ~1015 sequences, consisting of a random region (N30) bordered by fixed primers, was added to the immobilised VLPs. After 10 rounds the 17 successfully binding sequences (16 of which were unique) were eluted, reverse transcribed and amplified prior to sequencing.

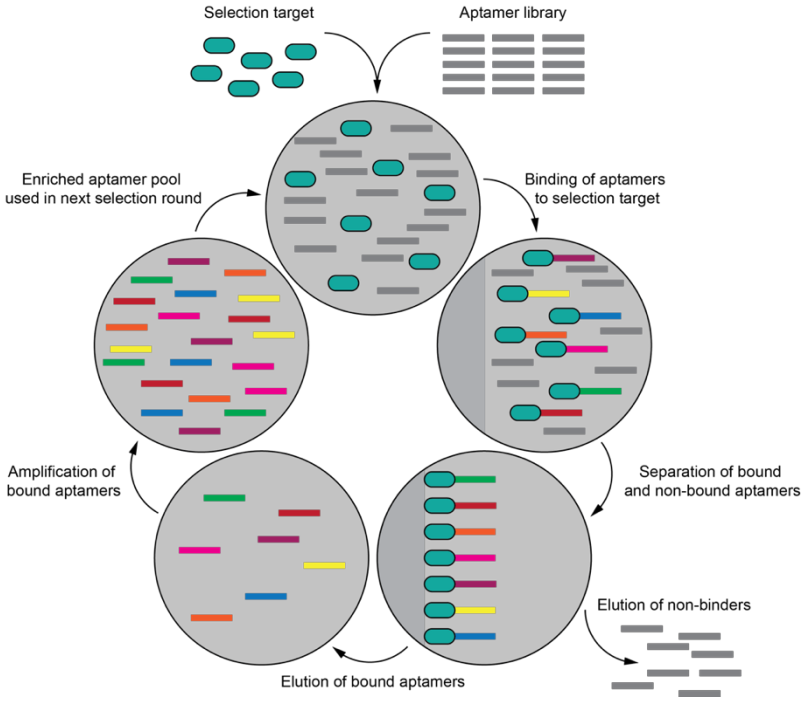


Figure 1.16: SELEX schematic. Diagram showing the stages of SELEX. Target is exposed to a naïve pool of aptamers a subset of which bind and remain bound after elution. Bound aptamers are eluted and amplified, and this enriched pool is used for the next round of selection.

A multiple sequence alignment algorithm (Alibee) identified several motifs; however, there was no clear consensus until the 16 candidates were compared to the STNV-1 genome. Statistically significant matches were discerned including one aptamer sequence (B3) that matched the STNV-1 genome in 16/25 nt, including a 10 nt stretch. When B3 and the corresponding region of STNV RNA were analysed using Mfold it was found that both sequences form 3 bp stem-loops. This can be extended to a 4 bp stem as the nucleotides flanking the region can also base pair, U-A (B3) and C-G (STNV-1). The degree of sequence similarity between B3 and positions 57-81 of the STNV-1 genome is unlikely to be coincidental suggesting that the 'B3-like' area of the STNV-1 genome is a CP binding site. 10 of the 16 identified sequences were able to form stem loops, a known protein recognition motif, presenting AXXA (where A is adenine and X is any residue) in a 4-, 5- or 6- nucleotide loop. When the adjacent sequences were considered, a further 4 aptamers conform to this pattern. Using this information, a set of six stem loops incorporating the AXXA motif at different positions in a 4-, 5- or 6- nucleotide loop, was used to probe the STNV-1 genome. This identified 30 stem-loop sites all presenting an AXXA motif (Bunka *et al*., 2011).

1.9.2 Effect of CP concentration on packaging signal mediated assembly

Data gathered both via computer simulations and experimentally has identified nucleation and subsequent growth as the predominant assembly mechanism (Zandi *et al*., 2006). *In silico* investigations carried out by Dykeman *et al* highlights the importance of coat protein concentration during viral assembly.

Dykeman's model employs the simplest analogue of capsid formation, 12 pentagons which can join together forming a dodecahedron. Hypothetical RNA molecules containing 12 "packaging signals" of varying affinities bind to the centre of these pentagons allowing CP-CP interactions to occur (Figure 1.17) (Dykeman *et al*., 2014). Nucleation via CP-CP interactions of the initial CP-RNA contacts starts assembly in this model, consistent with the two-phase model suggested by Borodavka *et al* in 2012 (Borodavka *et al.*, 2012) (Figure 1.17).

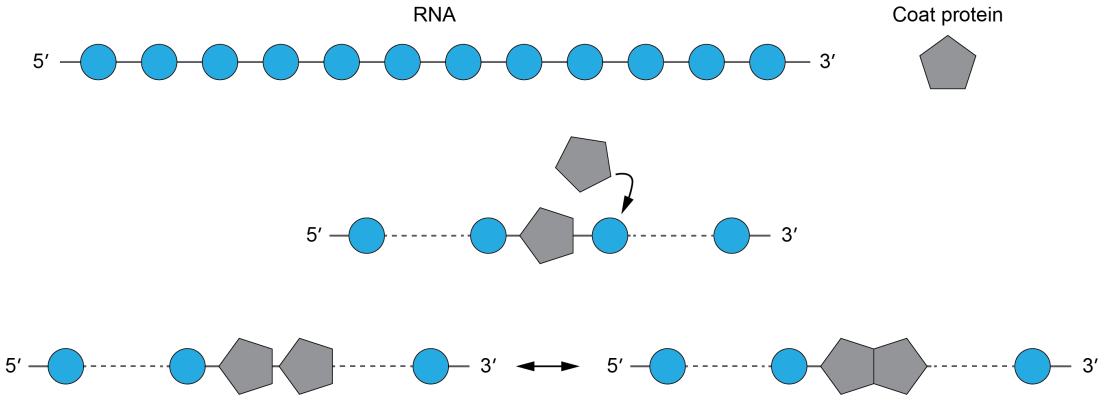


Figure 1.17: Dodecahedral model of capsid assembly. Components of the assembly system are a 'viral RNA' sequence with 12 CP binding sites (blue circles) and grey pentagons (CP pentamers). CP binds to the PS and allows for the formation of subsequent CP-CP interactions which causes a collapse in the hydrodynamic radius of the RNA. Adapted from Dykeman *et al*., 2014.

At each simulation stage all possible reactions between assembly intermediates are calculated based on their relative concentration and CP-RNA/CP-CP affinities. By analysing all the potential interactions between RNA-CP and CP-CP relative probabilities of subsequent reactions an assembly pathway for each hypothetical RNA can be calculated. A smaller volume of building blocks in *in vitro* assembly decreases the number of intermediates that could occur during the assembly process. A definitive order of sequence-specific RNA-CP binding sites, or packaging signals, would lead

to a defined pathway for assembly. The addition of protein over a period of time biases the assembly routes to a small group of stable intermediates. It is only now, with novel techniques, that viral genomes are being searched for other putative packaging signals. The nucleation point is defined as the location on the RNA where the initial CP-CP complex takes place. When CP is added all at once, nucleation can occur on any of the PS sites regardless of affinity. When CP is added gradually, with a low initial starting concentration mirroring *in vivo* conditions, nucleation occurs predominantly on an adjacent high affinity pair of PS. This suggests that gradually increasing CP concentration allows nucleation to occur in a controlled manner. Assembly with this method results in a higher particle yield and the absence of kinetic traps. When a sequence possesses three potential nucleation sites RNA encapsidation is less efficient suggesting that a single high affinity nucleation site is advantageous during assembly (Dykeman *et al*., 2014).

1.9.3 Co-operativity in PS function

MS2 assembly reactions at nanomolar concentrations show a sudden collapse in RNA hydrodynamic radius upon addition of CP to the cognate genome (Patel *et a*l., 2015; Borodavka *et al*., 2012) (Figure 1.18). This is thought to be due to a subset of CP binding the viral RNA, and then forming CP-CP interactions. This collapse is not seen with non-viral controls or with non-cognate genomes (Patel *et al*., 2015; Borodavka *et al*., 2013). This characteristic decrease of R_h is compatible with an assembly mechanism that is orchestrated by multiple packaging signals dispersed throughout the viral genome interacting with CP in a sequence specific manner (Patel *et a*l., 2015).

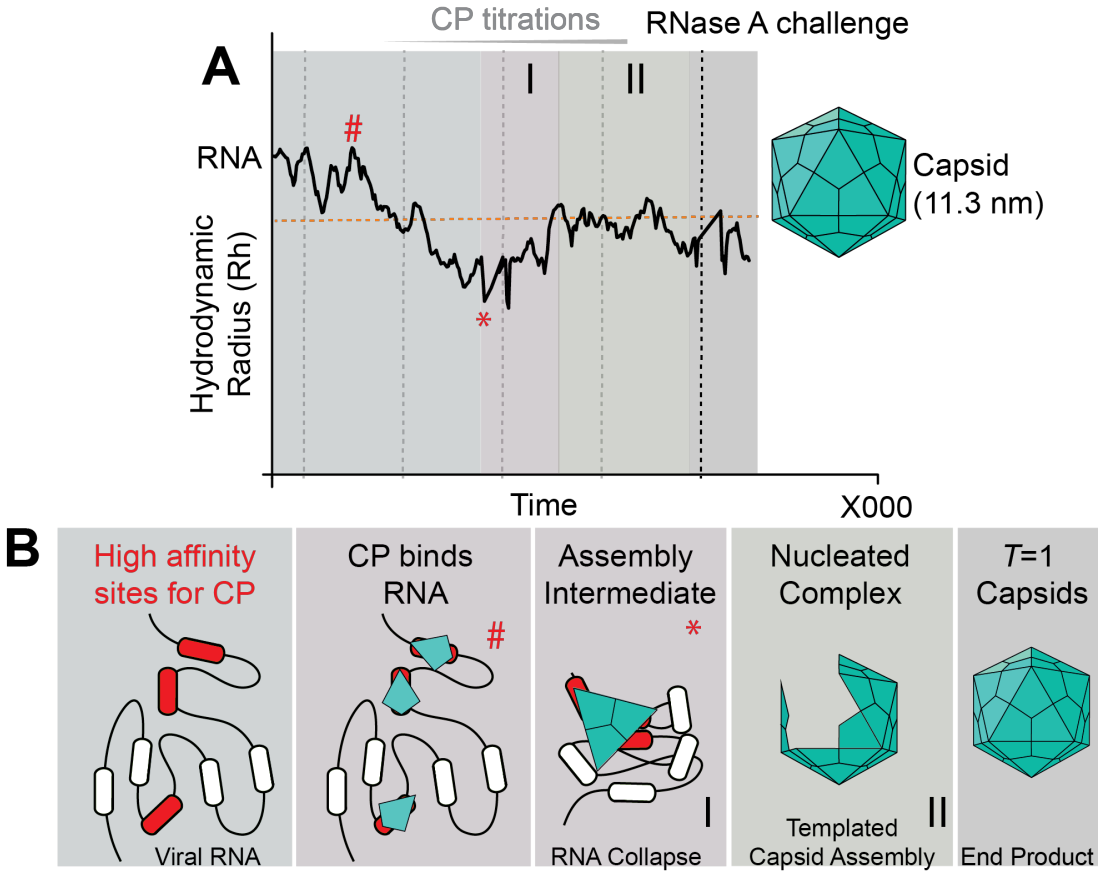


Figure 1.18: 2-Phase Viral Assembly Model. (A) Schematic of hydrodynamic radius changes during viral assembly adapted from (Borodavka *et al*., 2012). **(B)** Protein free genomic RNA with R_h ~20-30% larger than the interior of a capsid containing multiple high affinity sites for CP (Red). CP binds to these sites and subsequent CP-CP interactions trigger a collapse of the genomic RNA and reduce the R_h to fit into the capsid. During phase II, more CP is recruited to complete the capsid.

In the absence of CP, viral RNAs are larger than the internal volume of the capsid. Charge neutralisation due to electrostatics would cause gradual compaction of the RNA. However, the compaction seen in these assembly investigations is fast (Borodavka *et al*., 2012). The packaging signals with the highest affinities (i.e. MS2 TR), those that initiate the assembly process have long been accepted in the literature (Dykeman, Stockley and Twarock, 2013a). However, affinity between one packaging signal and RNA is not thought to be enough to achieve such levels of specificity as seen in vivo (Zlotnick and Mukhopadhyay, 2011). Assembly of viral capsids at low concentrations is

efficient and accurate and multiple contacts between coat protein and packaging signals spread throughout the genome would allow for this specificity (Stockley *et al*., 2013). This is seen when capsid assembly with a single packaging signal, the aptamer B3, is contrasted with a 5' 127nt fragment from the STNV-1 genome (Figure 1.19A). This sequence contains 5 stem-loop structures that all present the AXXA motif in 4, 5 or 6 nt loops (Figure 1.19A). When STNV CP is added to PS1-5 the characteristic hydrodynamic radius (R_h) drop is seen at 120 nM CP, at 200 nM CP rapidly forms nuclease resistant particles. Assembly with B3 is slower and capsid size particles are only achieved at 3000 nM CP (Figure 1.19B). As in the Dykeman model, the multiple sites on the PS1-5 fragment act cooperatively and form a nucleation point from which further CP-CP interactions can occur (Patel *et al*., 2015).

Both model systems (STNV and MS2) used to investigate assembly mechanisms show interaction between coat protein and RNA (Patel *et al.*, 2015; Ford *et al*., 2013). Stem-loops in STNV-1 contact coat protein subunits at the three-fold axes, allowing a CP-CP electrostatic barrier to be overcome. In MS2, stem-loops act as a conformational switch causing a shift from symmetric to asymmetric dimers. Both these co-assembly mechanisms conform to the same principle, that there are multiple packaging signals spanning the genome that interact with coat proteins and increase efficiency of assembly (Stockley *et al*., 2007).

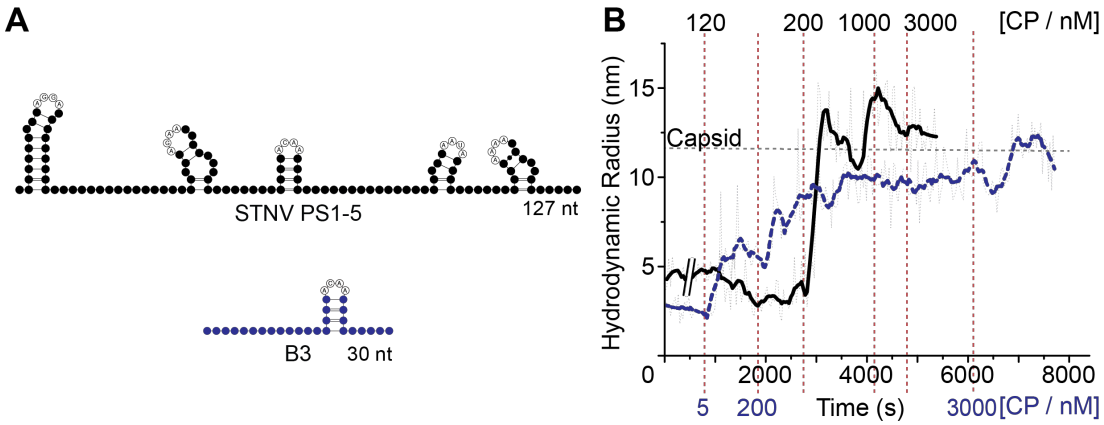


Figure 1.19: Cooperativity in Packaging Signals. (A) Secondary structures of PS1-5 and B3. **(B)** Coat protein titration into 10 nM Alexa-fluor-488 labelled PS1-5 (Black) or B3 (Blue). B3 forms a capsomere with a hydrodynamic radius of ~5 nm at 5 nM CP with VLPs beginning to form at 200 nM CP and progressing until 3000 nM. PS1-5 however, binds CP at 120 nM resulting in a hydrodynamic radius drop of 20-30%, at 200 nM it rapidly assembles into nuclease resistant capsids. Taken from Patel *et al*., 2015.

1.9.4 Importance of PSs in viral assembly

In order to investigate whether interaction between STNV CP and RNA were sequence specific or electrostatic the stem loop motif of the B3 aptamer (ACAA) was replaced with UUUU forming the mutant B34U. When B34U was incorporated into assembly reactions there was an initial increase in R_h, suggesting formation of a capsomere. This was however, observed at an approximate 2-fold increase in coat protein concentration. Assembly does not proceed beyond this point until the coat protein concentration reaches 3 μM where increase in R_h is followed by a fixed decline. Despite being the same length and having the same negative charge as B3, B34U is only capable of producing unstable aggregates that are susceptible to the action of RNase. As B34U is able to form a capsomere, this suggests that the initial stages of recognition involve interactions with the stem. These results suggest that B3 and ultimately the B3 like area of the STNV-1 genome have the characteristics

of a packaging signal (Patel *et al*., 2015). As the mutated motif is less capable of initiating assembly it has been proposed that the loop sequences allow the stem components to interact with the N-terminal helices of the coat protein, neutralising the positive charge and causing the region to become more ordered (Figure 1.20).

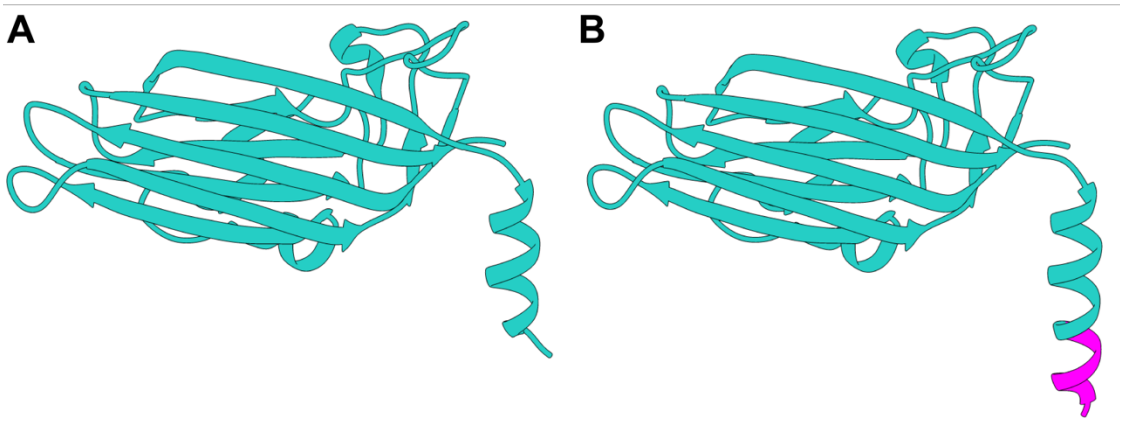


Figure 1.20: Recombinant STNV CP N-terminal ordering on interaction with RNA. (A) Recombinant STNV CP from *E.coli* (PDBID 3S4G). **(B)** Additional ordering of the N-terminus seen on interaction with RNA, highlighted in magenta (PDBID 4BCU).

To build on the theory of multiple packaging signals of varying affinity spanning the genome, further investigations were carried out with the PS1-5 STNV fragment. A series of mutants encompassing motif changes to all U's in all loops, all but 'B3 like' or 'B3 like' only were tested under similar conditions. With all motifs converted to UUUU, coat protein again bound at 120 nM but this time an increase in R_h was observed. When all sites but B3-like are mutated binding once again occurs at 120 nM and the decrease in R_h is observed although successive assembly is slow and only achieved at 3 μM coat protein. These results suggest that the loop motifs are also integral to *T*=1 production, that PS3 (B3-like) is responsible for initial decrease in R_h and cooperation between packaging signals makes assembly efficient.

Investigations also probed the relative spacing of these packaging signals and found that by increasing spacing between PS3 and the adjacent PSs by 10 nt 5´ and 3´ generated malformed capsids with few correctly assembled species. Combining all these results implies that in addition to initial binding at PS3 correct spacing between packaging signals is required for successful assembly (Patel et al., 2015). The packaging signals of MS2 (TR) and STNV-1 both form 4 nt stem loop structures that present motifs which are highly similar, AUUA (Rolfsson et al., 2008) and ACAA respectively. The 30 additional potential packaging signals identified by Bunka *et al* for STNV-1 also present AXXA motifs in 4, 5 and 6 nt loops. This is also the case with the genomes and mRNA of both STNV-2 and STNV-C, which incorporate multiple AXXA displaying stem-loops *(Bunka et al.*, 2011).

Countless cellular events are coordinated by specific interactions between RNA and proteins. These interactions remain however, less well characterised than DNA protein interactions. Previous work with MS2 has investigated the protein:RNA interactions in the MS2 translational repressor complex. The area of the MS2 genome that encompasses the CP recognition sequence, TR, is a 19 nt stem loop. The entire stem loop sequence however, can be altered without affecting binding so long as secondary structure remains intact and certain rules are obeyed at certain positions (Haneef *et al*., 1989). The interaction between CP and RNA was investigated via the introduction of single, double or triple random coding mutations into the MS2 CP gene and altering the TR sequence. RNA binding and assembly capacity was then analysed. At concentrations where binding of RNA:CP is specific a mutation

at -10, the location of the ‘A bulge’ did not promote binding of RNA to coat protein (Talbot *et al*., 1990). Vital residues in the loop are adenines at positions -4 and -7 along with uridine at -5 in relation to the replicase cistron. These key residues can be changed to guanosine resulting in lower affinity however, when -5 is changed from U to C affinity is significantly higher than wild type (Stockley *et al*., 1995). When other RNAs are added in high concentrations to CP-TR mixtures very little competition is observed suggesting that the binding of RNA by MS2 is highly specific. This is further highlighted by reduction of affinity, in some cases 100 fold or more, when point mutations are introduced in the RNA binding site (Johansson *et al*., 1997).

1.9.5 Defining the STNV Packaging Signal Motif

The rate of viral mutation and emergence means that novel methods of controlling spread are needed. The building of a protective coat is on the whole a conserved process however; it remains a largely unexplored drug target. As initiation of viral assembly, at nanomolar concentrations, seems to rely on CP interactions with one/few high affinity sites on the genome, any therapy that interrupts these interactions could be deleterious to viral assembly and by extension viral spread. Further *in silico* investigations by Dykeman *et al* showed that in the presence of hypothetical cellular RNA (12 sites of 3.0 kcal/M affinity) in at 300:1 ratio (cellular: viral) and 1 mM PS-binding drug, viral load dropped by ~20%. Other observed effects were increasing population of malformed capsids and increasing incorporation of non-viral RNAs. The formation of irregular capsids in natural infection could potentially trigger host immune responses leading to clearance of infection (Dykeman *et al*., 2014).

To interrupt these high-affinity packaging signals their recognition motif needs to be defined further. It has been shown that when the 4 nt loop of the B3 aptamer is changed to UUUU (B34U) RNA:CP binding is affected. A higher CP concentration is required to form a capsomere-like structure and after the addition of 3 μM CP only unstable, RNase accessible aggregates are formed. The 5 putative packaging signals present on the PS1-5 fragment, including the B3-like PS3, all present the AXXA motif in 4-, 5- or 6 nt loops (Patel *et al*., 2015). When original SELEX results for STNV are opened up to allow for non-Watson-Crick base pairing in the stem, 30 stem-loops are identified which all present the AXXA motif. This number of packaging signals would account for the RNA density seen in crystallographic maps (Bunka *et al*., 2011).

1.10 Thesis overview

In this thesis I will describe work I have done to explore RNA and genome encapsidation in STNV (Chapters 3-5) and expression and characterisation of alphavirus core proteins in both *E.coli* and insect cells systems as a prelude to identification of their PSs (Chapters 6 and 7).

[This page has been left intentionally blank]

Chapter 2: Materials and Methods

[This page has been left intentionally blank]

2.1 General Materials

Table 2.1: Competent cells

Cell type	Strain Genotype
BL21 (DE3) Gold pLysS	*B F– dcm+ Hte ompT hsdS(r – m–) galλ (DE3) [pLyss Cam^r]* endA Tet$^{r\ a}$*
NEB 5-alpha	*fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
Rosetta cells	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^r)
BL21-CodonPlus (DE3)-RIL	*B F– ompT hsdS(r – m –) dcm$^+$ Tet^r gal λ(DE3) endA Hte [argU ileY BB leuW Cam^r]*
Lemo 21 (DE3) Cells	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(Cam^R) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*

Table 2.2: Antibiotic stock and working concentrations

Antibiotic	Stock concentration	Working concentration
Carbenicillin	100 mg/mL in 70% Ethanol	100 µg/mL or 50 µg/mL*
Chloramphenicol	34 mg/mL in 100% ethanol	34 µg/mL
Tetracycline	10 mg/mL in 70% ethanol	10 µg/mL
Gentamicin	7 mg/mL in autoclaved 18.2 MΩ water	7 µg/mL
Kanamycin	50 mg/mL in autoclaved 18.2 MΩ water	50 µg/mL

*as specified in individual experiments.

All antibiotics were filtered (0.2 µm) and stored at -20˚C. All antibiotic stocks were made at 1000x concentration, 1 µL was added per mL of media.

Table 2.3: Buffer recipe Table

Buffer	Recipe
10X STNV Native buffer	500 mM HEPES, 2 M NaCl pH 7.5
10X STNV Dissociation buffer	500 mM Tris, 100 mM EDTA pH 8.5
10X STNV Reassembly buffer	500 mM HEPES, 30 mM $CaCl_2$* pH 7.5
STNV Buffer B	150 mM HEPES, 9 mM $CaCl_2$, 2 M NaCl pH 7.5
10X Alphavirus Native buffer (1 L)	200 mM HEPES, 2.5 M NaCl pH 7.4
10X Alphavirus solubilisation buffer (1 L)	200 mM HEPES, 2.5 M NaCl, 500 mM $MgCl_2$ pH 7.4
10X Alphavirus reassembly buffer (1 L)	250 mM HEPES, 1 M KOAc, 1.7 M $Mg(OAc)_2$ pH 7.4
Alphavirus B (1 L)	20 mM HEPES, 1.5 M NaCl, 50 mM $MgCl_2$ pH 7.4
10X Tris-Glycine running buffer	250 mM Tris, 1.92 M Glycine, 1% SDS
1.5 M Tris	1.5 M Tris, pH 8.8 adjusted to 500 mL 18.2 MΩ water
1 M Tris	1 M Tris, pH 6.8, adjusted to 500 mL 18.2 MΩ water
1X RNA folding buffer	50 mM NaCl, 10 mM Tris-HCl, 1mM DTT pH 7.9
Luria-Bertani (LB) medium	10 g NaCl, 10 g Tryptone, 5 g yeast extract in 1 L 18.2 MΩ water and autoclaved
Terrific Broth (TB) medium	12 g Tryptone, 24 g yeast extract, 4 mL glycerol in 900 mL 18.2 MΩ water. Autoclaved and adjusted to 1L with 100ml of filter sterile 0.17 M KH_2PO_4, 0.72M K_2HPO_4
LB-Agar	5 g NaCl, 5 g tryptone, 2.5 g yeast extract, 7.5 g agar in 500 mL 18.2 MΩ water. Autoclaved
10X MOPS buffer (1 L)	400 mM MOPS free acid, 100 mM NaOAc, 10 mM EDTA, pH 7
10X TBE buffer (1L)	108 g Tris, 55 g Boric acid, 40 ml 50 mM EDTA soltuon pH 8.0
5X TBE buffer (1 L)	54 g Tris, 27.5 g Boric acid, 20 mL 50 mM EDTA solution pH 8

* $CaCl_2$ stored as frozen 1 M stocks and added to 1X buffer before use.

2.2 General methods

2.2.1 Growth of o/n *E. coli* culture

A single colony of *E. coli* was picked and placed into a 50 mL falcon tube containing 10 mL LB and appropriate antibiotics. Culture was grown at 37˚C o/n with shaking at 220 rpm.

2.2.2 Preparation of bacterial stocks for freezing

100 µL, 80% (0.22 µm filtered) glycerol was added to 900 µL of o/n *E. coli* culture in a 1.5 mL cryovial and stored at -80˚C.

2.2.3 Electrophoresis

2.2.3.1 SDS-PAGE preparation

Polyacrylamide gels were made using the recipes as in Table 2.4. SDS-PAGE were typically 15% resolving unless otherwise stated. 0.75 mm spacer glass plates and short plates (BioRad) were cleaned with ethanol before use and assembled with casting stands and clamps. Appropriate resolving solution was poured into the plates and topped with isopropanol to form a stacking/resolving interface. Upon polymerisation of the resolving gel the isopropanol was removed, and the 5% stacking solution poured on top. A 10/15 well comb (Biorad) was added before the stacking gel set.

Table 2.4: SDS-PAGE recipes

Resolving Gel

	Solution components	5 mL	10 mL	15 mL	20 mL	25 mL	30 mL
12%	H_2O	1.6	3.3	4.9	6.6	8.2	9.9
	30% Acrylamide Mix	2.0	4.0	6.0	8.0	10.0	12.0
	1.5 M Tris (pH 8.8)	1.3	2.5	3.8	5	6.3	7.5
	10% SDS	0.05	0.1	0.15	0.2	0.25	0.3
	10% APS	0.05	0.1	0.15	0.2	0.25	0.3
	TEMED	0.002	0.004	0.006	0.008	0.01	0.012
15%	H_2O	1.1	2.3	3.4	4.6	5.7	6.9
	30% Acrylamide Mix	2.5	5.0	7.5	10.0	12.5	15.0
	1.5 M Tris (pH 8.8)	1.3	2.5	3.8	5	6.3	7.5
	10% SDS	0.05	0.1	0.15	0.2	0.25	0.3
	10% APS	0.05	0.1	0.15	0.2	0.25	0.3
	TEMED	0.002	0.004	0.006	0.008	0.01	0.012

Stacking Gel

Solution components	1 mL	2 mL	3 mL	4 mL	5 mL	6 mL
H_2O	0.68	1.4	2.1	2.7	3.4	4.1
30% Acrylamide Mix	0.17	0.33	0.5	0.67	0.83	1.0
1.5 M Tris (pH 6.8)	0.13	0.25	0.38	0.5	0.63	0.75
10% SDS	0.01	0.02	0.03	0.04	0.05	0.06
10% APS	0.01	0.02	0.03	0.04	0.05	0.06
TEMED	0.001	0.002	0.003	0.004	0.005	0.006

2.2.3.2 SDS-PAGE sample preparation for Protein

Samples were prepared by adding an equal volume of 2X Laemmli sample loading buffer (Sigma-Aldrich) and incubating at 95˚C for 5 minutes. 10/15 µL of appropriate sample was loaded into each well before being run at 180 V for ~60 minutes, in tris-glycine running buffer. Gels were stained using QuickBlue Coomassie stain (TripleRed). Gel images were captured using a Syngene Gel Doc.

2.2.3.3 Native Polyacrylamide Gel Electrophoresis for DNA

Native PAGE were prepared as follows: 12 mL 30 bis-acrylamide (29:1 v/v 30% solution), 7.2 mL 5X Tris-borate-EDTA (TBE) buffer (PanReac AppliChem), 16.4 mL ddH_2O, 360 µL Ammonium persulphate (APS) and 36

µL N,N,N',N'-tetramethylethane-1,2-diamine (TEMED). 1 µL sample and 1 µL 6X loading buffer (NEB) was added to appropriate well and run at 280 V for approximately 90 minutes. Gel was stained with ethidium bromide (0.5 µg/mL) and visualised using a Syngene Gel Doc transilluminator filter.

2.2.3.4 Denaturing RNA Polyacrylamide Gel Electrophoresis for RNA

10% denaturing PAGE were prepared using the UreaGel system (29:1 SequaGel National Diagnostics). Gels were pre-run at 15 W for 3 hours in TBE. Wells were rinsed before loading of samples to remove urea. Samples were prepared with 2X gel loading buffer (Ambion), incubated at 95˚C for approximately 5 minutes and loaded into appropriate wells. The gel was run at 15 W for ~ 90 minutes in TBE and stained with ethidium bromide (0.5 µg/mL). The gel was rinsed three times with autoclaved, deionised H_2O and visualised using the transilluminator function on a Syngene Gel Doc.

2.2.3.5 1 % (w/v) Agarose gel for Plasmids/DNA

0.5 g of agarose was dissolved in 50 mL of TBE buffer with gentle heating. Agarose was allowed to cool slightly then poured into a tray and left to set with a comb inserted. Samples were loaded with Midori Green 6X loading dye (Geneflow) and either a 10 bp (Invitrogen) or a 2-log (NEB) DNA ladder. The gel was run in TBE, at 100 V at room temperature and visualised using a Blue Green LED Transilluminator (Geneflow).

2.2.3.6 1% (w/w) Denaturing Agarose for RNA

0.5 g of agarose was dissolved in 31 mL of 18.2 MΩ ddH_2O with gentle heating. In a fume cupboard, 9 mL formaldehyde was added followed by 10 mL 10X 3-(N-morpholino) propanesulfonic acid (MOPS) slowly. 2X volume, 50% glycerol was added to samples before loading. The gel was run in MOPS buffer at 80 V for 1 hour.

2.2.4 Preparation of grids for negative stain transmission electron microscopy (TEM)

Formvar/Carbon coated grids (Agar scientific) were glow discharged using a PELCO easiGlow (Ted Pella) or UV irradiated for approximately 15 minutes. 20 µL of sample was pipetted onto the grid and allowed to adsorb for 30 seconds. After this time, filter paper (Whatman No.1®) was used to blot the excess sample. 5 µL of uranyl acetate (2% w/v) was used to stain the grid and immediately blotted with filter paper. This staining process was repeated, and the grids were left to air dry. Images were captured using a JEOL JEM-1400. Magnification of individual images and scale bars are indicated in Figure legends.

2.2.5 Analytical ultracentrifugation – Sedimentation velocity

2.2.5.1 Background

Sedimentation velocity analytical ultracentrifugation (svAUC) measures the rate molecules move while under centrifugal force. The rate of sedimentation can be analysed by measuring the change in UV absorbance over time. This technique can be utilised to determine molecular weight and shape of

molecules in heterogeneous solution (Freundlich, 1940). A homogeneous sample solution is put into a 2-sector cell and an applied centrifugal force causes the solution to sediment to the bottom. The area near the meniscus becomes solute depleted and a sharp boundary forms between the sedimenting uniform sample and the depleted area (Figure 2.1). The movement of this boundary is measured, and a sedimentation coefficient (s) is determined which depends on particle mass, inversely on frictional coefficient and can be used as a measure of particle size (Ralston, 1993) see equation 2.1 (Freundlich, 1940). Previous investigations have shown that STNV VLPs sediment at approximately 42 S (Ford, 2012).

Equation 2.1

$$s = \frac{u}{\omega^2 r} = \frac{M(1 - \bar{\nu}\rho)}{Nf}$$

u = constant velocity

ω^2 = square of the angular velocity

r = distance from axis of rotation

M = molecular weight ($g.mol^{-1}$)

$\bar{\nu}$ = partial specific volume of molecules ($cm^3.g^{-1}$)

ρ = density of the solvent ($g.mL^{-1}$)

N= Avogadro's constant (6.022×10^{23} mol^{-1})

f = frictional coefficient of molecule

s = sedimentation coefficient (Svedberg 10^{-13} seconds)

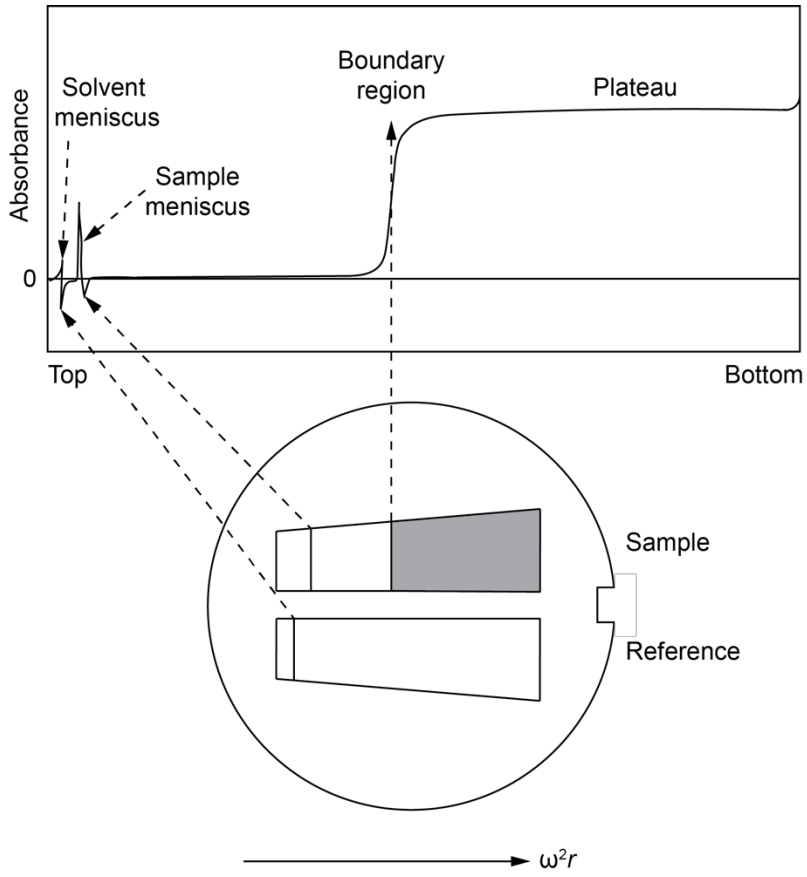


Figure 2.1: Schematic of boundary forming in svAUC. (Top) Formation of a sharp boundary when high angular velocity is applied to molecules in solution which is not seen in the reference sample. **(Bottom)** Illustration of 2-sector sedimentation velocity cell showing the formation of a boundary due to concentration gradient in the sample. Movement of this boundary is monitored and used to determine the sedimentation coefficient.

2.2.5.2 Method

0.32 mL of sample was placed in a 1.2 cm path length 2- sector meniscus matching epon centrepiece cell constructed with sapphire windows. The samples were centrifuged in an Optima XL-I analytical ultracentrifuge at 20°C in an An50-Ti rotor. Changes in absorbance at 260 nm were detected by absorbance optics. Buffer densities and viscosities were calculated using Sednterp version 1.09 (Laue et al., 1992). Data was fitted and analysed using the program Sedfit version 12.1b 2010, using a continuous distribution c(s) Lamm equation model (Schuck, 2000).

2.2.6 Fluorescence Anisotropy

2.2.6.1 Background

In a solution, molecules are randomly oriented therefore their ground state electrons are also randomly oriented. Fluorophores with transition dipole moments aligned with the electric vector of the exciting plane polarised light will be preferentially excited. Fluorescence anisotropy describes the extent to which the emitted light is polarised. It is quantified by measuring the difference between signal detected in the plane perpendicular to the excitation plane and signal detected parallel to the excitation plane. Rate of rotational diffusion or 'tumbling' of a molecule is dependent upon the size and shape of the rotating molecule and solvent viscosity. The larger the molecule, the slower the rotational diffusion and more emitted light will be polarised (Lakowicz, 2006).

2.2.6.2 Method

Satellite tobacco necrosis virus (STNV) coat protein was titrated into 5 nM Alexa-fluor-488 labelled B3 or TR and the polarisation was measured using an EnVision® multilabel plate reader. Samples were then taken for EM analysis and smFCS measurements.

2.2.7 Single Molecule Fluorescence Correlation Spectroscopy (smFCS)

2.2.7.1 Background

SmFCS is a technique that allows for the detection of single molecules *in vitro* and *in situ* through statistical analysis of equilibrium fluctuations in a small observed confocal volume (Lakowicz, 2006; Krichevsky and Bonnet, 2002).

By tracking these fluctuations over time and autocorrelating the intensity measurements hydrodynamic radii, mobility and diffusion can be determined (Fitzpatrick and Lillemeier, 2011). As FCS measures fluctuations caused by single molecules entering and leaving a small observable volume, a highly sensitive confocal microscope is essential for these investigations. In a typical FCS instrument (Figure 2.2) excitation of fluorophores in a sample is achieved with a laser. The beam is directed to the objective with a dichroic mirror. Photons emitted from the sample pass back through the mirror and are filtered in to the detector (Tian *et al*., 2011). The detection module typically includes an autocorrelator connected to a photomultiplier tube or an avalanche photodiode (APD), which using photomultiplication converts light into an electrical signal.

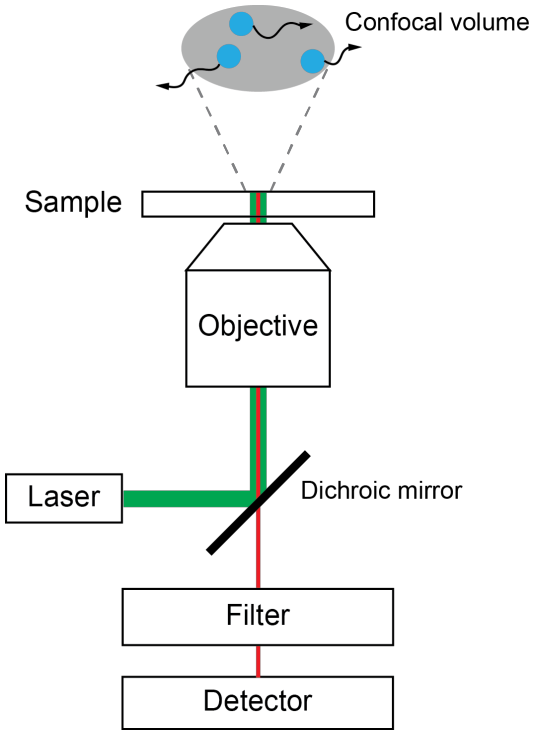


Figure 2.2: smFCS set up. Excitation laser is directed at a dichroic mirror which directs the beam into the objective where it is focused on the sample. As the laser excites fluorophores light is emmited and collected by the objective. The excitation light is filtered by the dichoric mirror and an additional filter towards the detector. The photon bursts are autocorrelated to give diffusion time which can be used to calculate hydrodynamic radius (R_h).

In FCS the principle measurement is fluorescence intensity over time as seen in an intensity/fluorescence trace. As illustrated in Figure 2.3, the amplitude of these fluctuations is equivalent to particle numbers. While the frequency of the fluctuations indicates the time molecules spend within the confocal volume. These readings are then quantified by autocorrelation analysis, which tests for self-similarity, and curve fitting (Leica Microsystems, 2010). During the process of autocorrelation, the intensity signal is compared to itself after the lag time (τ).

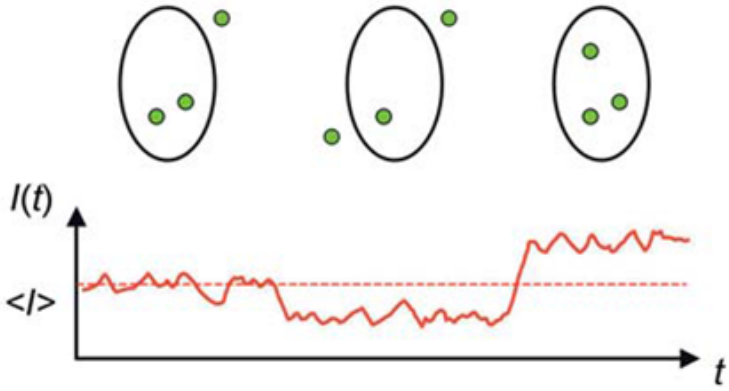


Figure 2.3: Fluorescence intensity trace. Fluorophores (green dots) moving in and out of the confocal volume cause fluctuations in fluorescence intensity. Taken from Leica Microsystems, 2010.

The signal is copied and shifted along the data set by the lag time (Figure 2.4A). When the data is only shifted a small amount the data is still self-similar and has a higher sum. The further the shift the lower the self-similarity sum. The autocorrelation sum is then plotted against lag time (Figure 2.4B). Larger particles will diffuse slowly in and out of the confocal volume. Slower diffusion causes slower fluctuations in the intensity of photon bursts. Faster fluctuations result in a steeper autocorrelation curve. By plotting the autocorrelation function against lag time, it is possible to determine particle numbers (N) and diffusion time (correlation time). From these values concentration and diffusion coefficient can be determined (Leica Microsystems, 2010).

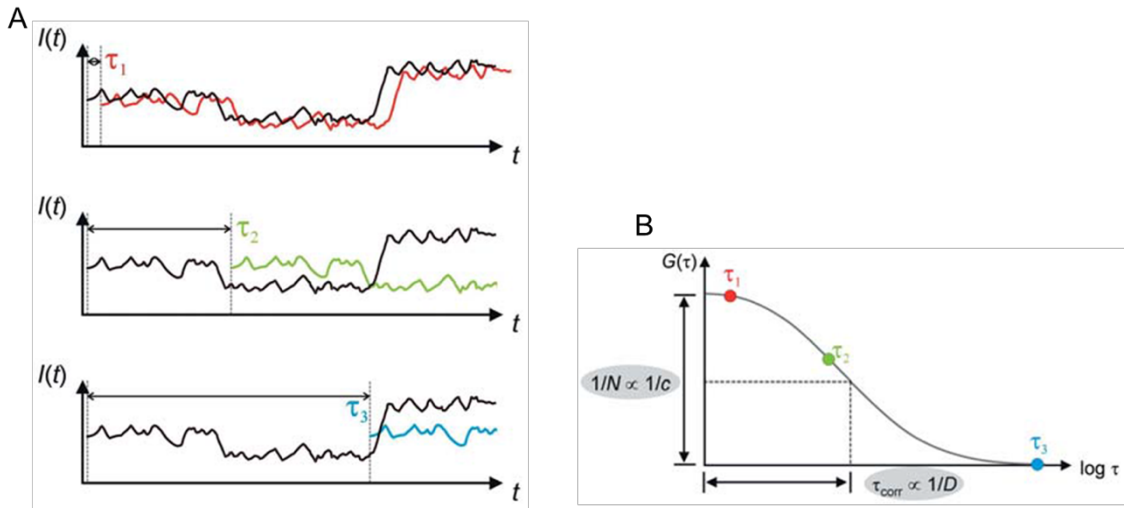


Figure 2.4: Autocorrelation function and curve fitting. (A) The intensity trace gets copied and multiplied by itself, this is done multiple times and tests for self-similarity. When the lag time is small the data is still self-similar and has a higher sum. As the lag time moves, the data becomes less self-similar and the sum drops. **(B)** Autocorrelation function from A is plotted over lag time, with coloured traces matching up to coloured dots on the curve. Taken from Leica Microsystems, 2010.

2.2.7.2 smFCS Data Collection and Analysis

FCS measurements were performed on a custom-built smFCS facility illustrated in Figure 2.2. Excitation laser (Sapphire CW blue laser, 488 nm, Coherent, USA) power was set to 65 µW. Focus position was adjusted to 20 µm from the cover slip inner surface (maintained by piezoelectric feedback loop, Piezosystems Jena, Germany). Immersion oil (refractive index 1.515, type DF, Cargille Laboratories, USA) was used with immersion oil objective (63 x magnification, numerical aperture 1.4). Photon count was recorded and analysed by an ALVL5000 multiple tau digital correlator (ALV-GmbH) in single channel mode. FCS data was analysed using non-linear Least-Squares fitting with a single component diffusion model autocorrelation function corrected for the triplet state in MATLAB. Scripts used for data processing were provided by Dr. Roman Tuma. Diffusion time was used in the calculation of apparent hydrodynamic radius (R_h) and plotted as a function of assembly time. R_h

calculations were based on the measured diffusion time for Alexa-Fluor-488 (AF488) dye with the estimated R_h = ~0.7 nm in assembly buffer (Doeven *et al*., 2005).

2.2.7.3 Equations for smFCS

Equation 2.2: Equation describing shape of confocal volume

$$V_0 = \pi^{\frac{3}{2}}\omega_r^2\omega_z$$

ω_r^2 = radial distance
ω_z^2 = axial distance
From (Borodavka, 2012.)

Equation 2.3: Normalised autocorrelation function

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2}$$

$G(\tau)$= autocorrelation function
F = intensity
δF= intensity fluctuations
t = time
τ = lag time
From (Schwille and Haustein, 2001)

Equation 2.4: Relationship between diffusion time and diffusion coefficient

$$D_{\text{trans}} = \frac{\omega_r^2}{4\tau_D}$$

D_{trans} = diffusion coefficient
τ_D= diffusion time
ω_r^2 = radial distance
From (Serdyuk *et al*., 2007)

Equation 2.5: Equation for 3D diffusion

$$G(\tau) = \frac{1}{N}\left(1 + \frac{4D_{\text{trans}}\tau}{\omega_r^2}\right)^{-1}\left(1 + \frac{4D_{\text{trans}}\tau}{\omega_z^2}\right)^{-\frac{1}{2}}$$

$G(\tau)$= autocorrelation function
N= number of particles
D_{trans} = diffusion coefficient
τ = lag time
ω_r^2 = radial distance
ω_z^2 = axial distance
From (Schwille *et al.*, 1996; Gosch and Rigler, 2005)

Equation 2.6: Fitting experimentally obtained autocorrelation function into 3D diffusion model

$$G(\tau) = \frac{1}{N}\left(1 + \frac{\tau}{\tau_D}\right)^{-1}\left(1 + \frac{\tau}{k^2\tau_D}\right)^{-\frac{1}{2}}$$

$G(\tau)$= autocorrelation function
N= number of particles
τ = lag time
τ_D= diffusion time
$k = \frac{\omega_z}{\omega_r}$ = confocal structural parameter*
From (Borodavka, 2012.)
* Confocal structural parameter is determined experimentally at FCS data collection by measuring diffusion properties of a well-characterised dye.

2.2.7.4 Preparation of samples for smFCS

RNA was annealed by heating to 80°C and cooling at 0.1°C/min to 4°C using a thermal cycler in RNA folding buffer. All smFCS experiments took place at 21°C in a sterile nuclease-free chamber (Lab-Tek II Chambered Coverglass, Nunc).

2.2.8 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

2.2.8.1 Background

The first instance of liquid based molecular size differentiation was in 1953 by Wheaton and Bauman during their investigations into ion-exclusion where they found that non-ionic species such as D-glucose and methanol were separated by size on a Dowex 50X8 resin (Wheaton and Bauman, 1953). This was quickly followed by the separation of amino acids and peptides on a starch column leading to the size exclusion techniques used widely today (Hong et al., 2012). Size exclusion chromatography (SEC), relies on the use of known standards for the estimation of molecular weight of a sample (Sahin and Roberts, 2012). When this technique is coupled with multi angle light scattering (MALS) to form SEC-MALS more accurate measurements can be obtained. Samples are first separated by hydrodynamic volume and subsequently MALS analyses the angular dependence and intensity of the light scattering, measuring the size and molar mass of the molecules. A quasi-elastic light scattering detector (QELS) in series allows for measurement of the hydrodynamic radius of a sample (Wyatt Technology, 2018b).

2.2.8.2 Method

500 µL of samples were loaded onto a pre-equilibrated SEC column (superdex200 increase) at a flow rate of 0.4 mL/min on an ÄKTA pure (GE Healthcare) with an associated DAWN8+, optilab T-rEX and QELS (Wyatt). Absorbance of samples was measured at both 260 and 280 nm using Unicorn software version 7 (GE Healthcare, 2014) and light scattering was measured and analysed using Astra software version 6.1.6.5 (Wyatt Technology, 2016).

2.2.8.3 Calculation of hydrodynamic Radius from LS

QELS (also known as light scattering) is the fluctuation and intensity of scattered light with time. This allows for the measurement of a translational diffusion coefficient (T) which can in turn be used to calculate hydrodynamic radius using the Stokes-Einstein equation (Equation 2.7).

Equation 2.7: Stokes-Einstein Equation

$$D_T = \frac{kT}{6\pi\eta \mathrm{R}_h}$$

R_h = Hydrodynamic radius
η = Solvent viscosity
D_T = Translational diffusion coefficient
K = Boltzmann constant
T = Temperature

2.2.9 DNA Sequencing

All Sanger sequencing was carried out by Source Bioscience. Primers not supplied by Source Bioscience are shown in appendix B.

2.2.10 RNA *in vitro* Transcriptions

Transcriptions were carried out using Highscribe™ T7 High yield RNA synthesis kit (NEB) following manufacturers protocol with the addition of 1 µL RNase inhibitor (Invitrogen). Products were run on a denaturing PAGE (2.2.3.6), stained with ethidium bromide (0.5 µg/mL) and visualised using a Genius with a transilluminator function (Syngene).

2.2.11 Gel Purification of RNA

Bands stained with ethidium bromide (10^5 v/v dilution) were visualised under UV light and removed from the gel using a scalpel and stored with 100 µL of elution buffer (EB) (Qiagen) at -20°C o/n. Sample was removed from the freezer and rolled at room temperature for 2 h. The EB was removed (stored at 4˚C) and a further 100 µL EB added. This was repeated for a total of 300 µL EB. 30 µL 3 M sodium acetate was added to the total EB and vortexed for ~10s before adding 3 µL glycogen (Invitrogen) and 300 µL 100% cold, filtered (0.22 µm) isopropanol. The RNA was incubated at -80°C for 30 minutes then centrifuged at 16,160 x *g* for 1 hour. Supernatant was removed, and the pellet was washed with 1 mL 70% filtered (0.22 µm) ice-cold ethanol and centrifuged for 20 minutes at 16,160 x *g*, 4°C. Pellets were then dried under vacuum (Genevac) and the pellet resuspended in 30 µL Diethyl pyrocarbonate (DPEC) treated, nuclease free water. RNA concentration was measured using a NanoDrop One spectrophotometer (Thermo Fisher).

2.2.12 Dye-Labelling of Oligonucleotides

6 µL of each 5′ amino modified oligonucleotide was incubated at room temperature, rolling, and protected from light with 3 µL (10 mM) AF488 and 1 µL (1 M) sodium borate pH 8.3. Samples were then analysed by denaturing PAGE (2.2.3.4) and purified using the ‘crush and soak’ method (2.2.11).

2.3 Specific Methods for Chapter 3

2.3.1 STNV expression

2.3.1.1 Transformation of competent BL21 (DE3) Cells

50 µL of competent *E. coli* cells (BL21 (DE3) Gold) (Agilent) were thawed on ice. Upon thawing, 200 ng of plasmid pET22b(+)STNV was added and incubated on ice for 30 minutes. The cells were heat shocked in a water bath at 42˚C for 45 seconds and incubated on ice for a further 5 minutes. 0.95 mL SOC media (NEB) was added to transformed *E. coli* and incubated for an hour at 37˚C, 220 rpm. Following incubation, 50 µL of the sample was removed and plated out onto LB agar (100 µL/mL Carbenicillin) and incubated o/n at 37˚C.

2.3.1.2 STNV protein expression

Single colonies of transformed *E. coli* were picked and grown in 5 mL of LB (Carbenicillin 100 µg/mL). Inoculations were incubated o/n at 37˚C, 180 rpm. The 5 mL cultures were used to inoculate 1 L flasks of LB (Carbenicillin 100 µg/mL). Flasks were placed in an orbital incubator (180 rpm) until the cultures reached an OD_{600} of 0.6. Following this, the culture was heat shocked at 42˚C for 30 minutes, then cooled to 24˚C for 20 minutes. Cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration 1 mM) and incubated at 24˚C o/n before harvesting.

2.3.1.3 Cell harvest and lysis

Samples were subjected to centrifugation at 4˚C and 5250 x *g* in a Beckman Coulter SX4750A rotor for 45 minutes. Cell pellets were resuspended in 100

mL STNV Native Buffer plus 1 EDTA-free protease inhibitor tablet (Thermo Fisher) homogenised and stored at -80˚C until needed. Samples were thawed and stirred on ice with 1 mg of lysozyme (Sigma) for 30 minutes before five cycles of 30 seconds on/off sonication on ice (Amplitude 15 microns). Following sonication, 1 mg DNase I (Sigma) and 1 EDTA-free protease inhibitor tablet was added, and cell suspension was incubated at 4˚C on ice for 1 hour. Cell suspension was then ultracentrifuged at 38,430 x *g* for 30 minutes in a SW32 ti rotor (Beckman Coulter).

2.3.1.4 Ammonium sulphate precipitation

The clarified lysate was ammonium sulphate precipitated by slowly dissolving ammonium sulphate until a final concentration of 30% w/v was reached (~1.75 M). The solution was then ultracentrifuged at 38,430 x *g* for 30 minutes in a SW32 ti rotor (Beckman Coulter). The resulting pellet was resuspended in STNV Native buffer and dialysed against 3 L of STNV Native Buffer at 4˚C o/n.

2.3.2 STNV VLP purification

2.3.2.1 Sucrose Density Gradient Sedimentation

15% Solution: 15 mL 10X STNV Native buffer, 22.5 mL 2 M Sucrose stock solution and 112.5 mL ddH_2O.

45% Solution: 15 mL 10X STNV Native buffer, 67.5 mL 2 M Sucrose stock solution and 67.5 mL ddH_2O.

15%-45% sucrose gradients were made by layering a 45% sucrose solution under a 15% sucrose solution and mixed using a gradient mixing station (Biocomp).

1 mL of STNV VLPs, obtained from ammonium sulphate precipitation, was pipetted onto each gradient (or until balanced) and ultracentrifuged using an SW32Ti rotor (Beckman Coulter) at 51,610 x *g* for 21 hours 15 minutes at 4˚C. 1 mL fractions were collected across the gradient and peaks that absorbed at 260 and 280 nm were analysed by 15% SDS-PAGE. Fractions containing STNV VLPs were pooled and dialysed against 3 L STNV Native buffer.

2.3.2.2 Q-Sepharose Anion Exchange Chromatography

STNV VLPs were loaded onto a pre-equilibrated (STNV Native buffer) XK 50 anion exchange Q-sepharose column (GE Healthcare) at a flow rate of 3 mL/min using the sample pump of an ÄKTA explorer (GE Healthcare) and the 'flow through' was collected in 15 mL fractions. Fractions that absorbed at 260 and 280 nm were analysed by SDS-PAGE and concentrated by ammonium sulphate precipitation as described in 2.3.1.7, before visualisation via negative stain TEM.

2.3.2.3 Size Exclusion Chromatography

STNV VLPs were loaded onto a pre-equilibrated (STNV Native buffer) Superdex-200 increase 10/300 GL (GE Healthcare) in 1 mL injections (flow rate of 0.5 mL/min) and 1 mL fractions were collected. Fractions that absorbed at 260 and 280 nm were analysed by SDS-PAGE and negative stain TEM. Fractions containing STNV VLPs were pooled and stored at 4˚C.

2.3.3 Disassembly and reassembly of Recombinant STNV

Dissociation of VLPs was achieved by dialysis (molecular weight cut-off (MWCO) 3 kDa) of 0.5 mL of recombinant STNV VLPs (~2 mg/mL) into 3 L of STNV Dissociation buffer (EDTA-free protease inhibitor tablet, 1 µg/mL Pepstatin A) for 5 hours at 4°C. Reassembly of VLPs was facilitated by dialysing dissociated VLPs into 3 L of STNV reassembly buffer (EDTA-free protease inhibitor tablet, 1 µg/mL Pepstatin A) at 4°C o/n.

2.3.4 SEC-MALS of recombinant STNV

Recombinant STNV was loaded onto a pre-equilibrated (STNV Native buffer) TSK gel g6000 PWXL 13 um, 7.8x300 mm (TOSOH), as in 2.2.8.2 and collected in 1 mL fractions.

2.3.5 Charge Detection Mass Spectrometry of STNV

Purified recombinant STNV VLPs in STNV Native buffer were concentrated using Ambion Ultra 0.5 mL centrifugal filters with a 10K MWCO. Concentrated samples (3.4 mg/mL) were analysed by charge detection mass spectrometry (CDMS) performed by Corrine Lutomski and Dr. Elizabeth Pierson from the Jarrold Lab, Indiana University, Bloomington. The sample was buffer exchanged into 50 mM ammonium acetate and ions produced using a nanoelectrospray source (Advion Biosciences, Ithaca, NY, USA). These ions are introduced to the mass spectrometer where the ions are separated from background gas. A narrow band of ion kinetic energies are selected in the dual hemispherical deflection analyser and focused into the ion trap. Trapped ions oscillate through the charge detector which induces a periodic charge. This

charge is detected and amplified with a charge-sensitive pre-amplifier (Amptek A250). The oscillation frequency of ions in the trap is related to the m/z and the charge is proportional to the amplitude at the fundamental frequency in FFT (Pierson *et al*., 2016).

2.3.6 STNV CP purification

STNV VLPs were dissociated in a 1:1 ratio with 10X STNV dissociation buffer (e.g. 10 mL STNV VLPs with 10 mL 10X STNV dissociation buffer) and Pepstatin A (final concentration 20 µM) for 1 hour rolling at 4˚C. The dissociated sample was loaded onto pre-equilibrated (STNV dissociation buffer) 5 mL Q-sepharose FF (Q-FF) (GE Healthcare) and 5 mL SP-Sepharose FF (SP-FF) (GE Healthcare) ion exchange columns in series at a flow rate of 1 mL/min in 10 mL injections. The loaded columns were washed with 25 mL STNV dissociation buffer before removal of Q-FF column. The SP-FF column was washed with a further 25 mL of STNV dissociation buffer. STNV CP was eluted from the column with a 20 minute, 0-100% gradient of 2M NaCl in STNV dissociation buffer collecting 0.5 mL fractions. Fractions that absorbed at 260 and 280 nm were analysed by SDS-PAGE and negative stain TEM. Pooled CP fractions were dialysed into STNV reassembly buffer for 36 hours at 4˚C. Dialysed coat protein was quantified using a NanoDrop One spectrophotometer (Thermo Scientific) and used in downstream investigations.

2.4 Specific methods for Chapter 4

2.4.1 Production of B3 variants

2.4.1.1 B3 Short Variant Sequences

Table 2.5: B3 Short Variant Sequences

Variant	Sequence 5′→3′	Loop Sequence	Extinction Coefficient ($M^{-1}cm^{-1}$)
1	ACAUGCAAUUAUGCACAC	AUUA	172,600
2	ACAUGCAAUUUUGCACAC	AUUU	165,400
3	ACAUGCAUUUAUGCACAC	UUUA	167,400
4	ACAUGCAAUUGUGCACAC	AUUG	169,300
5	ACAUGCAGUUAUGCACAC	GUUA	171,300
6	ACAUGCAGUUGUGCACAC	GUUG	168,000
7	ACAUGCAGUUUUGCACAC	GUUU	164,100
8	ACAUGCAUUUGUGCACAC	UUUG	164,100
B3 Short	ACAUGCAACAAUGCACAC	ACAA	176,600
B34U	ACAUGCAUUUUUGCACAC	UUUU	160,200
B3	CCUUUUCAAGACAUGCAACAAUGCACACAG	ACAA	282,700

Extinction coefficients were calculated using ITD oligo analyser.

2.4.1.2 Preparation of DNA Templates using PCR

B3 short variants were produced using primers shown in Appendix B.1 and the KAPA2G system (KAPA biosystems) following the manufacturer's protocols. Resulting PCR products were analysed by 1% native PAGE (2.2.3.3).

2.4.1.3 RNA *in vitro* Transcription

Transcription of B3 variants was carried out as in 2.2.10 and gel purified as in 2.2.11.

2.4.2 STNV reassembly in the presence/absence of aptamer B3

Reassembly reactions were carried out in the presence and absence of B3 short in a 1:3 RNA:CP molar ratio, final concentration of STNV CP was 6 μM. CP and RNA concentrations were calculated using the extinction coefficients for the STNV CP (Appendix A.3) and relevant RNAs (Table 2.5), absorbance at 280 nm and the Beer-Lambert equation. Samples were dialysed o/n in STNV reassembly buffer (EDTA-free protease inhibitor tablet, 1 μg/mL Pepstatin A) at 4˚C stirring. All samples were analysed by negative stain TEM.

2.4.3 STNV reassembly in the presence of B3 variants

Reassembly reactions were carried out in the presence of B3 short variants (Table 2.5) in a 1:3 RNA:CP ratio, final concentration of STNV CP was 6 μM. Samples were dialysed o/n in STNV reassembly buffer (EDTA-free protease inhibitor tablet, 1 μg/mL Pepstatin A) at 4˚C stirring. All samples were analysed by negative stain TEM and svAUC.

2.4.3.1 Sedimentation Velocity Analytical Ultracentrifugation

As in 2.2.5.2, samples were centrifuged at 18111.6 x *g* with 100 scans taken in approximately 11 hours 30 minutes. Changes in solute concentration were detected by interference and absorbance scans at 260 nm. Density and viscosity of buffers were calculated using Sednterp version 1.09 (Laue, Shah *et al*, 1992). Data was fitted and analysed using the program Sedfit version 12.1b, 2010, using a continuous distribution c(S) Lamm equation model (Schuck, 2000).

2.4.3.2 SEC-MALS of B3 variants

1 mL STNV reassembly reactions were carried out in the presence of B3 variants 1, 4, 6, B3 and B34U in a 1:3 RNA:CP molar ratio, final concentration of STNV CP was 6 μM. Samples were dialysed o/n in STNV reassembly buffer (EDTA-free protease inhibitor tablet, 1 μg/mL Pepstatin A) at 4˚C stirring. SEC-MALS of the samples was analysed using a pre-equilibrated (STNV reassembly buffer) superose-6 SEC column (GE Healthcare), as in 2.2.8.2.

2.4.4 smFCS displacement assays

Initial measurements of Alexa-fluor-488 labelled RNA was taken for at least 10 runs of 30s (5 minutes). Purified STNV CP was titrated into labelled RNA until the sample had formed a capsomere structure (R_h=~5 nm) the sample was monitored for a further 120 runs of 10s (20 minutes) to ensure stability. At this point, unlabelled B3 short/B3 short variant was added in 100-fold molar excess and measured for 120 runs of 10s. Diffusion time was used in the calculation of apparent R_h and plotted as a function of assembly time. R_h calculations were based on the measured diffusion time for AF488 dye with the estimated R_h = ~0.7 nm in assembly buffer.

2.5 Specific methods for Chapter 5

2.5.1 STNV CP purification

STNV CP purification was carried out as described previously in specific methods for Chapter 2.

2.5.2 Preparation of genomic length templates

WT, C4-WT and ΔAXXA templates were obtained from Dr. Nikesh Patel from (Patel et al., 2017). GBlock® coding for the WT-ΔAXXA genome length variant (Appendix D) was cloned into the BamHI (NEB) and HindIII (NEB) sites of a linearised pACYC1 vector using T4 DNA ligase (NEB) as per manufacturer's instructions. The plasmid was transformed into DH5α cells (NEB) and verified by sequencing using T7 forward primer from Source Bioscience and pBAMrev.

2.5.3 Transcription of genomic length RNAs

Transcription performed using a HighScribe T7 kit (NEB):

Table 2.6: T7 HiScribe Transcription of genomic length RNAs

Component	Volume for 1 reaction
GTP	2.2 µL
ATP	6 µL
UTP	6 µL
CTP	6 µL
Amino GMP (Trilink)	3.8 µL
10X Buffer	6 µL
T7 polymerase	6 µL
RNase Inhibitor (Invitrogen)	1 µL

Components were added to each transcription reaction along with 2 µL of 0.1 M DTT and 1 ng appropriate template in a total volume of 60 µL. Transcriptions were incubated at 37˚C for 3 hours in a thermal cycler and isolated using Qiagen RNeasy Kit as per manufacturer's instructions.

2.5.4 Labelling of genomic length RNAs

30 µL of genomic length variants (~2 µM) were incubated at room temperature, rolling, and protected from light with 2 µL (10 mM) AF488 and 4 µL (1 M) sodium borate pH 8.3. Samples were isolated using Qiagen RNeasy Kit as per manufacturer's instructions then precipitated by adding 3.6 µL 3 M filtered sodium acetate, 0.36 µL glycogen (Invitrogen) and 36 µL filtered (0.22 µm) 100% isopropanol and stored at -20˚C o/n. Samples were then centrifuged at 17,000 x *g* 1 hour in a benchtop centrifuge at 4˚C. Supernatant was removed and the pellet was washed twice with ice cold, filtered (0.22 µm) 70% ethanol, centrifuging for 20 minutes between washes. The ethanol was removed to minimal volume and the pellets were dried under vacuum (Genevac) for ~ 15 minutes or until dry. Pellets were resuspended in 100 µL of water and RNA isolated using the Qiagen RNeasy Kit as per manufacturer's instructions and analysed by denaturing agarose gel as in 2.2.3.6.

2.5.5 Single molecule scale reassembly reactions – 1:60 titration

Concentration of RNAs was adjusted to 2 nM per RNA in a total volume of 18 mL STNV reassembly buffer. Single molecule scale reassembly reactions were performed using a liquid handling robot, Biomek 4000 (Beckman Coulter) in a low binding 96 well plate (Corning). Total volume in each well was 200 µL, with 6x STNV CP titration steps of 40 nM, pipette mixing after each addition, to a final concentration of 240 nM. Samples were incubated at 4˚C for 18 hours post reassembly to equilibrate the reactions. Samples were the concentrated to ~500 µL using Vivaspin20 columns, MWCO 50 MDa (GE Healthcare) at 7500 x *g* in a FX6200 rotor at 4˚C (Beckman Coulter) for 2 X 30-minute runs.

2.5.6 Single molecule scale reassembly reactions – 2:60 titration

As in 2.5.5 but titration step of 20 nM to a final concentration of 120 nM.

2.5.7 Single molecule scale reassembly reactions – one step

Concentration of both competitor RNAs was adjusted to 2 nM per RNA in a total volume of 18 mL in STNV reassembly buffer. Single molecule scale reassembly reactions were performed using a liquid handling robot, Biomek 4000 (Beckman Coulter) in a low binding 96 well plate (Corning). Total volume in each well was 200 µL, with a single STNV CP addition of 120 nM. Samples were incubated at 4˚C for 18 hours post reassembly to equilibrate the reactions. Samples were then concentrated to ~500 µL using Vivaspin20 columns, MWCO 50 MDa (GE Healthcare) at 7500 x *g* in a FX6100 rotor at 4˚C (Beckman Coulter) for 2 X 30-minute runs.

2.5.8 SEC-MALS of reassembled products

Concentrated reassembly reactions were loaded onto a pre-equilibrated (STNV native buffer + 200 mM NaCl) TSK gel g6000 PWXL 13 um, 7.8x300 mm (TOSOH) as in 2.2.8.2, collecting 1 mL fractions. Fractions with absorbance at 260 and 280 nm were concentrated using centrifugal filters with 100 kDa MWCO (Amicon). The fluorescence of concentrated fractions (~20 µL volume) were measured using a Nanodrop 3300 (Thermo Fisher) at wavelengths of 510 nm and 617 nm and samples were taken for negative stain TEM.

2.5.9 Incubation with RNase and SEC-MALS of samples

Concentrated fractions were then merged and incubated overnight with 3 µL RNase A (10 mg/mL, Thermo Fisher) at 4°C o/n and loaded onto a pre-equilibrated (STNV native buffer + 200 mM NaCl) TSK column (TOSOH) as in 2.2.8.2, collecting 1 mL fractions. Fractions with absorbance at 260 and 280 nm were concentrated using centrifugal filters with 100 kDa MWCO (Amicon). The fluorescence of concentrated fractions (~20 µL volume) were measured using a Nanodrop 3300 (Thermo Fisher) at wavelengths of 510 nm and 617 nm and samples were taken for negative stain TEM.

2.6 Specific methods for Chapter 6

2.6.1 Production of Alphavirus Cp in *E. coli*

The coding sequence for 5 alphavirus Cps were codon optimised for expression in *E. coli* using the IDT codon optimisation software (Integrated DNA Technologies, 2018) and synthesised as gBlocks with a 5′ NdeI site and a 3′ HindIII site. These gBlocks were ligated into the NdeI/HindIII sites of pET22b(+), transformed into DH5α cells (NEB) and verified by sequencing using T7 forward and reverse primers from Source Bioscience. Alphavirus core protein plasmids were prepared by Dr. Chris Kirkham in our laboratory.

2.6.2 Core protein expression in *E. coli*

2.6.2.1 Transformation into RIL cells

50 µL of competent *E. coli* cells BL21-CodonPlus (DE3)-RIL (Agilent) were thawed on ice to which 200 ng of plasmid pET22b(+) (+insert of choice) was

added and incubated on ice for 30 minutes. The cells were heat shocked in a water bath at 42˚C for 45 seconds and incubated on ice for a further 5 minutes. 0.95 mL SOC media (NEB) was added to transformed *E. coli* and incubated for an hour at 37˚C, 220 rpm. Following incubation, 50 µL of the sample was removed and plated out onto LB agar (100 µg/mL Carbenicillin and 34 µg/mL Chloramphenicol) and incubated o/n at 37˚C.

2.6.2.2 Transformation into RIL cells

50 µL of competent *E. coli* cells BL21-CodonPlus (DE3)-RIL (Agilent) were thawed on ice to which 200 ng of plasmid pET22b(+) (+insert of choice) was added and incubated on ice for 30 minutes. The cells were heat shocked in a water bath at 42˚C for 45 seconds and incubated on ice for a further 5 minutes. 0.95 mL SOC media (NEB) was added to transformed *E. coli* and incubated for an hour at 37˚C, 220 rpm. Following incubation, 50 µL of the sample was removed and plated out onto LB agar (100 µg/mL Carbenicillin and 34 µg/mL Chloramphenicol) and incubated o/n at 37˚C.

2.6.2.3 Transformation into Rosetta cells

20 µL of competent Rosetta *E. coli* cells (Novagen) were thawed on ice, to which 1 µL of plasmid pET22b(+) (+insert of choice) was added and incubated on ice for 5 minutes. The cells were heat shocked in a water bath at 42˚C for 30 seconds and incubated on ice for a further 2 minutes. 80 µL SOC media (NEB) was added to transformed *E. coli* and incubated for an hour at 37˚C, 220 rpm. Following incubation, 50 µL of the sample was removed and plated

out onto LB agar (100 µg/mL Carbenicillin and 34 µg/mL Chloramphenicol) and incubated o/n at 37˚C.

2.6.2.4 Media selection and expression

Single colonies of transformed *E. coli* were picked and grown in 5 mL of LB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL). Inoculations were incubated o/n at 37˚C, 180 rpm. The 5 mL cultures were used to inoculate 500 mL flasks of LB or TB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL). Flasks were placed in an orbital incubator (180 rpm) until the cultures reached an OD_{600} of 0.6 (LB) or 2.5 (TB). All cultures were then induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration 1 mM) and incubated as shown in Table 2.7 before harvesting.

Table 2.7: Expression of Alphavirus Cp

Plasmid	**pET22b(+)-CHIKV**	**pET22b(+)-SFV**	**pET22b(+)-RRV**
Cells	RIL	Rosetta	Rosetta
Media	TB	LB	LB
Antibiotics	Carb/Cam	Carb/Cam	Carb/Cam
Heat shock	42ºC for 30 min 28ºC for 20 min	42ºC for 30 min 28ºC for 20 min	42ºC for 30 min 28ºC for 20 min
Expression Temp (ºC)	37	28	28
Time (h)	6	6	6

2.6.2.5 Expression time-course

A single colony of transformed *E. coli* was picked and grown in 5 mL of LB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL). The inoculation was incubated o/n at 37˚C, 180 rpm. The 5 mL culture was used to inoculate a 500 mL flask of LB or TB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL) and placed in an orbital incubator (180 rpm) until the cultures reached an

OD_{600} of ~0.6 or ~2.5. A sample of non-induced cells was taken before sample was induced with IPTG (final concentration 1 mM) at temperatures in Table 2.7. From this point on a sample was taken every hour and stored at -20˚C before being analysed by SDS-PAGE.

2.6.2.6 Large scale expression

Single colonies of transformed *E. coli* were picked and grown in 10 mL of LB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL). The inoculation was incubated o/n at 37˚C, 180 rpm. The 10 mL culture was used to inoculate a 1 L flask of LB or TB (Carbenicillin 100 µg/mL, Chloramphenicol 34 µg/mL) and placed in an orbital incubator (180 rpm) until the cultures reached an OD_{600} of ~0.6 (LB) or ~2.0 (TB). Flasks were induced with IPTG (final concentration 1 mM) and incubated as in Table 2.7.

2.6.3 Rhamnose Rescue

Following incubation, 50 µL of the sample was removed and plated out onto LB agar plates (100 µg/mL Carbenicillin and 34 µg/mL Chloramphenicol) using conditions shown in Table 2.8, then incubated o/n at 37˚C.

Table 2.8: Rhamnose rescue LB agar conditions

	[IPTG]	[L-Rhamnose]
1	0	0
2	1 mM	0
3	1 mM	200 µM
4	1 mM	500 µM
5	1 mM	1 mM
6	1 mM	2 mM

2.6.4 Purification of Alphavirus Cp

2.6.4.1 Alphavirus cell harvest and lysis

Contents of 1 L flasks of alphavirus Cp expressing cells were subjected to centrifugation at 3,000 x *g* for 45 minutes in a SX4750A rotor (Beckman Coulter) at 4˚C. Cell pellets were resuspended in alphavirus native buffer and stored at -80˚C. When needed, aliquots of resuspended cells were thawed and incubated with 1 mg of lysozyme (Sigma) and 1 EDTA-Free protease inhibitor tablet (Thermo Fisher), on ice for 30 minutes. Sample was then sonicated (5 rounds 30 seconds on/off) on ice and incubated at 4˚C with 1 mg DNase 1 (Sigma) and 1 EDTA-Free protease inhibitor tablet (Thermo-fisher), for 1 hour. Sample was then ultracentrifuged at 17,080 x *g* for 80 minutes in an SW32 ti rotor (Beckman Coulter). Supernatant was discarded, and pellet the resuspended in alphavirus solubilisation buffer.

2.6.4.2 Alphavirus solubility assay

180 µL aliquots of resuspended CHIKV pellet were incubated at room temperature with 20 µL of the components shown in Table 2.9. Samples were mixed gently and incubated at room temperature for 30 minutes before being centrifuged at 17,000 x *g* for 10 minutes. A sample of the supernatant was removed and analysed by SDS-PAGE.

Table 2.9: Solubility assay conditions

Condition	Component	Condition	Component
1	Water	10	50 mM EDTA
2	5% SDS	11	100 mM DTT
3	5% Cetrimonium Bromide	12	1 M Na_2CO_3
4	5% Deoxycholate	13	1 M Na_2HPO_4
5	5% Sarkocyl	14	2 M Na_2HPO_4
6	10% Tween-20	15	1 M Na_2SO_4
7	10% Triton X-100	16	4 M NaCl
8	Igepal	17	1 M $MgCl_2$
9	8 M Urea		

2.6.4.3 Magnesium incubations

180 µL aliquots of resuspended CHIKV pellet were incubated at room temperature for 30 minutes with varying $MgCl_2$ concentrations (shown in Table 2.10) and then centrifuged at 17,000 x *g* for 10 minutes. A sample of the supernatant was removed and analysed by SDS-PAGE.

Table 2.10: Magnesium solubilisation conditions

Condition	[$MgCl_2$] mM	Condition	[$MgCl_2$] mM
1	0	7	6.25
2	0.19	8	12.50
3	0.39	9	25.00
4	0.78	10	50.00
5	1.56	11	100.00
6	3.12		

2.6.4.4 Fast protein liquid chromatography (FPLC)

Cell pellets were resuspended in alphavirus solubilisation buffer and incubated at room temperature (or 4˚C) for 1 hour with rolling. Sample was centrifuged at 5,000 x *g* for 10 minutes in an FX6100 rotor (Beckman Coulter) and then loaded onto a pre-equilibrated (alphavirus solubilisation buffer) 5 mL SP-FF ion exchange column at a flow rate of 3 mL/min collecting 50 mL fractions. The column was washed with 25 mL alphavirus solubilisation buffer, collecting 50

mL fractions. Cp was eluted from the column with a 67 minute, 0-70% gradient of alphavirus B buffer, collecting 2 mL fractions. Fractions that absorbed at 260 and 280 nm were analysed by SDS-PAGE.

2.6.4.5 Liquid Chromatography mass Spectrometry (LC-MS)

Protein desalting and mass analysis was performed by LC-MS using an M-class ACQUITY UPLC (Waters UK, Manchester, UK) interfaced to a Synapt G2S Q-IMT-TOF mass spectrometer (Waters UK, Manchester, UK). Data processing was performed using the MassLynx v4.1 suite of software supplied with the mass spectrometer. Data was collected and analysed by Drs. James Ault and Rachel George at the University of Leeds Mass Spectrometry Facility.

2.6.4.6 Protein Identification mapping

Gel processing, tryptic digestion and LC-MS was performed and analysed by Dr Rachel George at the University of Leeds Mass Spectrometry Facility. LC separation of the peptide mixtures was performed on an ACQUITY M-Class UPLC (Waters UK). Data processing was performed using the MassLynx v4.1 and Peptide MS/MS data were processed with PEAKS Studio (Bioinformatic Solutions Inc) and searched against the Uniprot *E. coli* entry (release 2016_12). Carbamiodomethylation was selected as a fixed modification, variable modifications were set for oxidation of methionine and deamidation of glutamine and asparagine. MS mass tolerance was 20 ppm, fragment ion mass tolerance was 0.05 Da and the false discovery rate was set to 1%.

2.6.4.7 Chemical cleavage of ribosomal S4

Purified CHIKV Cp and ribosomal contaminant were dialysed into alphavirus resuspension buffer o/n at 4˚C. Sample was then frozen at -20˚C and lyophilized (HETO LyoPro 6000 Freeze Dryer). Lyophilized protein was resuspended in chemical cleavage buffer to give a final concentration of ~5 mg/mL. Mixture was incubated in a 45˚C water bath for 4 hours. The reaction was quenched by cooling and addition of 3 volumes of 2% (v/v) trifluoroacetic acid (TFA). Sample was desalted using a NAP-25 column (GE Healthcare) as per manufacturer's instructions, pre-equilibrated in alphavirus solubilisation buffer, taking 1 mL fractions and analysed by SDS-PAGE.

Chemical cleavage buffer (25 mL): 3.06 mL 16.3 M Hydroxylamine, 9 mL 7.5 M prechilled guanidine-HCL, 2 mL 50% w/v NaOH, 5 mL 1M potassium carbonate, pH 9. Volume made up to 25 mL with 7.5 M prechilled guanidine-HCL.

2.6.5 CD analysis of purified Cp

Purified Cp was buffer exchanged in 10 mM sodium phosphate buffer (pH 7.5) using a Bio-spin 6 column (BioRad) according to manufacturer's instructions. Concentration was determined by a Nanodrop One spectrophotometer (Thermo Fisher) and adjusted to 0.02 mg/mL. Samples were placed in quartz cuvettes with a pathlength of 1 mm at 21˚C. Circular dichroism (CD) spectra were recorded using a Chirascan CD spectrometer at wavelengths between 180-340 nm with 1 nm steps and a buffer blank was collected. Data was analysed using Dichoweb, CDSSTR method using reference set 4 (Sreerama and Woody, 2000).

Molar ellipticity (Θ) was calculated from experimental data (millidegrees, machine units) using Equation 2.8.

Equation 2.8: Conversion of millidegrees (machine units) to molar ellipticity

$$\Theta = \mathrm{m}^\circ \times \frac{M}{10LC}$$

m° = millidegrees,
M = average molecular weight (g/mol)
L = pathlength of cell (cm)
C = concentration in g/L.

2.6.6 svAUC analysis of purified Cp

As in 2.2.5.2, samples were centrifuged at 72,446.4 x *g* with 300 scans taken in approximately 16 hours. Data was fitted and analysed using the program Sedfit (Schuck, 2000).

2.6.7 Core particle reassembly

2.6.7.1 Reassembly of purified Cp with non-specific DNA

50 µL of purified alphavirus Cp (10 µM) was prewarmed at 30°C for 10 minutes, followed by addition of 50 µL of 10 µM non-specific DNA (DAR82, sequence in Table 2.11) resulting in an approximate 1:1 molar ratio of components. The sample was then incubated at 30°C for a further 30 minutes and samples were analysed by negative stain TEM. This method was taken from (Tellinghuisen *et al*., 1999). Reassemblies were incubated with 1 M NaCl (final concentration) at room temperature for 10 minutes before analysis by negative stain TEM.

Table 2.11: Oligonucleotide sequence for core particle reassembly

Oligo Name	Sequence 5′→3′
DAR82	CTAGAGCGGAGAGAATCCAATTAGGATATCTTGAGC AGGGGCATGGAGCTC

2.6.8 smFCS analysis of SFV core particle assembly with DAR82

2.6.8.1 Labelling of oligonucleotides

DAR82 was ordered from IDT with a 5´ Amino modifier C6 and labelled as in 2.2.12. Samples were then analysed by denaturing PAGE (2.2.3.4) and purified using the 'crush and soak' method (2.4.1.6).

2.6.8.2 smFCS analysis of SFV core assembly

Initial measurements of 1 nM Alexa-fluor-488 labelled DAR82 were taken for 10 runs of 30s (5 minutes). Purified SFV Cp was titrated into labelled DNA/RNA in 40 nM steps and mixed, with measurements taken for 10 runs of 30s for each titration step. Purified STNV Cp was used as a positive control for the reassembly of STNV VLPs. Diffusion time was used in the calculation of apparent R_h and plotted as a function of assembly time. R_h calculations were based on the measured diffusion time for AF488 dye with the estimated R_h = ~0.7 nm in assembly buffer.

2.7 Specific methods for Chapter 7

2.7.1 Baculovirus expression of alphavirus coat proteins

2.7.1.1 Propagating pFastBac1 Vector

50 µL of DH5α cells (NEB) were thawed on ice. 2 µL pFastBac1 plasmid (Invitrogen) added and mixed by flicking 4-5 times. The mixture was then placed on ice for 30 minutes. Following this, the mixture was heat shocked at 42˚C and placed on ice for a further 5 minutes. 950 µL of room temperature SOC was added and the mixture incubated at 37˚C, 250 rpm for 60 minutes. 50 µL of sample was plated onto LB agar (100 µg/mL carbenicillin) and incubated o/n at 37˚C. Single colonies were picked and used to inoculate 10 mL LB (100 µg/mL carbenicillin) and incubated at 37˚C, 200 rpm, o/n. The sample was then centrifuged at 5,000 x *g* for 10 minutes and DNA isolated using a QIAprep spin mini kit (Qiagen) following the manufacturers protocol. The concentration of purified vector was measured on a NanoDrop One spectrophotometer (Thermo Scientific) and stored at -20˚C.

2.7.1.2 Production of pFastBac1 compatible inserts

Primers were designed using the NEBuilder Assembly tool v1.12.15, shown in Appendix B.2.

2.7.1.3 Q5 PCR of Alphavirus inserts

PCR reaction was set up with the following components:

Table 2.12: Q5 PCR of Alphavirus inserts

Component	Volume for 1 reaction
Q5 reaction mix	5 µL
10 µM dNTPs	0.5 µL
10 µM Forward primer	1.25 µL
10 µM Reverse primer	1.25 µL
Template plasmid (1ng/µL)	1 µL
Q5 High-fidelity Polymerase	0.25 µL
Water	15. 75 µL

Samples were placed into a thermal cycler on the following programme:

Table 2.13: Q5 PCR programme

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
25–35 Cycles	98°C	5–10 seconds
	*50–72°C	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	

2.7.1.4 Digestion of pFastbac1

pFastbac1 vector was linearised using both BamHI-HF and HindIII-HF restriction enzymes (NEB). 50 ng of vector was digested using 2 µL of both restriction enzymes simultaneously, with 5 µL of cutsmart buffer (NEB) in a total volume of 50 µL. Sample was incubated at 37˚C for 4 hours and isolated using a Monarch PCR and DNA clean up kit (NEB) as per manufacturer's instructions.

2.7.1.5 HiFi DNA assembly of pFastbac1 and compatible inserts

Mass of insert required for HiFI DNA assemblies was calculated using NEBioCalculator v1.6.0 (NEB) at ratios of 3:1, 5:1 and 7:1 (insert:vector). Assembly reactions were carried out with 10 ng of vector in a total reaction

volume of 20 µL. Reactions were incubated at 50˚C for 15 minutes and stored at -20˚C.

2.7.1.6 Transformation into DH5α cells

50 µL of competent NEB-5α cells (NEB) per transformation were thawed on ice. 2 µL of the HiFi DNA assembly mixture was added to the appropriate tube and mixed gently. Samples were incubated on ice for 30 minutes prior to heat shocking at 42˚C for 30 seconds. Cells were again incubated on ice for 2 minutes before the addition of 950 µL of room temperature SOC (NEB). Samples were incubated at 37˚C, 250 rpm for 60 minutes. Transformed cells were spread onto LB agar plates (100 µg/mL carbenicillin). Plates were incubated at 37˚C o/n.

2.7.1.6 Purification and Sequence analysis of plasmids

Single colonies from HiFi DNA assembly were picked and used to inoculate 10 mL LB (100 µg/mL carbenicillin) and incubated o/n at 37˚C, 200 rpm. Plasmid DNA was extracted from the samples using QIAprep Spin Miniprep Kit (Qiagen) as per manufacturer's instructions. Extracted DNA was sequenced by Source Bioscience with Polyhedrin forward and SV40 reverse primers, see Appendix B.3.

2.7.1.7 Transformation into DH10Bac cells

DH10Bac cells (Invitrogen) were thawed on ice and 100 µL aliquot transferred to a pre-chilled 15 mL falcon tube (100 µL per transformation). 1 ng of purified plasmid DNA added to each tube and mixture incubated on ice for 30 minutes.

Cells were heat shocked to 42˚C for 45 seconds and incubated on ice for a further 2 minutes. 900 µL SOC (Invitrogen) was added and the sample was incubated at 37˚C, 225 rpm for 4 hours. After incubation, a 3-fold serial dilution was prepared and 100 µL of cells were plated out onto LB agar plates containing kanamycin (50 µg/mL), gentamicin (7 µg/mL), tetracycline (10 µg/mL), Bluo-gal (100 µg/mL) and IPTG (40 µg/mL). Plates were incubated at 37˚C for 48 hours. 10 white colonies from each construct were picked and streaked out again on LB agar plates containing kanamycin (50 µg/mL), gentamicin (7 µg/mL), tetracycline (10 µg/mL), Bluo-gal (100 µg/mL) and IPTG (40 µg/mL). Plates were incubated at 37˚C for 24 hours.

2.7.1.8 Purification of Bacmids

Twice white colonies were picked and used to inoculate 2 mL o/n cultures of LB 50 µg/mL (kanamycin, 7 µg/mL gentamicin and 10 µg/mL tetracycline). Bacmid DNA was isolated from DH10Bac cells using PureLink™ HiPure Plasmid DNA Miniprep Kit (Invitrogen) as per manufacturer's instructions.

2.7.1.9 Analysis of recombinant Bacmids

Purified Bacmid DNA was used as a template in PCR reactions to check for the presence of insert using pUC/M13 forward and reverse primers (Appendix B.3). A reaction was set up for each isolate as in Table 2.14.

Table 2.14: pUC/M13 PCR of bacmids

Component	Volume for 1 reaction
Q5 reaction mix	5 µL
10 µM dNTPs	0.5 µL
10 µM pUC/M13 Forward	1.25 µL
10 µM pUC/M13 Reverse	1.25 µL
Template Bacmid	1 µL
Q5 High-fidelity Polymerase	0.25 µL
Water	15. 75 µL

Samples were placed into a thermal cycler on the following programme:

Table 2.15: Programme for pUC/M13 PCR

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
25–35 Cycles	98°C	5–10 seconds
	66°C	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	

Midori green loading buffer (Nippon genetics) was added to each sample and run on a 1% agarose gel (2.2.3.5) at 100V for ~60 minutes. Bacmid alone will produce a PCR product of approximately 300 bp, a bacmid transposed with pFastBac1 will produce a product of 2300 bp + size of insert.

2.7.2 Maintenance of SF9 insect cells

2.7.2.1 Storage of SF9 insect cells

On receipt of SF9 cells they were placed in liquid nitrogen for long term storage (6 months). Alternatively, if cells were to be used within 2 weeks they were stored at -80˚C.

2.7.2.2 Thawing SF9 cells

To thaw SF9 insect cells a vial was retrieved and immediately transferred to a 37˚C water bath until thawed. Working in a laminar flow hood, the vial was opened, and cells were slowly pipetted into a sterile 50 mL falcon tube. 10 mL

pre-warmed SF900II media (Gibco) (28˚C) containing no antibiotics was added dropwise. The suspension was pipetted 3-5 times and transferred to a T-75 flask (Corning) which was rocked to distribute cells evenly. The flask was incubated at 28˚C for 60 minutes. After incubation, the SF900II media was removed and immediately replaced with 5 mL of fresh pre-warmed (28˚C) SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco). The flask was incubated at 28˚C until 90% confluency by eye.

2.7.2.3 Counting SF9 insect cells

450 µL of PBS (Gibco) and 50 µL of trypan blue (Sigma Aldrich) were mixed in a bijou. 500 µL of insect cells were removed from an adherent or suspension cell culture and aliquoted into a bijou. 20 µL of insect cells were mixed with 20 µL of trypan blue dilution, 10 µL of this mixture was applied to the chambers of a haemocytometer (10 µL each side). Cells were counted using a light microscope in the outer 8 squares of a haemocytometer as shown in Figure 2.5, Total cells were counted and recorded for each square, along with blue cells for each counted square. The average cell count across the 8 counted squares was used to calculate cell density using the following formula.

Cell density (cells/mL)= ((Average counted cells X 10000) X 2) – ((Average blue cells X10000) X 2)

Blue cells were also counted and used to calculate cell viability.

Cell viability = (Average blue cells/ Average cells) X 100

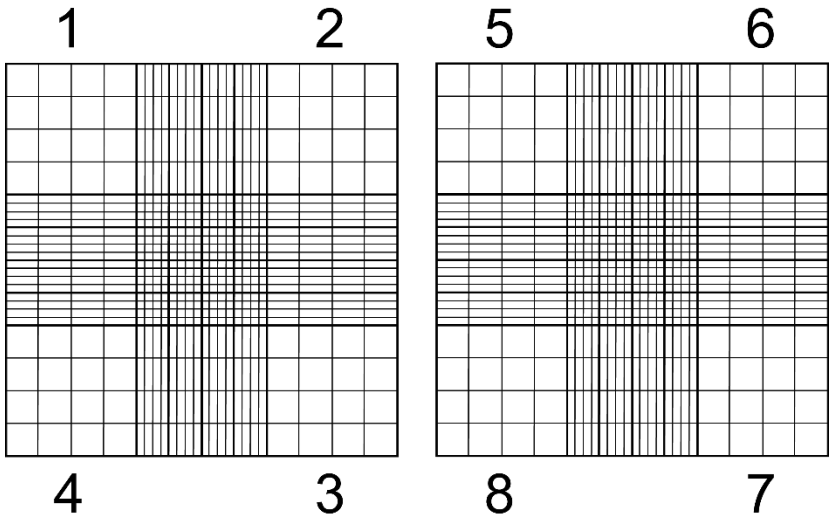


Figure 2.5:Haemocytometer for counting of cells. Cells within the boundaries of the 8 labelled squares were used to calculate the cell density and viability of SF9 cultures.

2.7.2.4 Passaging adherent cultures

Working in a laminar flow hood, the media was removed from a T-75 flask (Corning) with a ~90% confluent monolayer and replaced with 12 mL of prewarmed (28˚C) SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco). Using a cell scraper, the cells were removed from the base of the flask. 500 µL of cell suspension was transferred to a bijou. Cell density and viability was calculated as in 2.7.2.3. Cells were diluted to $5x10^5$ cell/mL and 12 mL was used to seed a new adherent culture in a T-75 flask, incubated at 28˚C.

2.7.2.5 Generation and passaging of suspension SF9 cultures

Working in a laminar flow hood, the media was removed from a T-75 flask (Corning) with a ~90% confluent monolayer and replaced with 12 mL of prewarmed (28˚C) SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco). Using a cell scraper, the cells were removed from the base of the flask. 500 µL of cell suspension was transferred to a bijou. Cell density and viability was

calculated as in 2.6.2.3. For suspension cultures cells were diluted to $5x10^5$ cells/mL in a total volume of 50 mL SF900II (Gibco) plus Antibiotic-Antimycotic (Gibco) in an Erlenmeyer flask (SIMAX) and incubated at 28˚C, 120 rpm. Cells were split when density reached >$4X10^6$ cells/mL.

2.7.2.2 Cleaning of SF9 insect cells glassware

Glassware for SF9 growth was thoroughly cleaned and free from any residues before use using the following protocol:

1. Immediately after being used to grow cells, sterilise glassware with 2% (w/v) Virkon solution for 30 minutes.
2. Rinse thoroughly with water.
3. Clean glassware with a bottle brush and a 1% (v/v) Triton X-100 solution.
4. Visually check that all residue has been removed.
5. Rinse glassware 5 times in tap water and 3 times in deionised water.
6. Clean glassware in a 10% (v/v) acetic acid solution for ≥ 1 hour (or o/n), with foil covering the top. The liquid level should be at least 50% higher than the maximum amount of media/cells that will be used.
7. Rinse glassware thoroughly 5 times in tap water and 5 times in 18.2 MΩ water. Visually check that the glassware has no residue left. If any residue is present, repeat cleaning process.
8. Allow glassware to air dry.
9. Autoclave glassware on a standard sterilisation cycle.
10. Store glassware in a clean room and use within 3 months.

11. Glassware older than 3 months should be rinsed 2X with deionised water and re-autoclaved.

2.7.3 Protein expression in SF9 insect cells

2.7.3.1 Transfection of SF9 insect cells

2 mL of $2x10^6$ cells/mL were seeded into the wells of a 6-well plate (Corning) and adhered for an hour at room temperature in a laminar flow hood. The following 3 tubes were set up:

1) 500 µL of SF900II media (Gibco), 40 µL (50 ng/µL) purified Bacmid DNA and 8 µL p3000 (Invitrogen).
2) 250 µL of SF900II media (Gibco) and 7.5 µL lipofectamine (Invitrogen).
3) 250 µL of SF900II media (Gibco) and 15 µL lipofectamine (Invitrogen).

274 µL of tube 1 was added to both tubes 2 and 3, mixed and incubated at room temperature for 10 minutes. 265.5 µL of tube 2 (3.25 µL lipofectamine) was added dropwise to appropriate well in duplicate. 269.5 µL of tube 3 (7.5 µL lipofectamine) was added dropwise to the appropriate well in duplicate. A negative control of lipofectamine only (250 µL SF900II media and 7.5 µL lipofectamine) was added to the appropriate wells. Samples were rocked to distribute the mixtures evenly and incubated at 28°C, checking cells every 24 hours using a light microscope. Once cytopathic effect was visible (typically 72-96 hours) media was removed from cells, clarified using a 0.22 µm filter and stored at 4°C protected from light.

2.7.3.2 Amplification of viral stock (Formation of P2 viral stock)

In a laminar flow hood 30 mL of $2x10^6$ cells/mL were transferred to a sterile glass Erlenmeyer shake flask (SIMAX) and incubated at 28°C, 120 rpm for 30 minutes. After this time, the cells were infected with 2 mL of the clarified P1 viral stock and incubated at 28°C, 120 rpm. Cells were counted every 24 hours and cell number and viability monitored. Once cell numbers and viability diverged from the control, the sample was centrifuged at 500 x *g* for 5 minutes and media separated from pelleted cells. Media was filtered through a 0.22 µm filter and stored at 4˚C protected from light.

2.7.3.3 Performing a viral plaque assay

In a laminar flow hood 2 mL of $5x10^5$ cells/mL were seeded into the wells of 2, 6 well plates (Corning) and allowed to adhere at room temperature for 1 hour. An 8-log serial dilution was prepared from the clarified P2 viral stock in serum free SF900II media (Gibco) containing no antibiotics. Media was removed from the cells and replaced with 1 mL of the appropriate serial dilution (-8 to -4) or 1 mL of SF900II (Gibco) containing no antibiotics (control). Cells were incubated at 28°C, checking every 24 hours until plaques became visible.

2.7.3.4 Calculating viral titre

The number of plaques in each dilution were counted and the following formula was used to calculate viral titre.

titre (pfu/mL) = number of plaques X dilution factor X 1/(mL of inoculum / well)

2.7.3.5 Western Blot of CHIKV Core protein

Table 2.16: Western Blotting buffers

Buffer	Recipe
Transfer Buffer (1 L)	500 mL MeOH, 2.9 g Glycerine, 5.8 g Tris, 0.37 g SDS
Blocking Buffer	1.875 g Milk Powder in 50 mL TBS
TBS (1 L)	6.05 g Tris, 8.76 g NaCl pH 7.5
TBS-T (1 L)	500 µL Tween-20 in 495 mL TBS

A 15% SDS-PAGE was prepared and run as in 2.2.3.2. Gel was then layered between sheets of wadding and nitrocellulose paper (BioRad) pre-soaked in transfer buffer and run at 100V for 1 hour. Nitrocellulose paper was removed and put in 50 mL blocking buffer rolling, at 4°C o/n. Paper was washed with 15 mL TBS, followed by 2 x 15-minute washes of 15 mL TBS-T. Paper was then incubated with 1/10,000 dilution of sheep anti-chikungunya antibodies (Native antigen company) in 15 mL TBS-T and rolled at room temperature for 1 hour. Paper was then washed twice for 15 minutes with 15 mL TBS-T and incubated with 1/20,000 dilution of rabbit anti-sheep IgG (H/L):Biotin (BioRad) in 15 mL TBS-T rolling for 1 hour at room temperature. Paper was then washed twice for 15 minutes with 15 mL TBS-T and incubated with 1/1,000 dilution of Streptavidin Alkaline Phosphatase (Promega) in 15 mL TBS-T rolling for 1 hour at room temperature. Paper was then washed twice for 15 minutes with 15 mL TBS-T and incubated with 5 mL Western Blue (Promega) until bands appeared.

2.7.3.6 Multiplicity of infection (MOI) optimisation for CHIKV protein expression

1 mL of $6x10^5$ cells/mL were seeded into the wells of a 24 well plate (Corning) and allowed to adhere for 30 minutes at room temperature in a laminar flow hood. The media was removed, cells were washed once with fresh SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco) and a further 300 µL of

SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco) added. Baculoviral stock was added to the cells at MOI of 0, 1, 2, 5, 10 and 20. Cells were incubated at 28°C for 144 hours at which time media was removed, and cells harvested and stored separately at -20 until analysis by SDS-PAGE and western blot as in 2.7.3.5.

2.7.3.7 Time-course optimisation for protein expression

1 mL of $6x10^5$ cells/mL were seeded into the wells of a 24 well plate (Corning) and allowed to adhere for 30 minutes at room temperature in a laminar flow hood. The media was removed, cells were washed once with fresh SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco) and a further 300 µL of SF900II media (Gibco) plus Antibiotic-Antimycotic (Gibco) added. Baculoviral stock was added to the cells at MOI 5 and cells were incubated at 28°C. At the appropriate time point (24h, 48h, 72h, 96h, 120h, 144h, 168h, 192h) media was removed, and cells harvested from appropriate well and stored separately at -20 until analysis by SDS-PAGE and western blot as in 2.7.3.

Chapter 3: Production of STNV VLPs and protein purification

[This page has been left intentionally blank]

3.1 Introduction

In 2011, Lane *et al* designed a synthetic, codon optimised, STNV CP gene for expression in *E.coli*. Coat protein produced in this expression system spontaneously forms VLPs and when purified, VLPs were shown to be of comparable size and with a similar RNA content, by UV $OD_{260/280}$, to wild type STNV (Lane *et al*., 2011; Ford, 2012; Ford *et al*., 2013). These VLPs can be produced in large quantities (mg/mL) and have previously been purified for use in structural studies (Ford *et al*., 2013; Ford, 2012) and assembly investigations (Bunka *et al*., 2011; Patel *et al.*, 2015). The nucleotide sequence of the codon optimised STNV CP gene is similar to that of the wild type. Sequence alignment using Clustal Omega (Sievers *et al*., 2011) of both wild type and recombinant STNV CP is shown in Figure 3.1, where there is 76.48% sequence identity between the two nucleotide sequences. Additionally, investigations in the original Lane *et al* paper also confirm that the structure of the recombinant protein does not differ significantly from the 1984 wild type structure from Jones and Liljas (Jones and Liljas, 1984) Figure 3.2. The VLPs formed *in vitro* from the recombinant protein can be readily purified, dissociated into CP subunits and reassembled using a specific buffering system involving the chelation of internal divalent metal ions located near the 3 and 5-fold axes (Lane *et al*., 2011; Ford, 2012).

This chapter describes improvements to the purification of STNV VLPs regarding prevention of proteolysis, and investigations into the dissociation, reassembly and packaged nucleic acid component of these *E.coli* derived VLPs, and the selective RNA binding abilities of purified proteins.

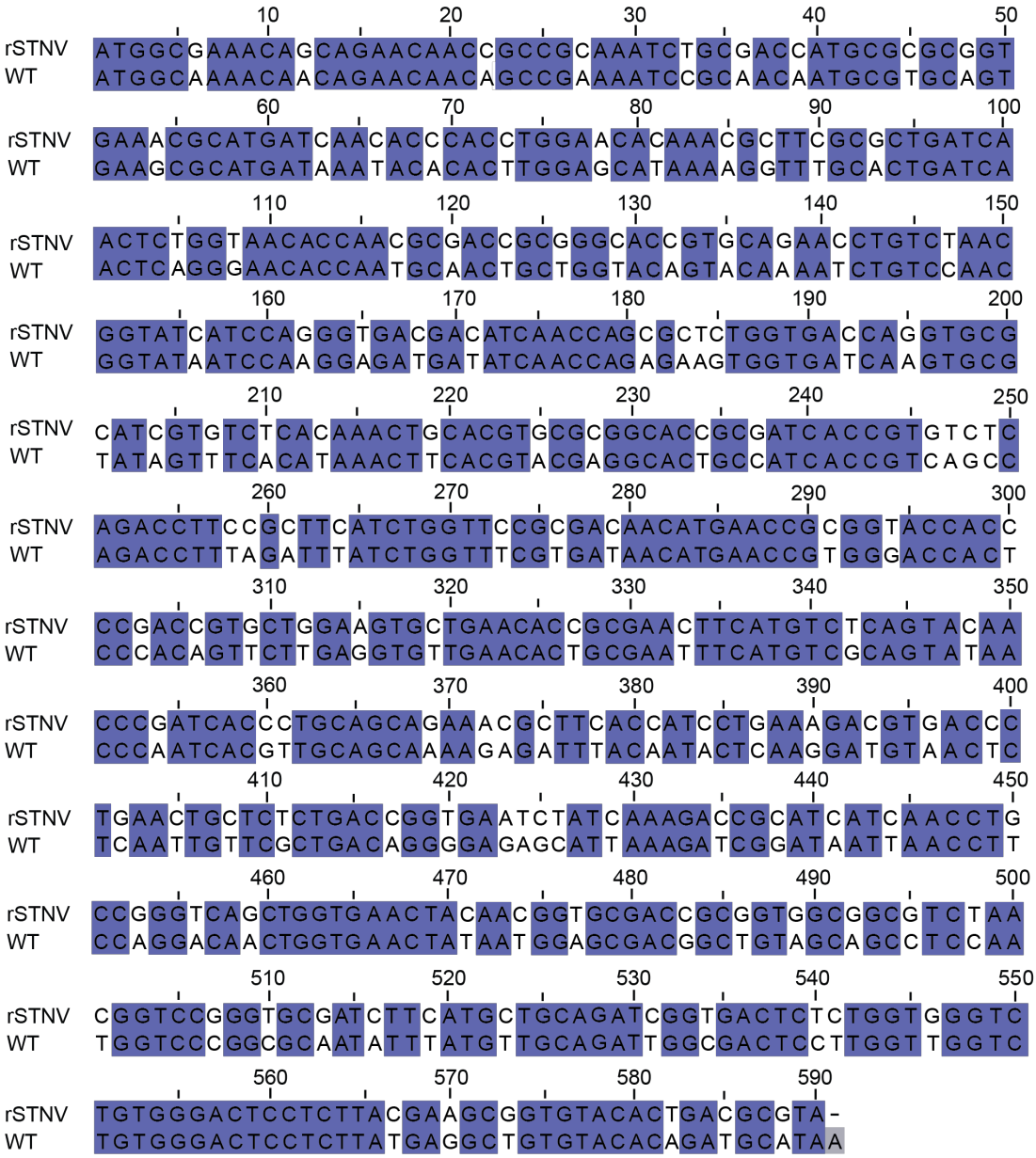


Figure 3.1: Alignment of recombinant (rSTNV) and wild type STNV (WT) coat protein genetic sequences. Sequences were aligned with Clustal Omega (Sievers *et al*., 2011) using default parameters and coloured for sequence identity. There is 76.48% homolgy betweeen the two sequences.

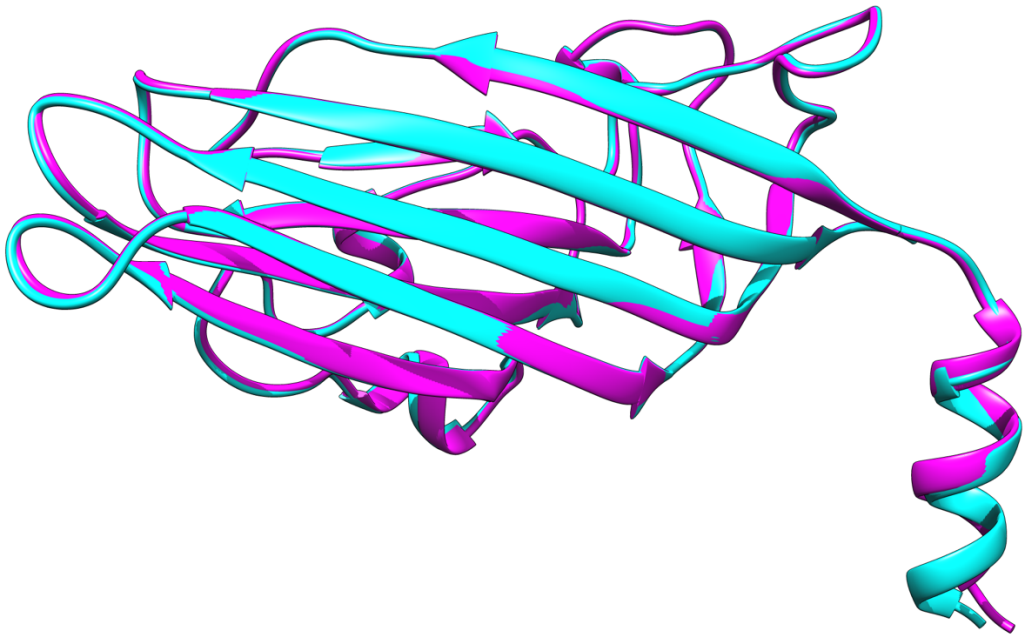


Figure 3.2: Comparison of wild type and recombinant STNV coat protein structures. Recombinant and wild type STNV coat protein structures from PDB, 3SRV (Magenta) and 2BUK (Cyan) respectively, aligned using ChimeraX (Goddard *et al*., 2018). Structures of the 2 proteins, as noted in Lane *et al*, do not differ significantly (Lane *et al*., 2011).

3.2 Results

3.2.1 STNV CP Expression

The STNV CP gene (Materials and Methods) cloned in a pET22b(+) vector (pET22b(+)-STNV) was transformed into BL-21 (DE3) Gold cells (as described in Methods). This *E.coli* strain was selected due to the reduction in 'leaky' expression compared to other systems (Ford, 2012). In this construct the STNV CP gene is under control of a *Lac* UV5 promoter (Novagen, 2003), where a 2 bp mutation in the *lac* -10 region leads to more effective recruitment of RNA polymerase to the promoter increasing transcription of the target gene (Ross *et al*., 1998). Single colonies selected by ampicillin/carbenicillin-

resistance were used to inoculate 10 mL LB starter cultures (37˚C) which were used for subsequent inoculation of 1L flasks (100 µg/mL ampicillin/carbenicillin).

When the optical density at 600 nm (OD_{600}) of the cultures was approximately 0.6, cells were heat-shocked at 42˚C. The heat shock reaction in *E.coli* has two major pathways and leads to expression of *E.coli* heat shock proteins, including DnaK and DnaJ. These proteins are chaperones that aid in protein folding and prevent aggregation (Ellis, 1991; Ellis, 2006). Expression of STNV CP was induced with IPTG, which inhibits the action of the lac repressor (Marbach and Bettenbrock, 2012). Samples were then incubated at 24˚C overnight. STNV CP was detected in the sample via SDS-PAGE analysis of an aliquot taken from 1L cultures with a band at ~21 kDa (STNV CP molecular weight = 21 kDa) indicating induction was successful (Figure 3.3). No STNV CP was detected in the sample taken before the addition of IPTG. After overnight incubation the cultures were then centrifuged in order to pellet the *E.coli* containing the STNV VLPs. A sample taken from the supernatant indicates no STNV is present after centrifugation confirming successful pelleting of the *E.coli*. Coat protein is observed in the resuspended pellet (STNV native buffer) however, due to the reduction in volume it appears more concentrated. The resuspended pellet was incubated with 1 mg of lysozyme (Sigma) and 1 EDTA-Free protease inhibitor cocktail tablet (Roche/Fisher) for 30 minutes on ice, prior to sonication on ice.

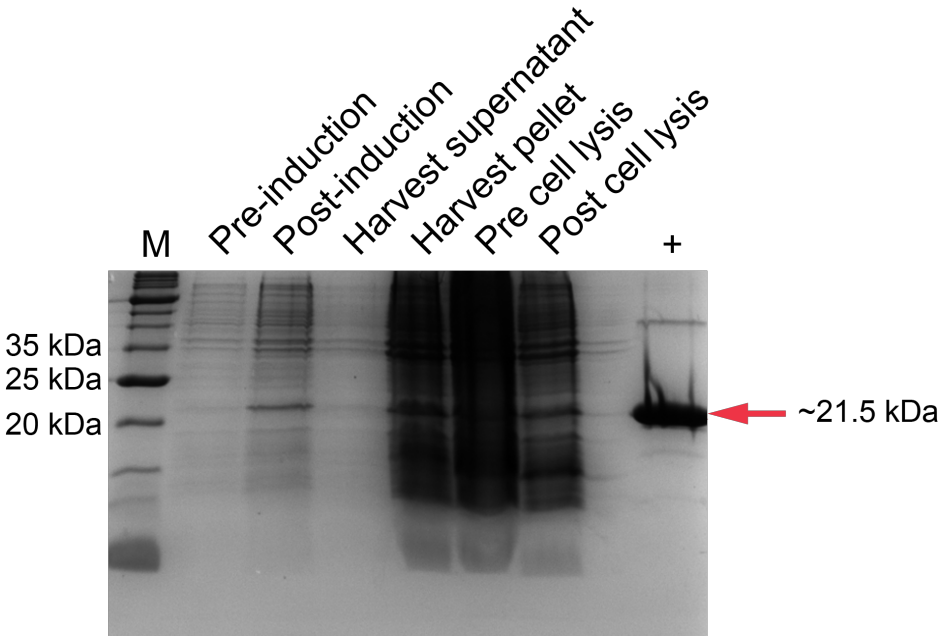


Figure 3.3: SDS-PAGE of STNV expression and purification. SDS-PAGE of STNV coat protein preparation from IPTG induction to cell lysis. Recombinant STNV CP is seen at ~21.5 kDa. + indicates positive control, M indicates NEB broad range protein marker.

Lysozyme incubation was performed in order to make cell lysis more effective by hydrolysing the glycosidic bonds in bacterial cell walls. The cell lysate was then incubated with DNase 1 (Sigma) for one hour at 4˚C, mixed and centrifuged to remove the insoluble *E.coli* components and DNA. Previous work has shown that approximately 50% of the expressed coat protein is present in the soluble fraction while the remainder is pelleted with the cellular debris (Ford, 2012). STNV is detected in the supernatant from cell lysis to a greater extent than the pre-lysed sample suggesting that sonication and incubation with lysozyme caused disruption of the cell wall and release of VLPs. Increasing the salt concentration of the supernatant (30% w/v ammonium sulphate) precipitates the soluble STNV VLPs, causing the proteins to bind to each other in large complexes that are easily precipitated out (Duong-Ly and Gabelli, 2014). STNV VLPs are subsequently collected via

centrifugation and the supernatant discarded as it contains no appreciable STNV VLPs. The VLPs were then resuspended in STNV native buffer (1 mM DTT and protease inhibitors) and dialysed overnight against 3 L of STNV native buffer (1 mM DTT and protease inhibitors) to remove the ammonium sulphate.

3.2.2 STNV VLP and coat protein purification 1

For VLP purification 6 linear sucrose density gradients were prepared using 15-45% (w/v) sucrose in 38 mL tubes (Seton). Approximately 1 mL of sample, post ammonium sulphate precipitation and dialysis, was loaded on top of each gradient. Gradients were centrifuged for 21 hours 15 minutes at 51610 x *g* at 4˚C in a SW32 ti rotor and allowed to stop without braking to avoid gradient disruption. The gradients were subsequently fractionated using a Biocomp gradient fractionator where sample absorbance at 260 nm was measured (Figure 3.4A). The absorbance trace shows two peaks that absorbed at 260 nm, which were subsequently analysed via SDS-PAGE (Figure 3.4B). An STNV CP positive control is indicated (+) and used for comparison. Contaminating *E.coli* proteins carried over from ammonium sulphate precipitation, including some VLPs are collected in fraction 1, whereas purified VLPs are seen in fraction 2 and to a lesser extent fraction 3. Samples were dialysed against 5 L, STNV native buffer to remove sucrose from the sample. Following dialysis, the sample was concentrated using Vivaspin20 columns prior to further experimentation.

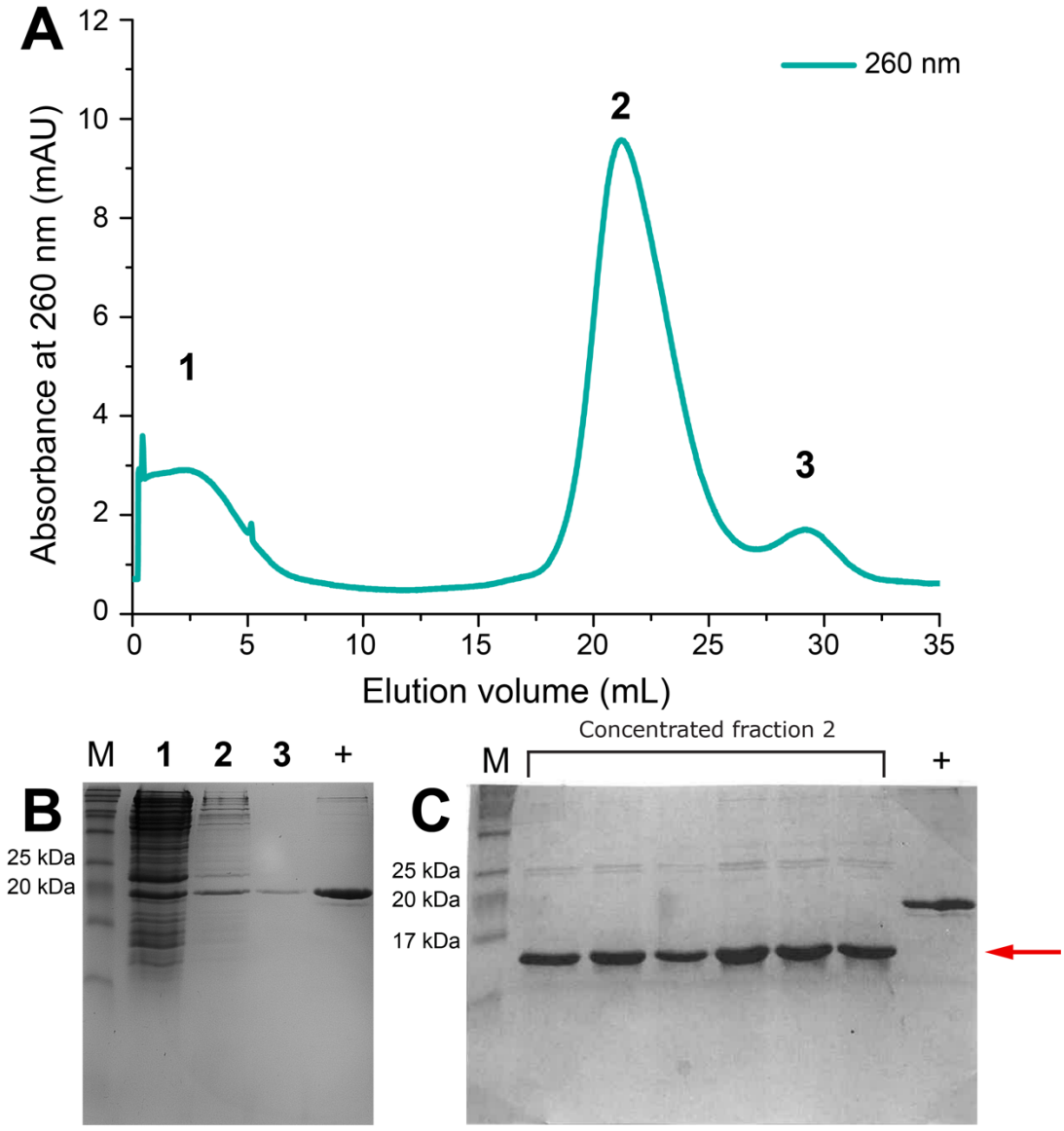


Figure 3.4: Sucrose density gradient frationation of STNV VLPs. (A) Sucrose density gradient fractionation profile of STNV VLP purification. Overlaid absorbance traces at 260 nm for 6 STNV sucrose density gradients. **(B)** SDS-PAGE analysis of peaks 1-3. M indicates NEB broad range protein marker. **(C)** SDS-PAGE of concentrated fraction 2 from all gradients after sucrose density purification. M indicates NEB broad range protein marker and + indicates positive control.

3.2.3 Strategies to prevent cleavage

When the concentrated protein samples were again analysed by SDS-PAGE cleavage of the CP was observed samples (Figure 3.4C). Previously purified protein stored in glycerol is used for the positive control, from which cleavage is evident as the major product at approximately 15 kDa in all samples. Cleavage of the 26 amino acids that make up the N-terminal arm would result in an 18481.86 Da fragment, therefore cleavage of a larger fragment (~60 amino acids) seems to occur. The STNV CP is very proteolytically sensitive. Using the peptide cutter tool from ExPASy (Gasteiger *et al*., 2005) the amino acid sequence can be cleaved at 159 different points by 21 enzymes.

A number of different measures were taken to prevent coat protein cleavage during the purification protocol. Firstly, EDTA-free protease inhibitor tablets (Roche) were added to all buffers however this, was not sufficient to eliminate protein cleavage in the samples (not shown). EDTA-free protease inhibitor cocktails were switched from Roche to Pierce as the latter contained Aprotinin, AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride), Bestatin, Leupeptin, E-64 and Pepstatin A. These components are effective inhibitors of serine proteases, cysteine proteases, aspartic acid proteases and aminopeptidases (Pierce, 2014), whereas the Roche tablets are only effective against serine and cysteine proteases (Roche, 2004). Cleavage of samples however, remained an issue. This could be due, in part, to the concentration of protease inhibitors used. The Roche and Fisher recommended guidelines state that one tablet is to be used per 50 mL (Pierce, 2014; Roche, 2004). Protease inhibitor tablets were used in this instance at 1 tablet per litre of buffer

as the recommended concentration was prohibitively expensive. Additional protease inhibitors were therefore added to buffers to supplement the cocktail tablets. PMSF, aprotinin and pepstatin A were added to all buffers at the working concentrations shown in Table 3.1. Communal equipment in use by other lab members/faculty for protein purification i.e. AKTA explorer and Biocomp fractionator were cleaned thoroughly prior to use and new cleaning protocols were implemented. Additionally, an alternative purification protocol was devised utilising both size exclusion and ion exchange chromatography which proved successful in purifying STNV VLPs and coat proteins without cleavage.

Table 3.1: Protease inhibitor activities and working concentrations

Inhibitor	Active Against	Working concentration
Roche Complete-EDTA free	Serine and cysteine	1 tablet/50 mL (Roche, 2004)
Pepstatin A	Aspartyl petidases	0.5-1 µg/mL (Sigma, 2001)
Aprotinin	Serine proteases (Sigma, 2018)	1-2 µg/mL (GoldBiotechnology, 2015)
PMSF	Serine proteases	1.74 mg/mL (Thermo Fisher Scientific, 2015)
Thermo Fisher Complete EDTA free	Serine, Cysteine, theionine and aspartyl	1 tablet/50 mL (Pierce, 2014)

3.2.4 STNV VLP and protein purification 2

The following methods take advantage of the size and charge differences between *E.coli* and STNV VLPs. The molecular weight of a wild type STNV virion is approximately 1.7 MDa (Appendix C); as such they will elute in the void volume of most size exclusion columns (Lane *et al.*, 2011). The coat protein of STNV has an isoelectric point (pI) = 10 (Gasteiger *et al*., 2005) and as such the coat protein should bind to an SP-FF anion resin at neutral pH but should not bind to a Q-FF cation resin.

After concentration by ammonium sulphate precipitation, samples were resuspended in STNV native buffer and dialysed extensively to remove the ammonium sulphate. The sample was subsequently loaded onto an XK 50 Q-Sepharose anion exchange column which had been equilibrated in STNV native buffer. The previous stages in the purification process remained successful. SDS-PAGE analysis shows STNV CP is visible in samples taken after induction through to Q-Sepharose loading (Figure 3.5). The majority of the contaminating *E.coli* proteins that remained in the sample following the concentration step bind to the negatively charged resin, whereas approximately 50% of the STNV VLPs pass through the void volume and are collected in the fractions. The bound proteins, which contain a mixture of STNV VLPs and *E.coli* derived proteins, are eluted from the column with 2M NaCl.

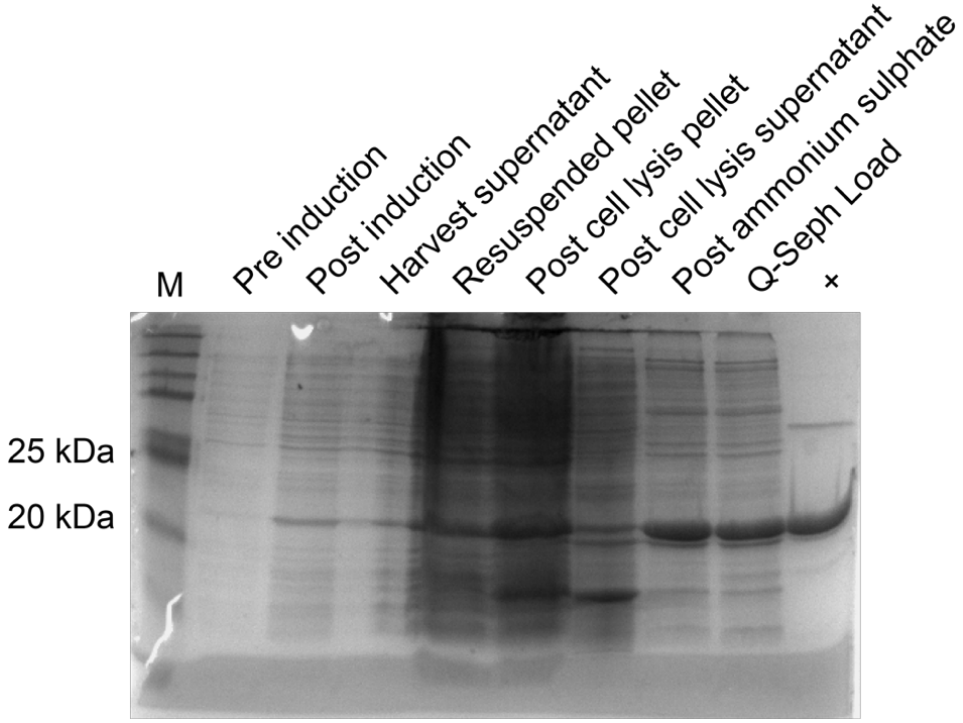


Figure 3.5: SDS-PAGE of STNV purification steps. SDS-PAGE of stages of STNV purification from pre-induction of protein to sample loaded onto XK50 Q-Sepharose column. + indicates positive control and M indicates pre-stained broad range marker (NEB).

Typically, Q-Sepharose anion exchange chromatography produced a 'flow-through' peak which absorbed at both 260 and 280 nm, appearing at ~ 30 mL post injection, followed by another larger peak which was only seen after the column was washed with 2M NaCl (Figure 3.6A). SDS-PAGE analysis of the fractions with 260 and 280 nm absorbance indicated that VLPs were eluted predominantly in the 'flow-through' peak (Figure 3.6B); this was confirmed with EM analysis of an eluted fraction (Figure 3.6C). The secondary peak, eluted with 2M NaCl also contained some STNV VLPs however, these fractions also contained contaminating proteins carried over from *E.coli*. Q-Sepharose anion exchange chromatography was an sufficient preparation step, removing the majority of the contaminating proteins with the added benefit that a large amount of sample could be purified in this manner using the sample pump. The sample loaded onto the column is shown in Figure 3.6B for comparison, from which it is evident that the majority of the *E.coli* cellular components have been separated from the STNV VLPs. These carried over proteins are also seen in EM images of the sample before Q-sepharose purification (Figure 3.7A and B). Fractions from the 'flow through' peak were pooled and concentrated via ammonium sulphate precipitation. The 2M NaCl elution peak was discarded as contaminating *E.coli* proteins were a potential source of proteases.

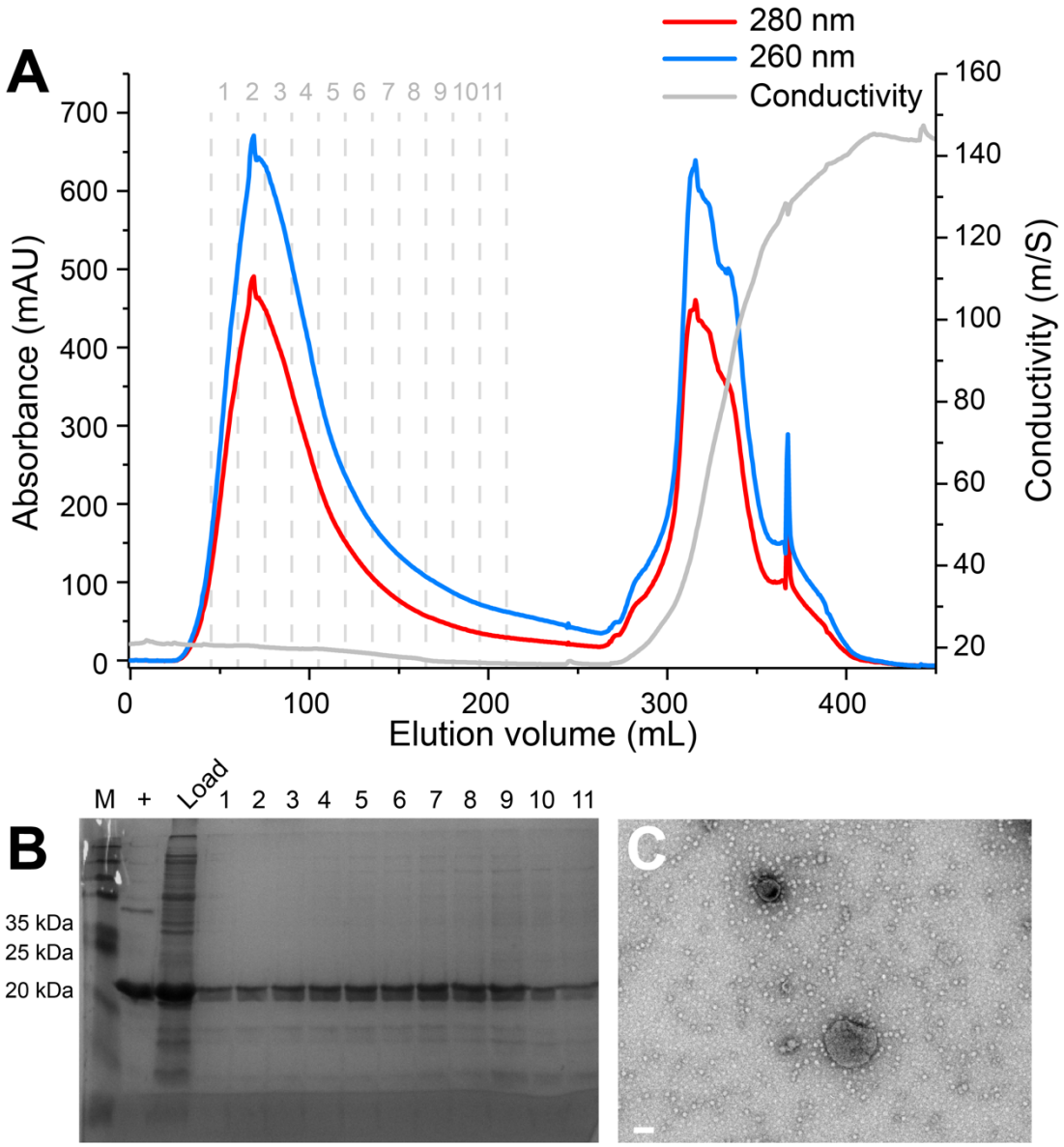


Figure 3.6: Q-Sepharose purification of STNV VLPs. (A) Absorbance traces at 260 and 280 nm of STNV VLP purification using an XK-50 Q-Sepharose anion exchange column. The initial peak, eluting from ~30mL to 200mL, is the 'flow through' peak containing the STNV VLPs. The second peak is eluted upon the addition of 2M NaCl. **(B)** SDS-PAGE showing fractions with absorbance at 260 and 280 nm where M indicates marker and + indicates STNV positive control. Fractions from the column are shown alongside an aliquot taken from the loaded sample for comparison. **(C)** Negative stain TEM image at 15,000X magnification showing STNV VLPs.

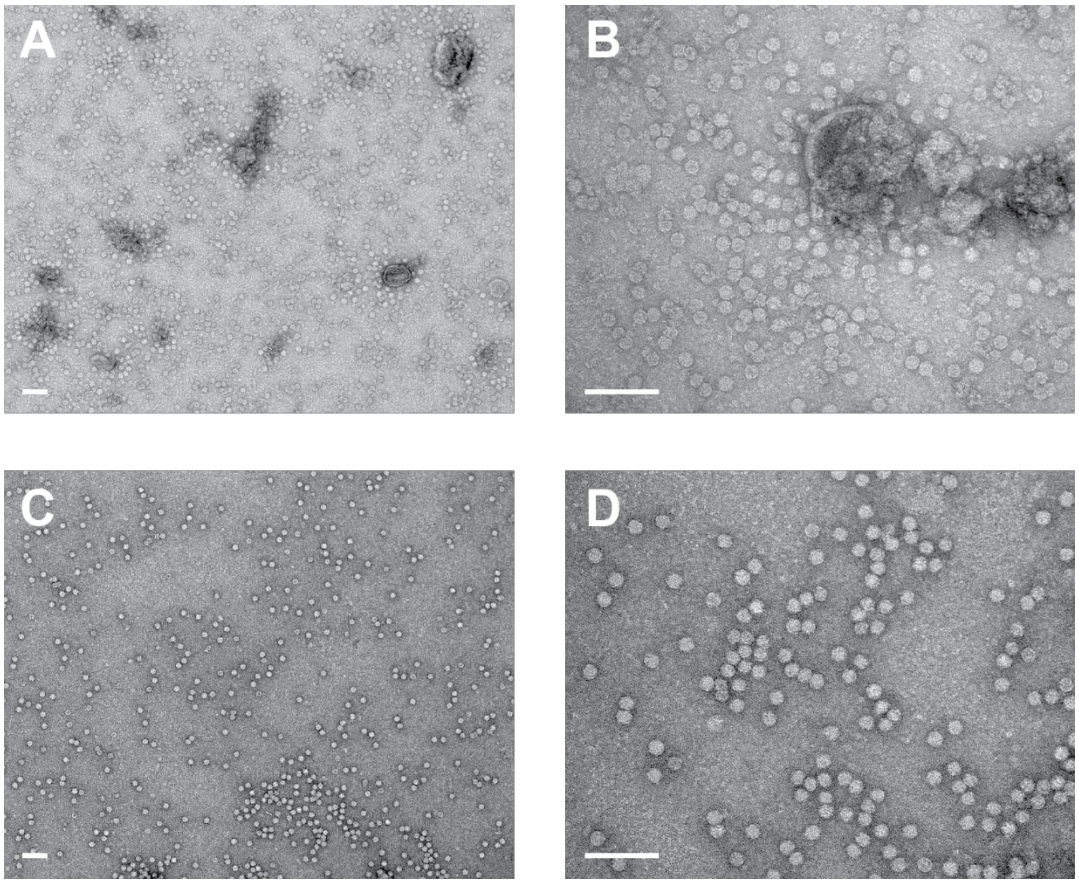


Figure 3.7: Negative stain TEM images of STNV VLPs during purification: (A) Negative stain TEM image of STNV VLPs at 15000X magnification after ammonium sulphate precipitation. **(B)** The same sample at 40000X magnification. **(C)** Negative stain TEM image at 15000X magnification of a Q-Sepharose fraction. **(D)** The same sample at 40000X magnification. Scale bars are 100 nm.

Size exclusion chromatography (SEC) was then carried out using 500 µL injections on a Superdex-200 increase 10/300 GL column at a flow rate of 0.3 mL/min and collected in 1.5 mL fractions (Figure 3.8A). Fractions that showed absorbance at 260 and 280 nm were analysed by 15% SDS-PAGE (Figure 3.8 B) and negative stain TEM to confirm the presence of VLPs (Figure 3.8C). The sample started eluting from the column ~ 9 mL post injection and stopped eluting at ~ 11 mL (Figure 3.8A). STNV containing fractions were pooled, quantified and stored for use in subsequent experiments. $A_{260/280}$ ratio of samples at this point was approximately 1.7 which is in accordance with previously stated values for STNV VLPs produced in *E.coli* (Lane *et al*., 2011; Ford *et al*., 2013; Patel *et al*., 2015). SDS-PAGE analysis (Figure 3.8B) carried out on fractions that absorbed at 260 and 280 nm shows one species and suggests that little or no proteolysis had taken place during the purification process. Additionally, negative stain TEM images of loaded sample, a single fraction and pooled fractions demonstrate purified VLPs are homogenous in size and shape (Figure 3.9).

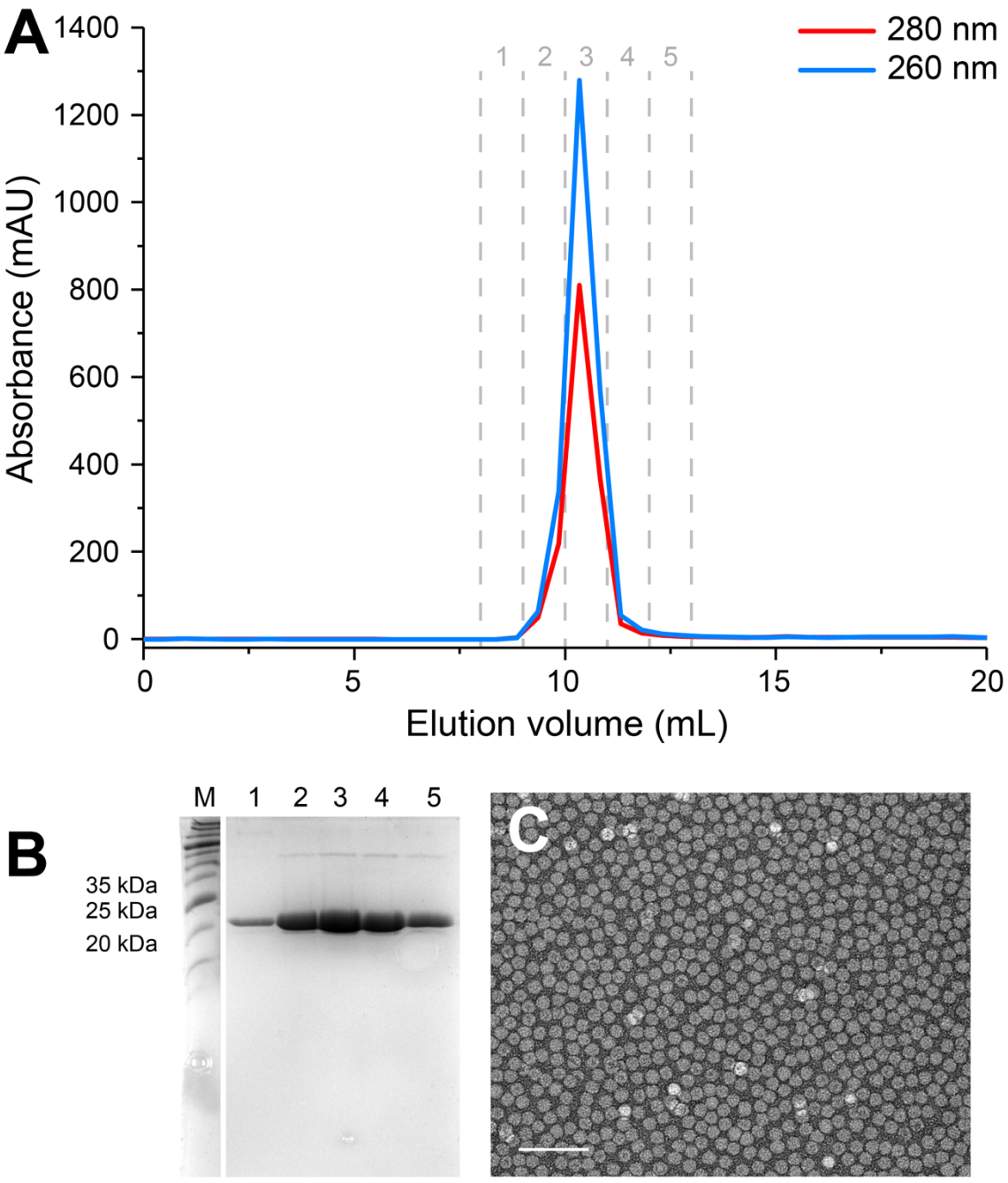


Figure 3.8: Size exclusion chromatography and characterisation of recombinant STNV VLPs. (A) Absorbance traces at 260 and 280 nm of STNV VLP purification using a superdex 200 increase 10/300. Peaks were observed approximately 9 mL post injection and finished eluting by 11 mL. **(B)** SDS-PAGE showing fractions with absorbance at 260 and 280 nm where M indicates marker. **(C)** Negative stain TEM image at 30,000X magnification showing STNV VLPs.

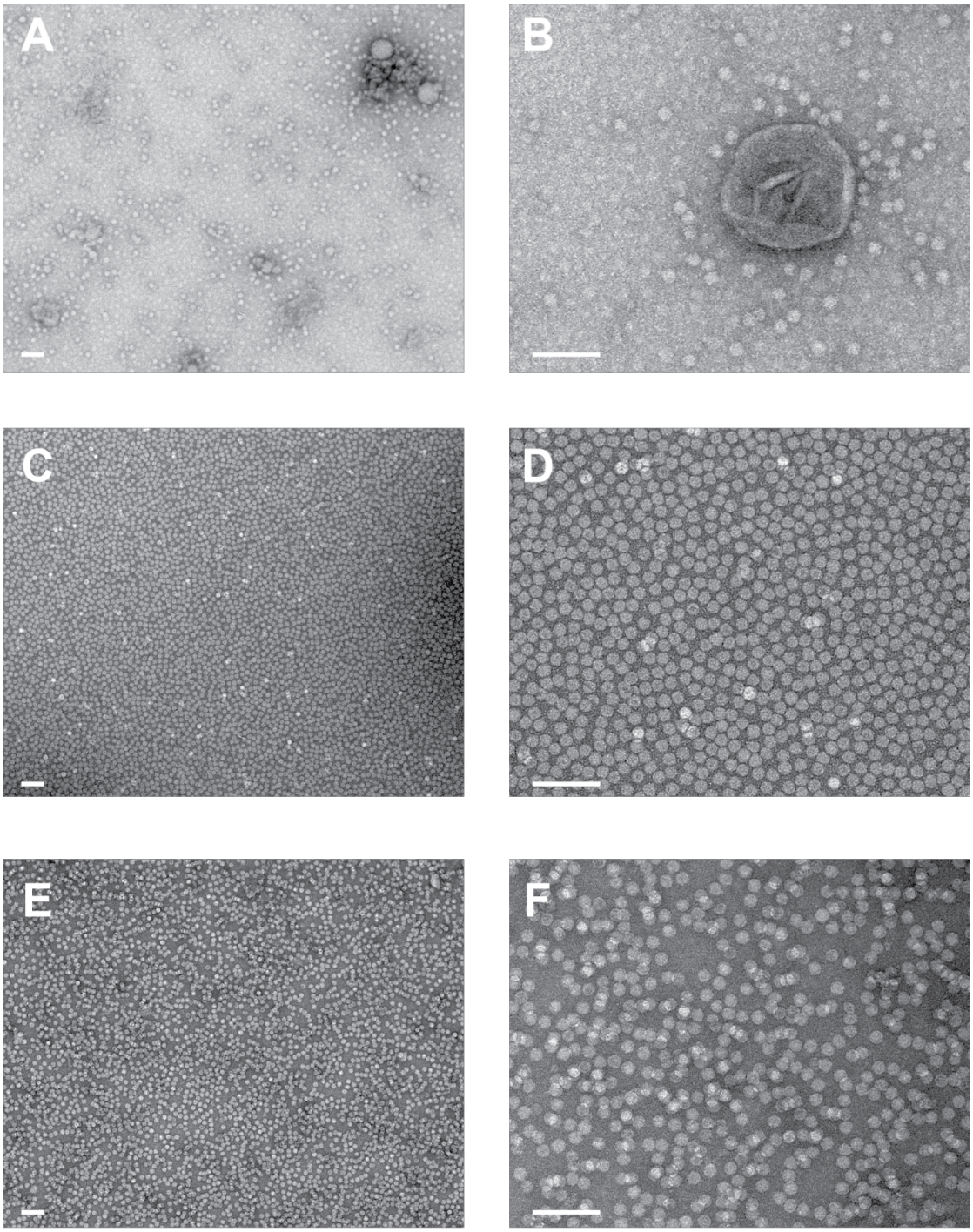


Figure 3.9: Negative stain TEM images of STNV VLPs during purification (2). (A) Negative stain TEM image of sample after ion exchange chromatography at 15000X magnification. **(B)** The same sample at 40000X magnification. Some contaminating *E.coli* proteins are carried over from the initial purification step. **(C)** Fraction after size exclusion chromatography at 15000X magnification. **(D)** The same sample at 40000X magnification. **(E)** Pooled SEC fractions at 15000X magnification. **(F)** The same sample at 40000X magnification. Scale bars = 100 nm.

3.2.5 Charge detection mass spectrometry

The $OD_{260/280}$ ratios of VLPs produced using the recombinant *E.coli* system are comparable to that of wild type STNV, which suggests that they probably contain similar amounts of RNA even though these particles do not contain the wild type STNV genome. Previously, nucleic acid, extracted from STNV VLPs and amplified using primers specific to the synthetic STNV gene, produced a product of approximately 500 bp; the approximate size of the STNV-1 CP mRNA. This product however, was stated to only account for an estimated 20% of the encapsidated material. The remainder was postulated to be smaller RNAs that could possibly be mRNA precursors, products of degradation or cellular (*E.coli)* RNAs (Lane *et al*., 2011). The synthetic STNV shares ~76% sequence homology at the nucleotide level (Goujon *et al*., 2010; Sievers *et al*., 2011) (Figure 3.1) suggesting that there is some sequence specificity to recombinant VLP assembly (Lane *et al*., 2011). The coat protein of WT STNV has a mass of 21,715 Da (Gasteiger *et al*., 2005) and as a *T*=1 capsid, each virus particle consists of 60 copies of this protein arranged around a 1,239 nt ssRNA genome stabilised by 92 internal calcium ions (Lane *et al*., 2011). This gives WT STNV particles a theoretical mass of ~1.7 MDa (Appendix C).

It has been suggested that the limit of detection when measuring the mass to charge ratio (m/z) is 20 MDa in standard MS (Snijder *et al*., 2013), this limit is reduced for samples which are inherently heterogeneous. As virus capsids have a consistent mass and are composed of a number of identical proteins multi-MDa range mass spectrometry (MS) has been performed on a select number of virus particles such as MS2 (Tito *et al.*, 2000), Hepatitis B virus

(Uetrecht *et al*., 2008) and Hong Kong 97 virus (HK97) (Snijder *et al*., 2013); though, for both MS2 and HK97 the m/z spectra had poor resolution. Electrospray is the method of choice for larger species as it produces ions with higher charge and as a consequence a lower m/z. However, highly charged ions are more difficult to resolve and generate a large amount of kinetic energy. In addition, larger species with higher m/z are not well suited to the detectors used in conventional instruments (Keifer *et al.*, 2017). However, charge detection mass spectrometry (CDMS) allows for mass determination of biological systems that are typically out of reach for traditional electrospray MS techniques. Purified STNV VLPs from *E.coli* were analysed by Dr Elizabeth Pierson and Corinne Lutomski via CDMS at the University of Indiana (Bloomington) where a sharp peak was observed at 1.683 MDa (Figure 3.10). As previously mentioned, WT STNV particles have an estimated mass of ~1.7 MDa.

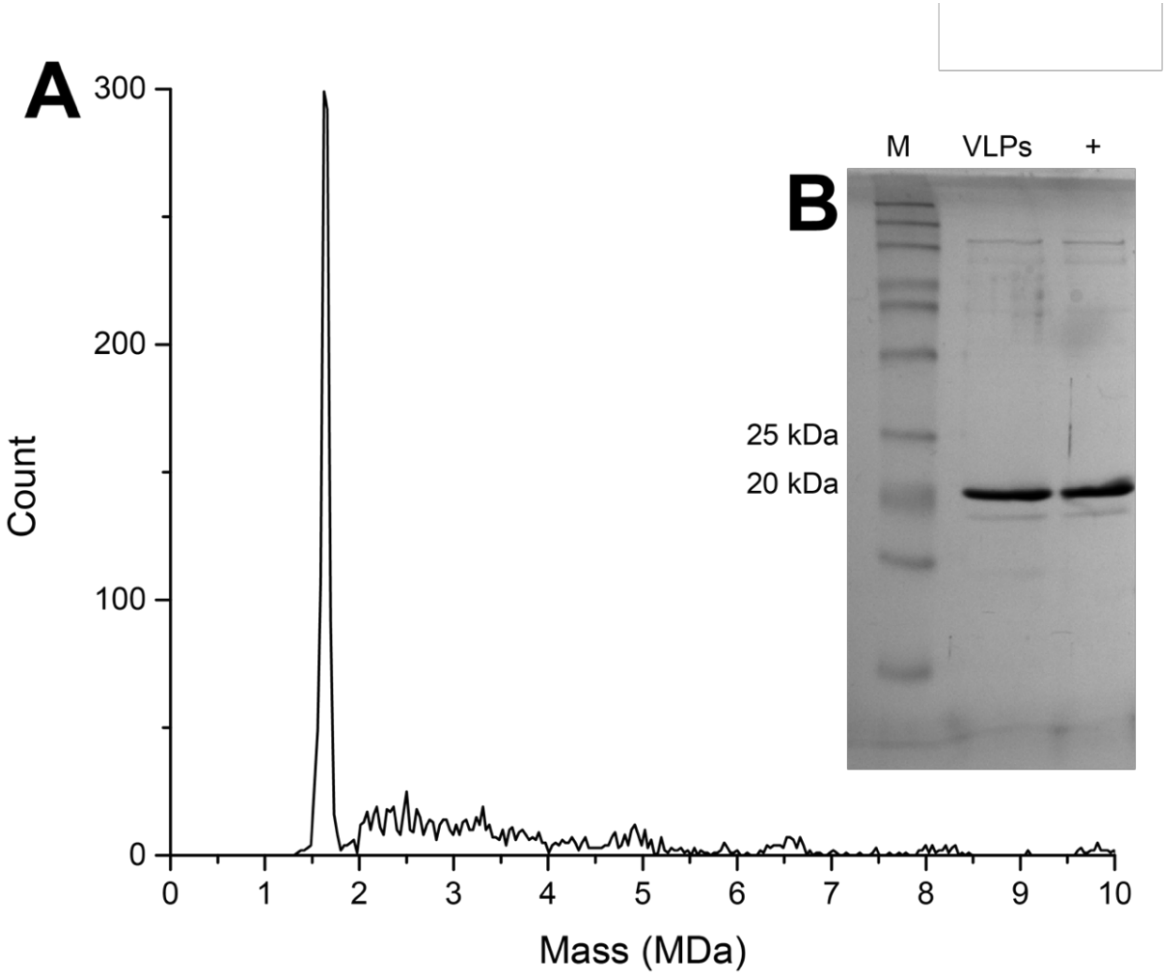


Figure 3.10: Charge detection mass spectrometry results for recombinant STNV. **(A)** Charge detection mass spectrometry (CDMS) of purified recombinant STNV VLPs. **(B)** SDS-PAGE of recombinant STNV sent for CDMS analysis. M indicates prestained protein marker (Geneflow) and + indicates purified recombinant STNV CP stored in glycerol.

3.2.6 Disassembly and reassembly of recombinant STNV

Recombinant *E.coli* STNV VLPs can be disassembled and reassembled *in vitro*. When VLPs are dialysed into a buffer containing EDTA pH 8.5 the *T*=1 particles expand and calcium ions that stabilise the interior are removed. Once subjected to dialysis in reassembly buffer containing 3 mM $CaCl_2$, the *T*=1 VLPs reform. This is similar to the dissociation and reassembly protocols devised for other plant viruses such as turnip crinkle virus (TCV) (Sorger *et al*., 1986). Once STNV is dissociated however, the coat protein becomes very proteolytically sensitive. EDTA-expanded STNV VLPs treated with trypsin

produce one cleavage product as a result of cleavage of Arg28 just beyond the N-terminal helix (Unge *et al*., 1980). STNV CP truncated mutants which lack these initial 12 residues (STNVΔ12) are assembly incompetent, highlighting the importance of the N-terminal region during capsid assembly (Bunka *et al*., 2011).

3.2.6.1 One pot disassembly and reassembly of recombinant STNV VLPs

0.5 mL of recombinant STNV VLPs (2.12 mg/mL) were dialysed into STNV disassembly buffer. After the sample had dialysed for 5 hours at 4˚C, reassembly was achieved by transferring the dissociated particles into reassembly buffer and dialysing for a further 18 hours at 4˚C. Disassembled recombinant STNV and reassembled recombinant STNV samples were analysed by negative stain TEM (Figure 3.11A and B). No VLPs are present in the sample removed after 5 hours in disassembly buffer, indicating that dissociation of recombinant particles was successful. After 18 hours in reassembly buffer VLPs are visible with a diameter of ~ 20 nm suggesting that reassembly of previously dissociated particles was achieved.

3.2.6.2 Dissociation and reassembly in the presence of α,2-macroglobulin (α2M)

Found in the circulation of vertebrates and in the whites of bird and reptile eggs, α2M is an 800 kDa homotetrameric plasma protein which functions as a molecular proteinase trap. A-macroglobulins possess a ‘bait region’ which contains several motifs recognised by proteases as targets for cleavage. As a protease recognises and subsequently cleaves this area, α2M undergoes a

conformational change, trapping the protease and rendering it inactive (Borth, 1992). α2M can capture any protease regardless of class and is therefore described as a 'panprotease inhibitor' (Sottrup-Jensen, 1989).

To determine whether α2M effectively inhibited the proteolysis of STNV CP, 0.25 mL of STNV VLPs (2.12 mg/mL) were dialysed in STNV dissociation buffer containing additional protease inhibitors, detailed in Table 3.1. After dialysis for 5 hours at 4˚C, reassembly was achieved by taking the disassembled particles, adding 1 unit of α2M, and dialysing the sample in STNV reassembly buffer (plus additional protease inhibitors) for 18 hours at 4˚C. Dissociated VLPs and reassembled VLPs were analysed by negative stain TEM (Figure 3.11C and D). A sample reassembled using the same method in the absence of α2M forms particles when dialysed in reassembly buffer (Figure 3.11A and B). However, in the presence of α2M there appear to be no particles. As these experiments were carried out simultaneously, this difference suggests that the presence of α2M in the sample causes this difference in behaviour. This concentration of α2M was selected as it had previously been used in STNV reassembly. Results from these investigations using EM analysis suggest that in the presence of α2M, STNV becomes aggregated. A2M's targets are not always proteases as it has been suggested that α2M and STNV CP aggregate under these conditions (Ford, 2012).

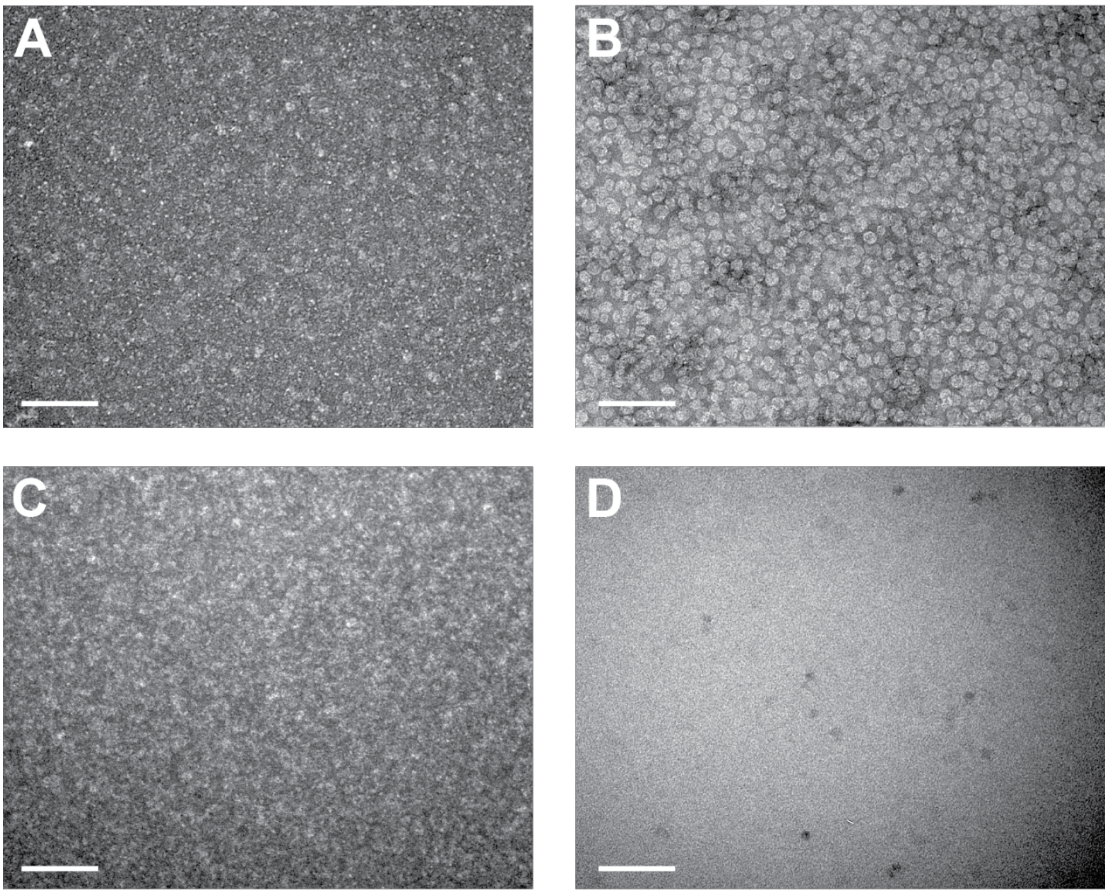


Figure 3.11: Negative stain TEM images of STNV dissociation and assembly in the presence and absence of α,2-macroglobulin (α2M). (A) Negative stain TEM image of dissociated sample in the absence of α2M. **(B)** Negative stain TEM image of reassembled STNV VLPs in the absence of α2M. **(C)** Negative stain TEM image of dissociated sample in the presence of α2M. **(D)** Negative stain TEM image of reassembled STNV VLPs in the presence of α2M. Scale bars = 100 nm

3.2.7 STNV CP purification

To purify the CP for reassembly reactions and investigate how RNA is linked to capsid assembly, this packaged material must first be removed. Reducing the calcium ion concentration with 100 mM EDTA (pH 8.5) had previously been shown to cause STNV VLP disassembly (Ford, 2012). 15 mL STNV VLPs (1 mg/mL) was incubated with 15 mL 10X dissociation buffer for 1 hour at 4˚C, rolling and 1µg/mL pepstatin A. The dissociated sample was loaded onto pre-equilibrated (1X dissociation) Q-FF and SP-FF ion exchange columns in series at a flow rate of 1 mL/min in 10 mL injections. The columns were then flushed with 15 mL STNV dissociation buffer. The use of these columns in series allows for the separation of STNV CP from packaged genetic material. As the Q-FF column is a strong anion exchanger, RNA will bind to the resin. As the protein has an overall positive charge it will not bind to the Q-FF but will bind to the cationic SP-FF column. The Q-FF column is always placed above the SP-FF column. After washing, the Q-FF column was then removed to avoid contamination with RNA. The SP-FF column was then washed with a further 15 mL of STNV dissociation buffer. CP was eluted from the SP-FF column with a 20 mL gradient of STNV buffer B. STNV CP eluted from the column at approximately 40% ‘B’ and was collected in 0.5 mL fractions. Figure 3.12A shows the absorbance trace at 280 nm for one injection and the peak of eluted coat protein. Samples that absorbed 280 nm were analysed by SDS-PAGE (Figure 3.12B) and negative stain TEM of an eluted fraction (Figure 3.12C) indicate that the VLPs were successfully dissociated and 260/280 ratios of ~0.5 indicate pure protein.

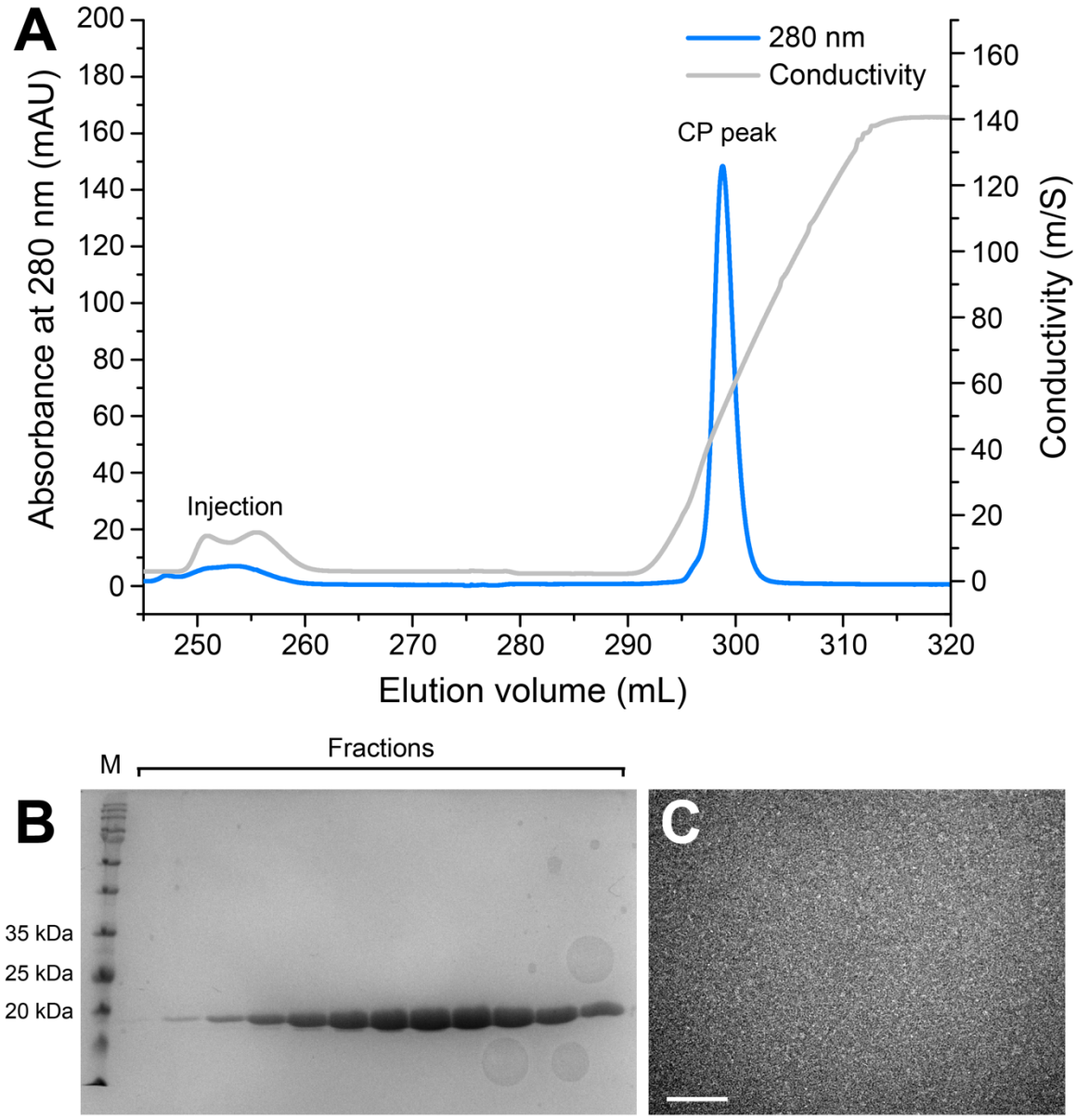


Figure 3.12: Purification of STNV coat protein. STNV VLPs were disassembled and purified using Q-Sepharose and SP ion exchange columns in series. The RNA will bind to the Q-sepharose resin and the STNV CP will bind to the SP resin. The protein is eluted from the column with a gradient of 2M NaCl in dissociation buffer. **(A)** Chromatogram showing absorbance at 280 nm of one injection and gradient of eluted protein. **(B)** SDS-PAGE of fractions with absorbance at 280 nm. M indicates prestained ladder (Geneflow). **(C)** Negative stain TEM image at 40,000X magnification of fraction with 280 nm absorbance. Scale bar = 100 nm.

3.2.8 Reassembly with B3 aptamer

In order to determine if the purified CP was assembly competent, STNV CP was used for reassembly reactions with B3 in different RNA:CP ratios and in the absence of RNA. 1 nM B3 and either 3 or 5 nM purified STNV CP and 3 nM CP only (100 µL reaction volume), was dialysed in STNV reassembly buffer at 4˚C, for 18 hours. Samples were visualised using negative stain TEM at 40000x magnification (Figure 3.13A and B). Reassembled VLPs have a similar morphology and diameter (~20 nm) to that of wild type STNV and recombinant VLPs produced in *E.coli*. In the absence of RNA (Figure 3.13C) VLPs were not observed, this is in accordance with past results and further suggesting that reassembly is dependent on RNA binding. Between the two coat protein-RNA ratios (3 or 5:1) there appeared to be no difference in the assembled particles.

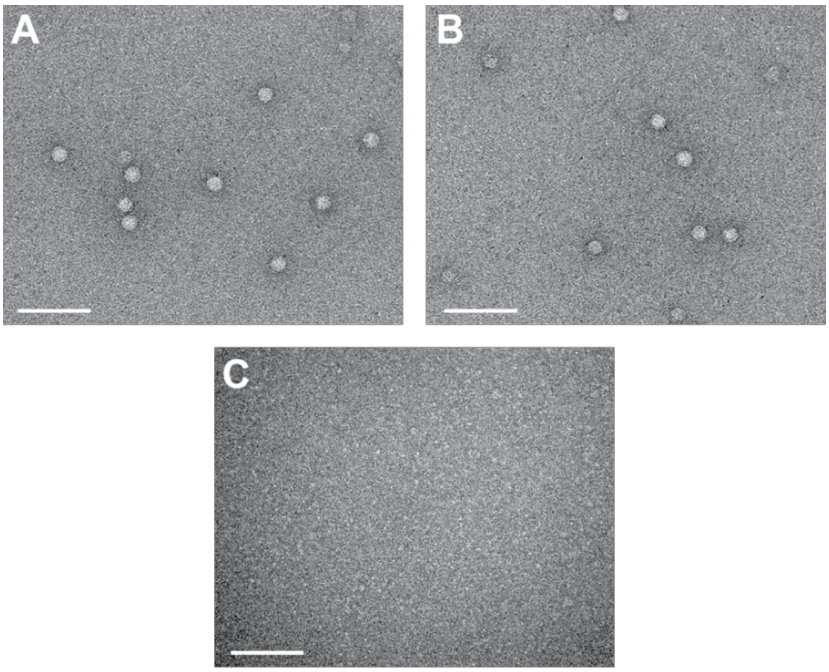


Figure 3.13: Negative stain TEM images of STNV VLP reassemblies with B3 aptamer at 1:3 and 1:5 RNA:CP ratios. (A) Image of reassembled STNV VLPs in a 1:3 RNA:CP ratio at 40000X magnificaiton. **(B)** Image of reassembled STNV VLPs in a 1:5 RNA:CP ratio at 40000X magnificaiton. **(C)** Image of STNV CP dialysed into reassembly buffer without RNA. Scale bars = 100 nm.

3.2.9 Anisotropy B3 vs TR

Anisotropy investigations were carried out with B3 and MS2 TR. The anisotropy of 5 nM B3 and TR was measured before the addition of STNV CP. Both samples initially bound to the STNV CP however, on the titration of further protein the TR-CP complex appears to dissociate whereas the B3-CP remains stable (Figure 3.14A). This suggests that TR is initially able to bind the STNV-CP, which may be due to the similar motif presented on the stem loop but is unable to assemble beyond this point and dissociates. It may be that although TR is capable of binding it cannot fulfil certain conditions that facilitate capsid assembly, for example correct orientation of stem loops. B3 however, binds STNV CP at ~ 5 nM. The R_h of an STNV virion is ~11.3 nm (Patel *et al*., 2015) and the values obtained in smFCS suggest that full capsid assembly was achieved within the coat protein titration (Figure 3.14B). The end products of the B3 titration were analysed by negative stain TEM where particles of a comparable size to wild type and recombinant STNV were visible (Figure 3.14C).

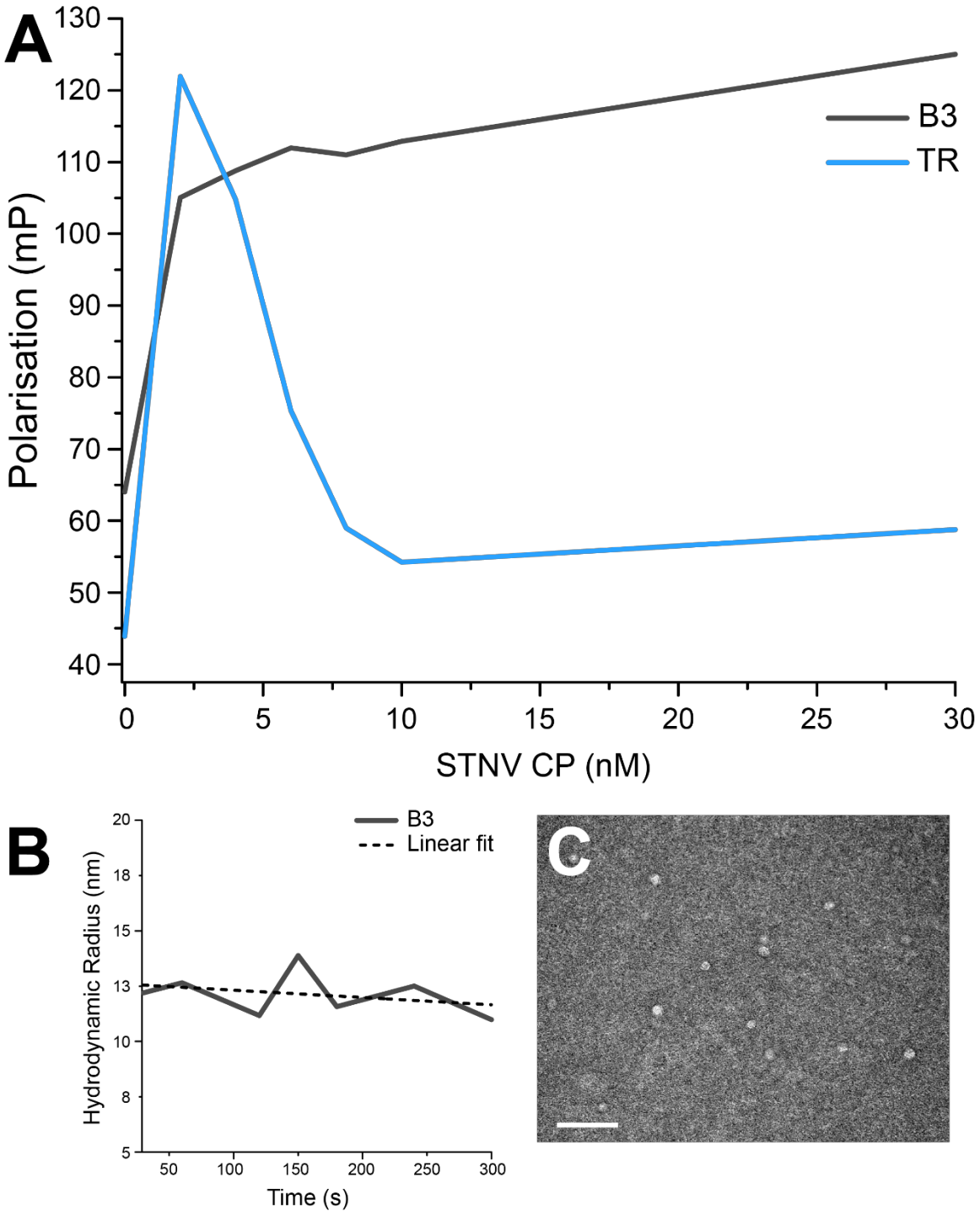


Figure 3.14: Fluorescence anisotropy of STNV CP titrated into A448-labelled B3. **(A)** Plots of fluorescence polarisation against STNV CP concentration for B3 and TR. Although both B3 and TR bind the STNV CP initially, TR dissociates while B3 remains bound. **(B)** Hydrodynamic radius measurements of B3 titration sample. R_h of STNV VLP = 11.3 nm. **(C)** Negatively stained TEM image of the B3 titration sample at 40,000X magnification. Scale bar = 100 nm.

3.2.10 Light scattering of recombinant VLPs and Purified CP

Synge and Tiselius were the first to postulate separation on the basis of size using chromatography after observing exclusion of molecules from the pores of zeolites. This 'molecular sieve' was the forerunner to SEC that is routinely used today for the characterisation of proteins and other macromolecules (Hong *et al*., 2012). SEC however, requires reference standards of known size and molecular weight to put results in context. A SEC system coupled with multiangle light scattering (MALS) allows for the analysis of each eluted fraction by three methods sequentially. Whereas in traditional batch MALS the final measurements are an average of the components in mixed samples. In SEC-MALS the sample is initially separated by virtue of size and absorbance is measured for each eluting species. Subsequently molar mass and size can then be determined for each individual element (Wyatt Technology, 2018a).

A mixture of purified recombinant VLPs and coat protein were loaded onto a pre-equilibrated Superdex 200 10/300 GL size exclusion column (GE Healthcare) connected to a HELEOS II (Wyatt technology) at a flow rate of 0.5 mL/min and collected in 0.5 mL fractions. As shown in figure 3.15 the Superdex 200 column is able to separate the mixed species into intact VLPs in peak 1 eluting at ~8-10 mLs, dissociated coat protein in peak 2 eluting at ~12.5-14 mLs and protease inhibitors in peak 3. Absorbance measurements of the VLP peak show a shoulder at the leading edge which is typical of STNV VLPs on this Superdex 200 column. From the light scattering it can be inferred that this peak likely contains aggregated particles however, this species makes

up a small percentage of the total peak. Comparatively, the dissociated coat protein peak has a lower absorbance at 280 nm which reflects the concentrations that were used in this investigation. The peak suggests that the protein has remained monomeric and the light scattering measurements confirm that as a smaller particle the purified coat protein scatters less light.

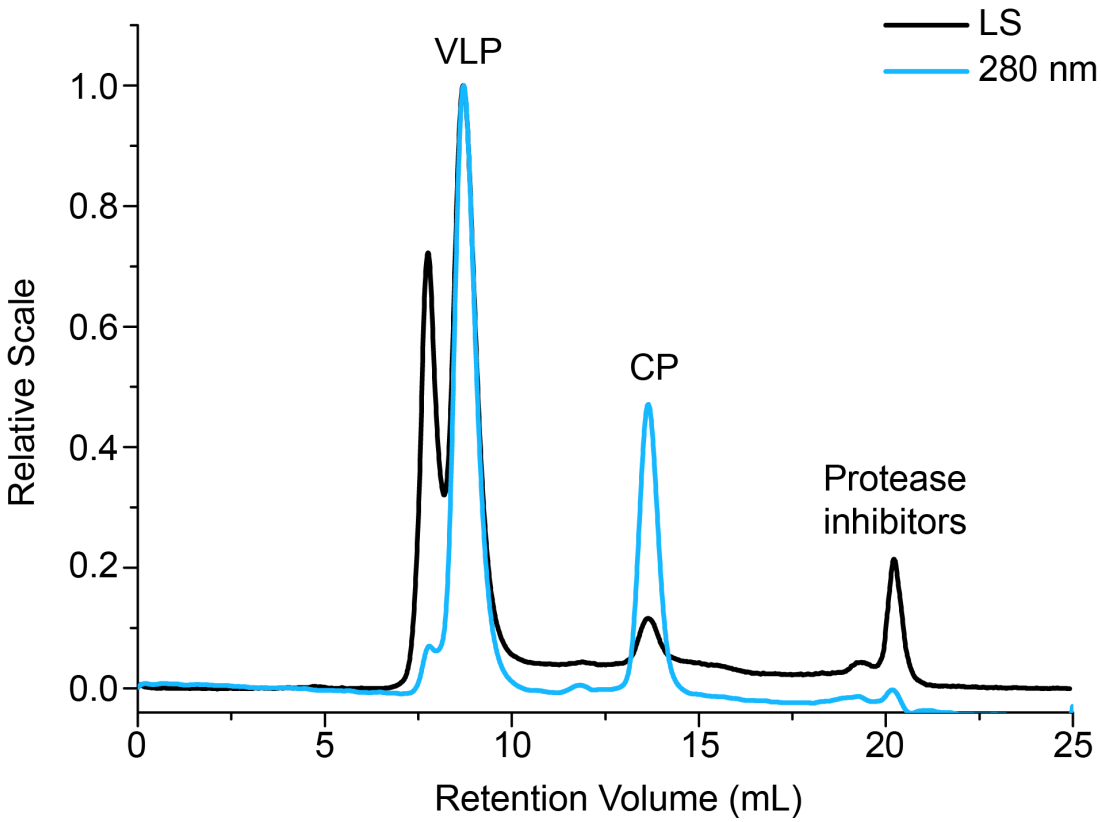


Figure 3.15: Dynamic light scattering of recombinant STNV VLPs and purified recombinant CP. Absorbance at 280 nm and LS of recombinant STNV VLPs and purified coat protein shown on a relative scale from Astra SEC-MALS software.

3.3 Discussion

This thesis utilises two distinct methods for the purification of STNV VLPs. These two methods diverge after ammonium sulphate precipitation of VLPs from the cell lysis supernatant. In the original purification method, the resuspended pellet was purified using sucrose density gradient sedimentation and the VLPs were dissociated and separated using both Q-FF and SP-FF columns in series. As seen in figure 3.4 cleavage of expressed protein was observed after the initial sucrose density purification step. Methods to suppress this cleavage were undertaken in the cleaning of communal equipment and the use of different protease inhibitors. Cleavage however, remained an issue and an alternative purification method was undertaken. Purification of recombinant STNV using Q-Sepharose ion exchange followed by SEC in combination with pepstatin A was altogether more successful in eliminating the cleavage of the protein and allowed particles and proteins to be used in downstream processes. Q-Sepharose ion exchange removed many of the contaminating proteins carried over from expression in *E.coli* with SDS-PAGE (Figure 3.6) and EM images (Figure 3.7) indicating that although some contaminants remained, when compared to pre-purification images loaded samples the majority of these components were removed. Subsequent SEC (Figures 3.8 and 3.9) was sufficient in purifying the recombinant particles.

When recombinant particles were analysed via CDMS (Figure 3.10) a sharp peak was observed at 1.683 MDa similar to the theoretical mass of wild type STNV of ~1.7 MDa (Appendix C). Nucleic acid, extracted from recombinant STNV VLPs and amplified using primers specific to the synthetic STNV gene,

had been shown to produce a product of approximately 500 bp which is approximately the size of the STNV-1 CP mRNA. This product however, was stated to only account for an estimated 20% of the encapsidated material, the remainder being mRNA precursors, products of degradation or cellular (*E.coli*) RNAs (Lane *et al*., 2011). The codon optimised STNV CP used for expression in *E.coli* has a different molecular weight than that of WT STNV CP, 21,577.52 Da (Gasteiger *et al*., 2005) vs 21,715 Da (WT) (UniProt Consortium, 2018; Chen *et al*., 2017). By calculating the expected mass for empty recombinant STNV VLPs we can see how much remains and is occupied by encapsidated RNA. 60 recombinant STNV coat proteins and internal calcium ions has a calculated weight of 1,298,338.376 Da. To calculate the approximate weight of ssRNA the number of nucleotides is multiplied by 320.5. 159 is then added to this total to account for 5′ phosphorylation (Thermo Fisher Scientific, 2018). When a theoretical 1200 nt are packaged using this formula the calculated mass is 1683097.376 Da or ~1.683 MDa, packaging a theoretical 1239 nt mRNA would yield a mass of ~1.695 MDa. These results indicate that recombinant STNV VLPs package a similar amount of RNA to wildtype particles. If the ~500 bp fragment, corresponding to the STNV-1 CP mRNA, were to comprise only 20% of the contents of the recombinant VLPs the resulting calculated mass would be approximately 2.1 MDa, which is not observed in CDMS (Appendix C).

The purified VLPs were dissociated and reassembled using previously published methods both in the presence and absence of α,2-macroglobulin as a potential protease inhibitor. However, it appears that α,2-macroglobulin

detrimentally impacts the ability of the dissociated VLPs to reform as reassembly is seen in particles incubated in reassembly buffer alone. VLPs were also dissociated in order to purify the STNV coat protein away from the encapsidated RNA using Q-FF and SP-FF columns in series. Investigations with multiple protease inhibitors found that dissociation in the presence of pepstatin A was able to curtail the cleavage that had been occurring suggesting that an aspartic protease was the cause of the problem. When the amino acid sequence of recombinant STNV coat protein is entered into ExPASy peptide cutter 21 enzymes from the database are found to cleave the protein and identified pepsin as the only aspartic protease. It is however worth noting that the ExPASy database does not contain all known proteases in *E.coli* and by no means all proteases in existence (Gasteiger *et al*., 2005). Negative stain TEM images of the purified protein indicate that dissociation of the VLPs was successful, SDS-PAGEs and 260/280 ratios of ~0.5 indicate pure protein with no nucleic acid carry over. Both recombinant VLPs and purified protein were characterised by SEC-MALS (Figure 3.15) and these results will aid the analysis of further reassembly investigations.

This purified protein was reassembled with the aptamer B3 in both a 1:3 and a 1:5 RNA:CP ratio as shown in Figure 3.13. Both ratios resulted in the formation of particles and so as to maximise the potential of the protein future reassemblies were performed in a 1:3 RNA:CP ratio which also reflects the trimeric nature of binding to the small stem loops (Patel *et al*., 2015). Anisotropy investigations with labelled B3 and TR, a stem loop from the MS2 genome show that, while both RNA stem loops are able to initially bind the

STNV coat protein, only STNV is able to reassemble particles of a correct approximate hydrodynamic radius whereas the TR-STNV complex rapidly dissociates. This initial binding may be due to the similarity of the TR and B3 loop sequences, (AUUA and ACAA respectively) however, the drop in anisotropy could be due to aggregation and subsequent loss of solubility. The next chapter investigates the reassembly capabilities of mutated B3 loops to form particles.

[This page has been left intentionally blank]

Chapter 4: Defining the STNV coat protein recognition motif

[This page has been left intentionally blank]

4.1 Introduction

The purified coat protein of STNV is monomeric and in the absence of RNA, remains so until high concentration (15 μM) (Patel *et al*., 2017). When purified CP is titrated into B3 (figure 4.1), under single molecule conditions, the initial stage in the assembly process is the formation of a trimeric capsomer with a hydrodynamic radius of approximately 5 nm. As more CP is added gradually, *T*=1 capsids form. When this procedure is repeated with B34U, a B3 variant containing the loop sequence UUUU, CP binds the RNA and forms this intermediate structure. However, it does not proceed correctly from this point indicating that this loop sequence is important in the STNV assembly mechanism (Ford, 2012).

This Chapter describes investigation of the binding specificity of the B3 stem loop and the determination of the critical features of CP recognition. A series of stem loop sequence variants were produced which retained the stem sequence of PS3 but displayed altered nucleotides in either the 'inner' or 'outer' nucleotide positions in the loop (Figure 4.1). In the natural PS3 loop these two outer positions are occupied by adenines. The 'outer' variants created alter these positions to either a guanine or a uracil, effectively investigating if there is strict sequence specificity or if a purine can be substituted for a purine or a pyrimidine at these positions. A similar procedure was used with the inner variants however, as the recognition motif was thought to be restricted to the two 'outer' nucleotides in the loop, as this was the motif shared by all 30 identified sites in the STNV genome, only 5 'inner' variants were created from the possible 16.

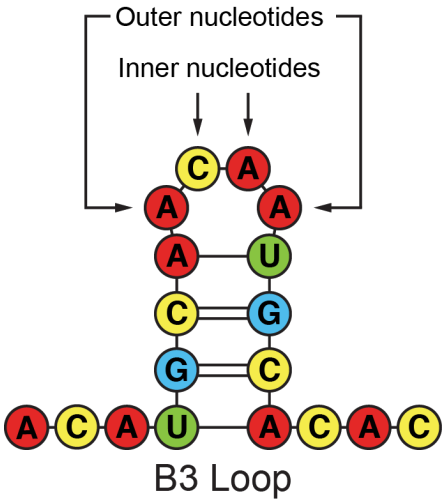


Figure 4.1: B3 Stem loop. Cartoon of the B3 loop indicating the inner and outer nucelotide positions which were altered for the PS motif investigations.

4.2 Results

4.2.1 Production of B3 variants

Primers to produce B3 short variant DNA templates were designed encompassing a T7 promoter and the desired loop sequence (2.2.5.2). The forward primer was designed to be universal for all variants. Using the mechanism shown in Figure 4.2, PCR and subsequent transcription produced the B3 short variants (Figure 4.3). PCR was carried out using the KAPA 2G system and resulting products were analysed by native PAGE (Figure 4.4). The KAPA 2G system utilises a second-generation DNA polymerase which has a faster extension rate that that of wild type *taq* polymerase (Biosystems K). Bands are seen at ~ 45 bp which is the approximate size of the PCR products (43 bp). The products obtained from PCR were then transcribed using a HighScribe T7 kit and analysed via urea denaturing PAGE (National Diagnostics) to determine whether transcription had been successful. When

bands of the correct size were observed, larger volumes of the samples were run on the urea denaturing PAGE and subsequently purified. Bands were extracted from the gel using the 'crush and soak' method. RNA transcript concentrations were determined using a nanodrop spectrophotometer and extinction coefficients shown in Table 2.5. The resulting RNAs were analysed by denaturing PAGE (Figure 4.5) to ensure they were the correct size and used in downstream investigations as described in the following sections.

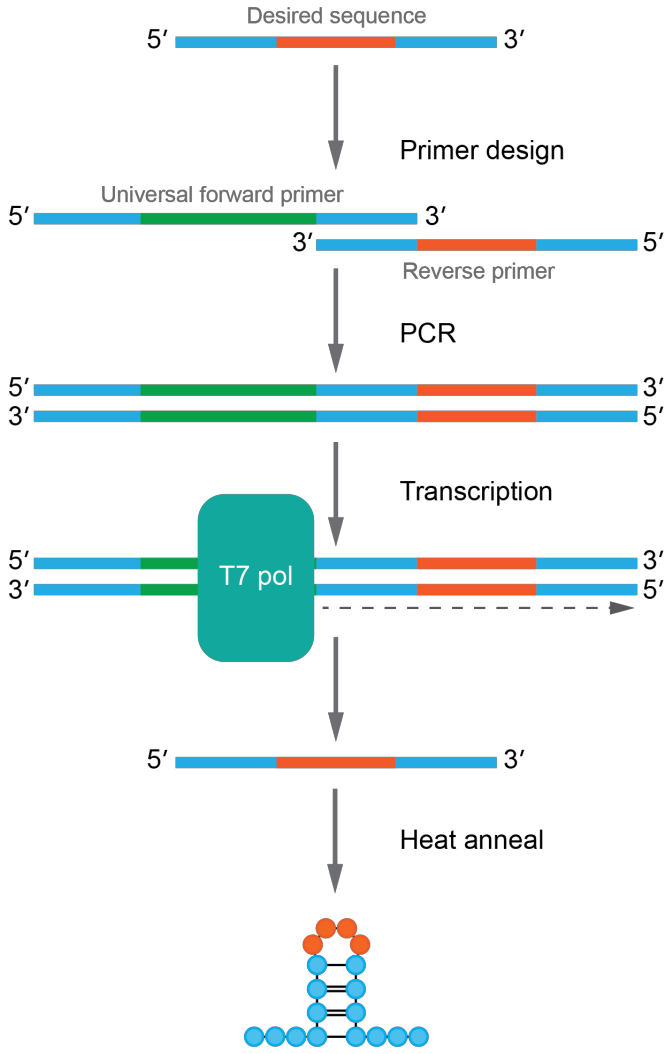


Figure 4.2: Schematic of B3 variant synthesis. Schematic showing the stages in B3 variant synthesis.

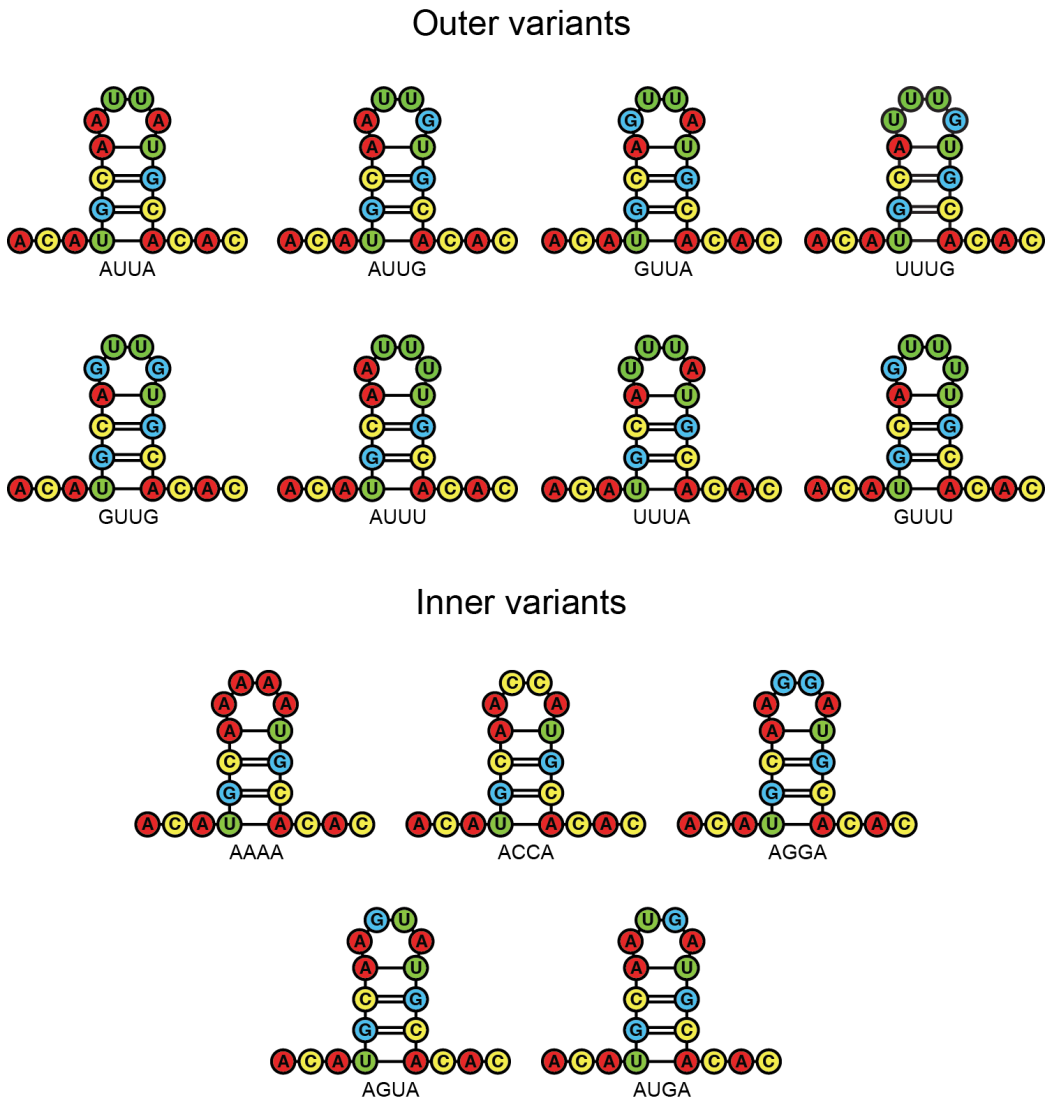


Figure 4.3: B3 variant cartoons. Mfold structures of all B3 variants with the loop sequence given underneath. For Mfold folding energies of all variants see Appendix B.

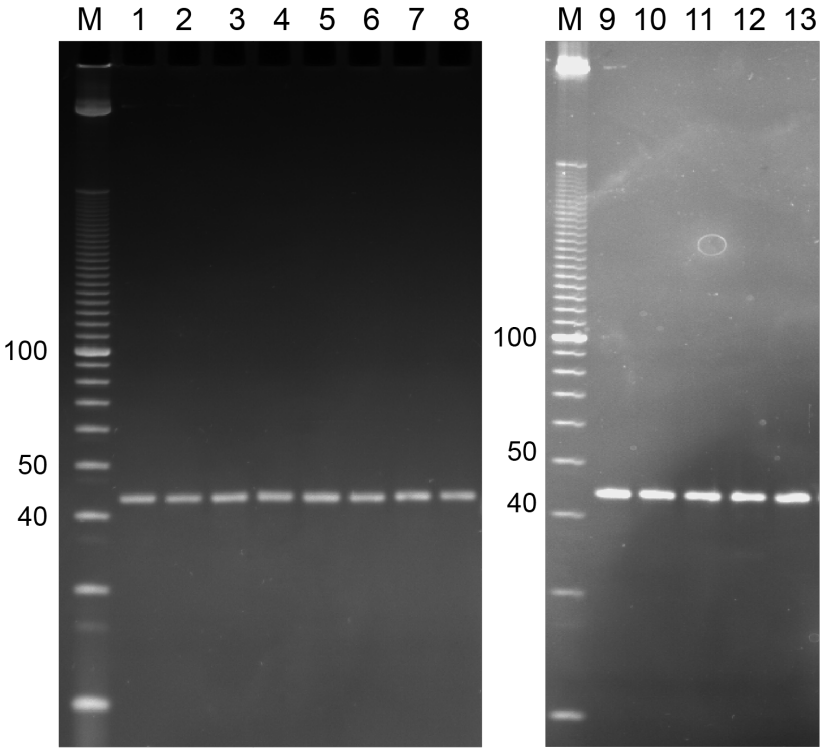


Figure 4.4: Native PAGEs of variant DNA templates. PCR products for all variants are seen at approximately 43 bp.

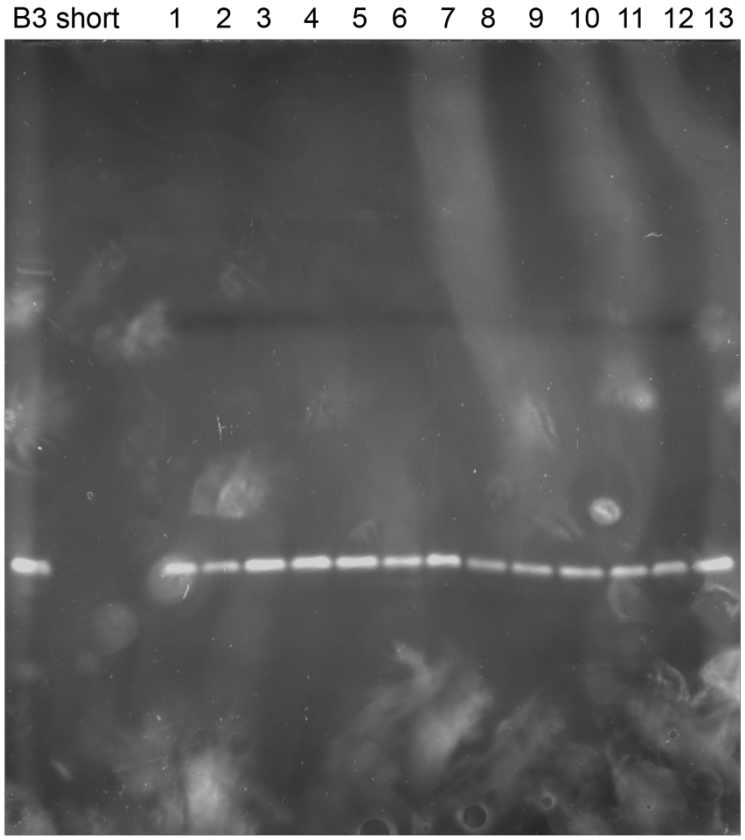


Figure 4.5: Denaturing PAGE of B3 short variants. Transcription products of B3 short variants (1-13) with B3 short, purchased from IDT, used as a size control.

4.2.2 Assembly with B3 variants

B3 variants were produced and purified as described in section 4.2.1. These variants were designed to test the optimal features of the STNV CP:RNA interaction and subsequently the effects these sequences have on assembly. It has been shown that the STNV-1 genome contains 30 stem loop structures that present this AXXA motif in 4-, 5- or 6 nt stem loops.

4.2.2.1 TEM analysis

All the B3 variants produced VLPs when incubated with dissociated coat protein in a 1:3 RNA: CP ratio and subjected to dialysis in STNV reassembly buffer for 18 hours at 4˚C. This behaviour was not unexpected, as it has previously been shown that B34U, which has a mutated loop sequence of 4 uridine residues, will make VLPs under these conditions (Ford, 2012). It was the extent of capsid assembly among the variants that was interesting. Negative stain TEM images of reassembly reactions are shown in Figure 4.6, recombinant VLPs from *E.coli* are shown in 3.7D for comparison. The AUUA variant, which contains the 'ideal' motif is depicted in Figure 4.6A. The VLPs produced in this reassembly reaction are of comparable morphology to the recombinant sample from *E.coli*, and to examples from the literature (Lane *et al*., 2011; Ford *et al*, 2013; Ford, 2012; Patel *et al*., 2015). The reassembled particles are spherical in shape and there are few aggregates or misshapen capsids. The inset image shows that VLPs assembled around the AUUA variant RNA are approximately 20 nm in diameter further suggesting they are correctly assembled STNV VLPs. The AUUU and UUUA (Figure 4.6B and C) variants also seem to contain fully formed particles which conform to the

known morphology of STNV VLPs suggesting that both of these RNAs are adequate templates for assembly at this concentration. The AUUG and GUUA variants seem to contain mixed species of both fully formed and partially formed capsids (Figure 4.6D and E). Many particles appear to be only partly assembled and these are interspersed with particles that appear to be the correct size and shape. However, it is obvious from the negative stain TEM images that the GUUG, GUUU and UUUG variants are less capable of forming VLPs. As seen in Figure 4.6 (F, G and H), intermediates of *T*=1 capsids are prevalent with very few, if any complete particles. Inset images show particles that are misshapen and rarely display complete *T*=1 architecture. It is evident from the TEM images that the further the variants deviate from this 'ideal' loop motif the more partially formed capsids are seen in samples analysed by negative stain TEM. These results confirm previous expectations and suggest that the outer adenines are components of the recognition motif.

The inner loop variants (Figure 4.7) however, all display similar architecture to that of *E.coli* derived STNV suggesting that the inner nucleotides in the loop do not affect capsid formation or general structure. Particles are overall spherical in shape and do not appreciably diverge from this template unlike the morphological range of particles seen in the outer loop variant reassemblies. As the 30 STNV genomic stem loops identified by SELEX all shared the same AXXA motif, it was expected that there was no base specificity at these inner nucleotides. The ability of these differing stem loops to form particles with comparable structure indicates that these central nucleotides are not components of the CP recognition motif.

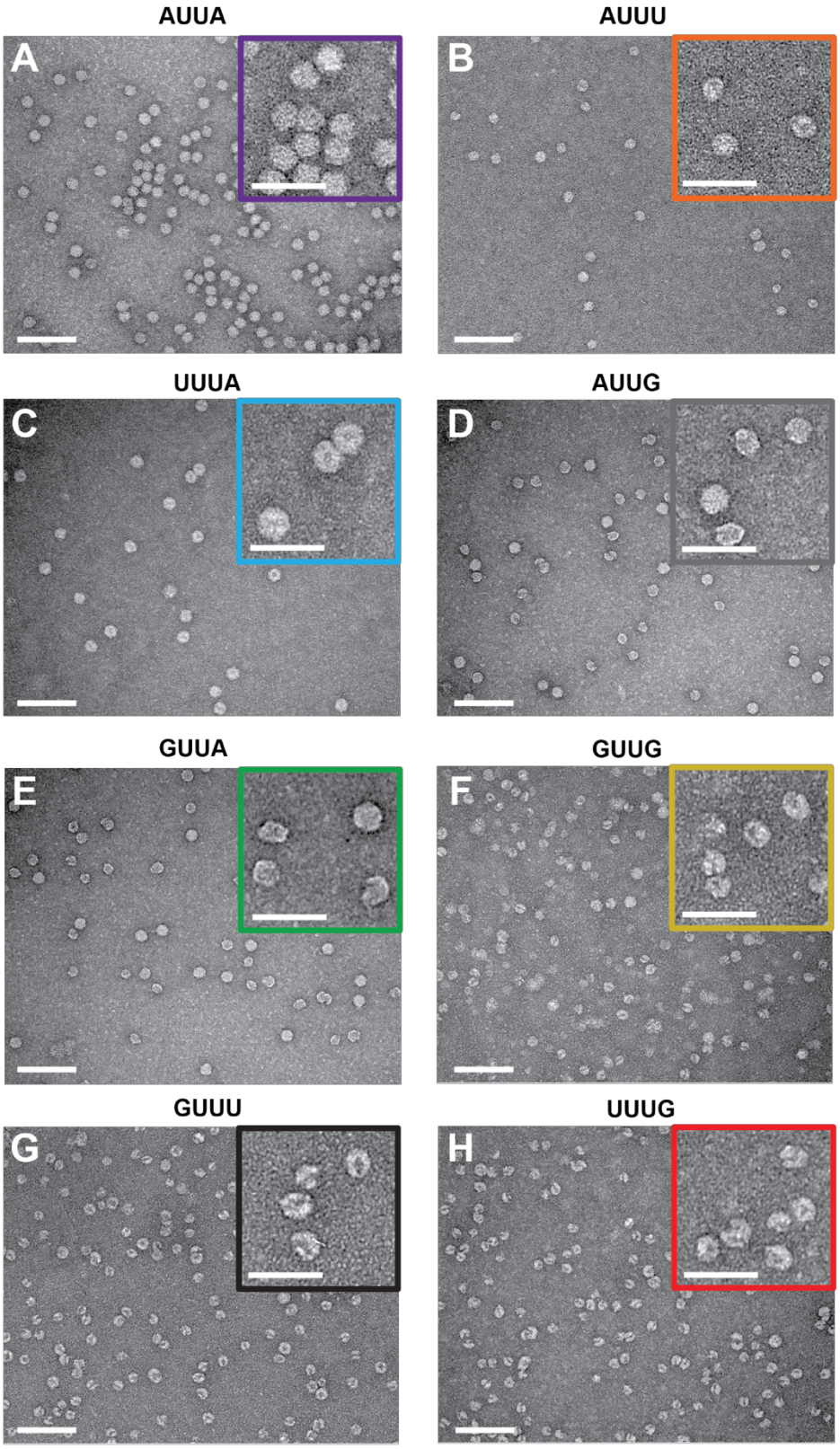


Figure 4.6: Outer variant reassembly negative stain TEM images. EM images of outer variant reassemblies. Reassemblies were carried out at a 1:3 RNA:CP with a final CP concentration of 6 μM. Images were obtained on a JEOL 1400 at 40,000 X magnification. Scale bar = 100 nm. Inset images are colour coded to variant and maintained throughout.

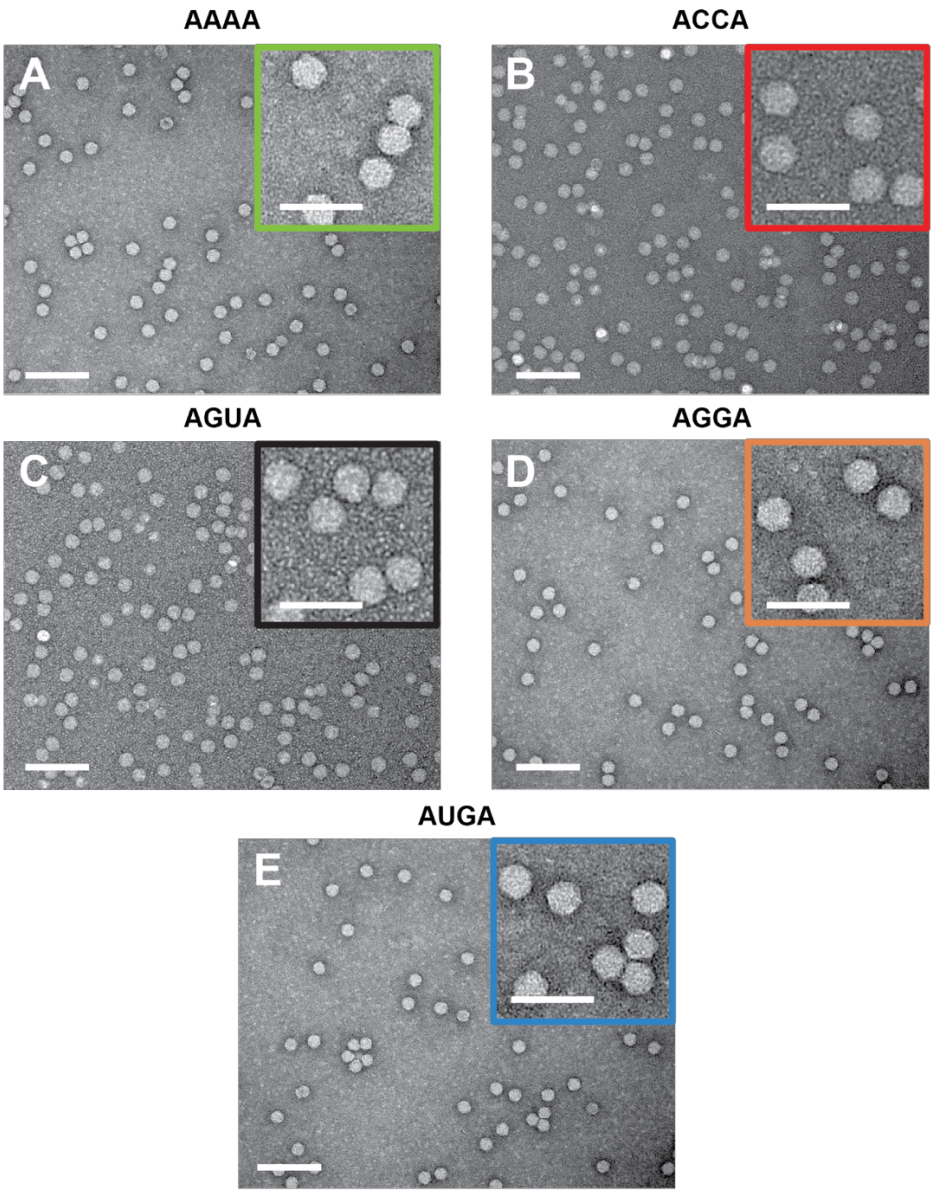


Figure 4.7: Inner variant reassembly negative stain TEM images. TEM images of inner variant reassemblies. Reassemblies were carried out at a 1:3 RNA:CP with a final CP concentration of 6 μM. Images were obtained on a JEOL 1400 at 40,000 X magnification. Scale bar = 100 nm. Inset images are colour coded to variant and maintained throughout.

4.2.2.2 Sedimentation velocity analytical ultracentrifugation (svAUC)

Electron microscopy was not thought to be enough to determine the assembly capacity of the B3 variants and as such the reassembly reactions underwent analysis via svAUC. To determine the sedimentation velocity, a homogenous solution is placed into a 2-sector cell and subjected to high angular velocity. The solute rapidly sediments to the bottom of the cell and causes a depletion about the meniscus. A boundary forms between this depleted area and the homogenous solution. The rate of movement of this boundary is measured by absorbance from which the sedimentation coefficient (s) can be determined. Sedimentation coefficient is measured in Svedbergs (1 Svedburg = 10^{-13} seconds) (Ralston, 1993).

STNV reassembly reactions around the B3 aptamer have been shown to produce a species that sediments at 42 S (*T*=1) when reassembled in a ratio of 1:1 (RNA:CP). B3 reassembly also produces another species that sediments at approximately 28 S. This species is thought to be incomplete (half) *T*=1 capsids (Ford, 2012) since by increasing the concentration of CP with respect to RNA concentration the level of this intermediate species is reduced. Knowing the sedimentation values of B3, B34U and recombinant STNV aided analysis of the reassembly reactions with the B3 variants. B34U however, is incapable of producing VLPs compared to B3, and even in a 1:5 RNA:CP ratio produces more intermediates sedimenting at 28 S, than fully formed VLPs (Ford, 2012). All variants (Figure 4.2) were reassembled with purified CP in a 1:3 RNA: CP ratio, final CP concentration of 6 μM, and subjected to dialysis in STNV reassembly buffer for 18 hours at 4˚C. These

samples were then analysed by svAUC and the data is presented in Figures 4.8-4.11.

The ability of the outer variant reassemblies to form particles is illustrated in Figure 4.6. These results further suggest that there is variation in the ability to assemble particles among the RNA templates and this variation is based on sequence differences. As previously mentioned, reassembly around the B3 aptamer results in particles that sediment at approximately 42 S. This is reflected in the sedimentation coefficient of the sample containing the AXXA motif (AUUA), and both the AUUU and UUUA variants. These results are consistent with the negative stain TEM images which show particles of the correct morphology and size for these variant reassemblies. Again, the GUUG, GUUU and UUUG variants are less capable of forming VLPs. Major peaks for these reassemblies are typically seen at ~30 S which indicates incomplete assembly products. In the case of GUUU this is accompanied by a minor peak closer to the correct sedimentation coefficient for *T*=1 particles however, it is far below the value seen for the correctly assembling species. Both GUUA and AUUG variants display this double peak trend. This double peak behaviour is seen for the B34U variant which suggests that formation of these correctly sedimenting species seen in these minor peaks is independent of the loop sequence and may be due to the concentration at which the reassemblies are performed. When the results are plotted together on one graph (Figure 4.9) these differences become more evident. Assembly capacity splits the outer variants into two distinct families, particles that sediment like wild type (AUUA, AUUU and UUUA), and those that do not sediment like wild type.

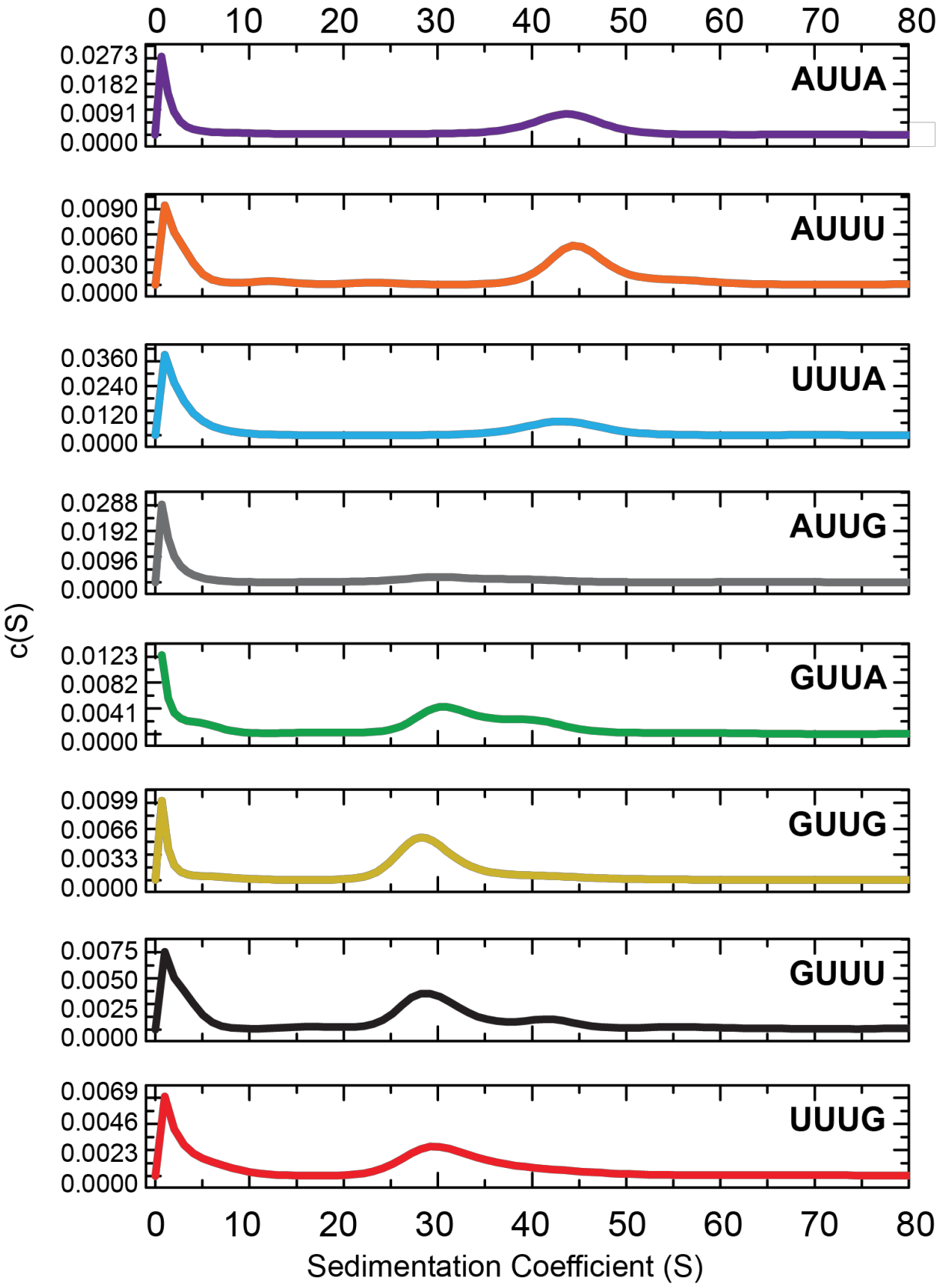


Figure 4.8: Outer variant svAUC results. B3 outer variant reassemblies analysed by svAUC. Variant loop sequence is given in the top right hand corner of each and colour coding is maintained throughout. Reassemblies were carried out at a 1:3 RNA:CP with a final CP concentration of 6 μM. Sedimentation of *T*=1 VLPs is typially seen at ~42S.

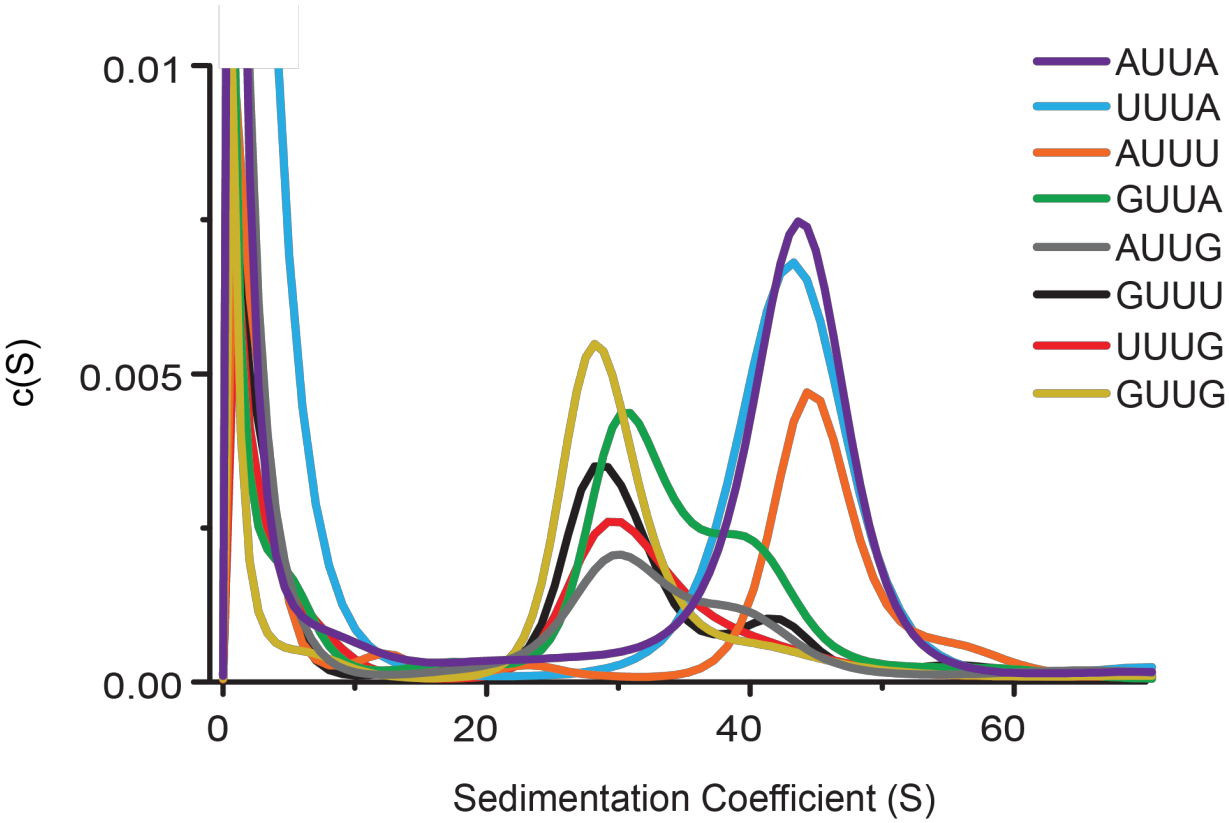


Figure 4.9: Composite outer variant svAUC. Ensemble reassembly of outer variant B3 RNAs, analysed by sedimentation velocity (variant RNAs are color-coded as given in Inset). Reassemblies were carried out at a 1:3 RNA:CP with a final CP concentration of 6 μM. The expected *T*=1 VLP sediments at ~42 S, where S is the sedimentation (Sed) coefficient. c(S) is a continuous distribution from the Lamm equation model (Schuck *et al.*, 2015).

The inner variants however, do not show this diverse behaviour (Figure 4.10). All chosen variants have one peak that sediments at approximately 42 S. This aligns with the TEM images where the particles appear to be fully formed and of the correct size and shape. When these individual results are combined into one graph (Figure 4.11) all the peaks align over the previously published sedimentation coefficient for STNV assembled with B3 and B3 short (Ford, 2012). This indicates that the particles produced in these reassemblies are the approximate size and shape of the wild type and B3 particles.

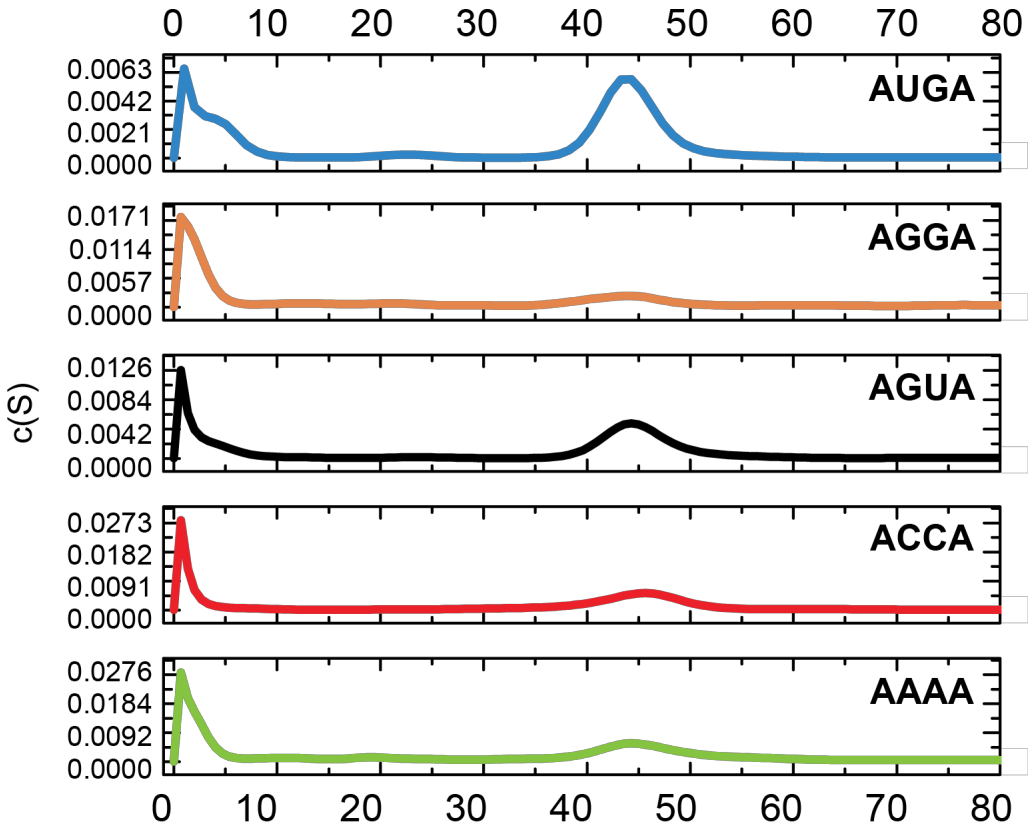


Figure 4.10: Inner variant svAUC results. Inner variant reassemblies were carried out at a 1:3 RNA:CP with a final CP concentration of 6 μM and analysed by svAUC. Variant loop sequence is given in the top right hand corner of each pannel.

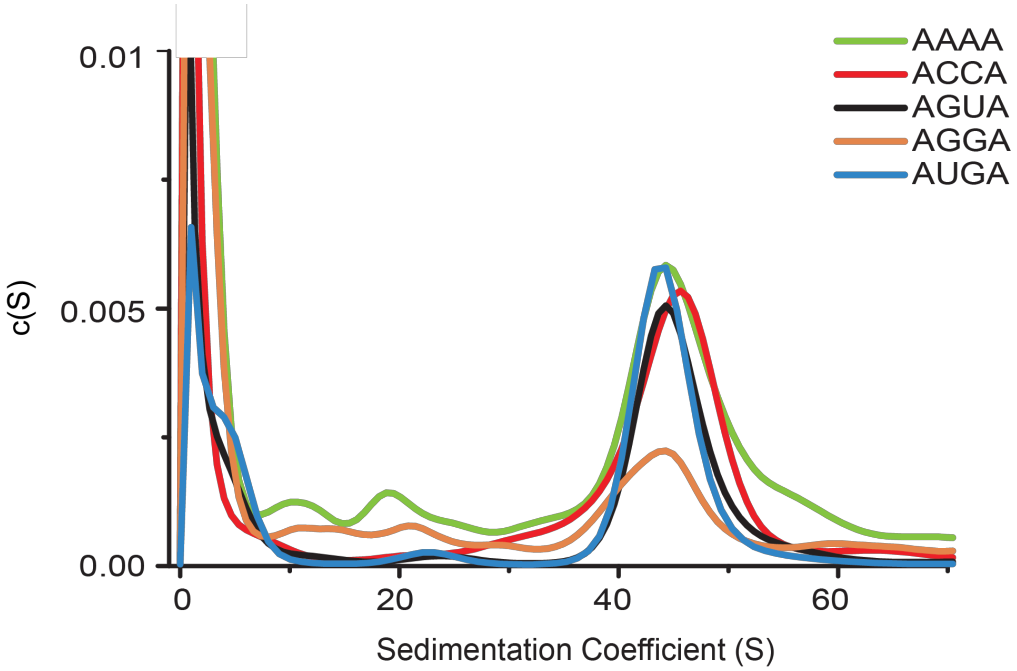


Figure 4.11: Composite inner variant svAUC. Ensemble reassembly of inner variant B3 RNAs, analysed by sedimentation velocity (variant RNAs are color-coded as given in Inset). As in 4.7 the expected *T*=1 VLP sediments at ~42 S, where S is the sedimentation (Sed) coefficient. c(S) is a continuous distribution from the Lamm equation model (Schuck *et al*., 2015).

By integrating the svAUC traces it is possible to identify the proportion of the samples that resembled into each species or failed to assemble. Using Origin (2017 version), the data was segregated into three sections; 0-10 S, 20-35 S and 35-55 S using the integrate function, data is presented in Table 4.1. These S designations are categorised as unassembled/small products (0-10 S), incomplete particles (20-35 S) and fully assembled particles (35-55 S).

Table 4.1: Integration of svAUC reassembly data.

	GUUU	UUUG	UUUA	AUUU	GUUA	GUUG	AUUG
% 0-10 S	34.54	38.83	36.61	36.58	26.55	23.14	43.35
% 20-35 S	43.97	41.18	1.78	2.49	42.54	63.09	32.34
% 35-55 S	16.79	17.73	60.33	54.59	29.54	12.74	23.24

	AUUA	ACCA	AGUA	AGGA	AUGA	AAAA
% 0-10 S	37.29	40.84	37.51	26.01	34.29	38.91
% 20-35 S	4.76	4.745	1.44	7.28	2.04	7.06
% 35-55 S	55.10	52.75	57.98	57.14	62.12	43.45

4.2.2.3 Light scattering

The analysis of these reassemblies up until this point has not been quantitative so light scattering (LS) was used to benchmark some of the assembled products and determine the efficiency with which a selection of the B3 variants formed *T*=1 capsids. Selected variants were reassembled with purified CP in a 1:3 RNA: CP ratio, final CP concentration of 6 μM, and subjected to dialysis in STNV reassembly buffer for 18 hours at 4˚C. The aptamer B3 was used as the positive control and B43U as the negative control. Samples were separated on a Superdex 200 10/300 GL size exclusion column (GE Healthcare). As seen in Figure 4.12, variants have different size exclusion profiles. B3, the positive control (4.12A) contains 2 peaks, one from 7-9.6 mL and a second from 12.5-15 mL. Previous investigations on the light scattering of both *E.coli* derived VLPs and purified STNV Cp (Chapter 3, Figure 3.15) allow the two peaks in these traces to be assigned to VLP (7-9.6 mL) and free Cp (12.5-15 mL). B3 and AUUA both display similar size exclusion profiles suggesting that their ability to form *T*=1 capsids is similar. However, both the GUUG and AUUG variants perform in a comparable fashion to B34U, which suggests non-specific reassembly events as this loop contains only uridines. Areas under the curves was calculated for the assembled VLP peaks seen between 7.0 and 9.6 mL, B3 = 0.139, AUUA = 0.137, B34U = 0.085, GUUG = 0.059 and AUUG = 0.088. LS-QELS was carried out with recombinant VLP and purified CP in Chapter 3 (3.2.10) where peaks eluted are consistent with those seen for these reassembly reactions.

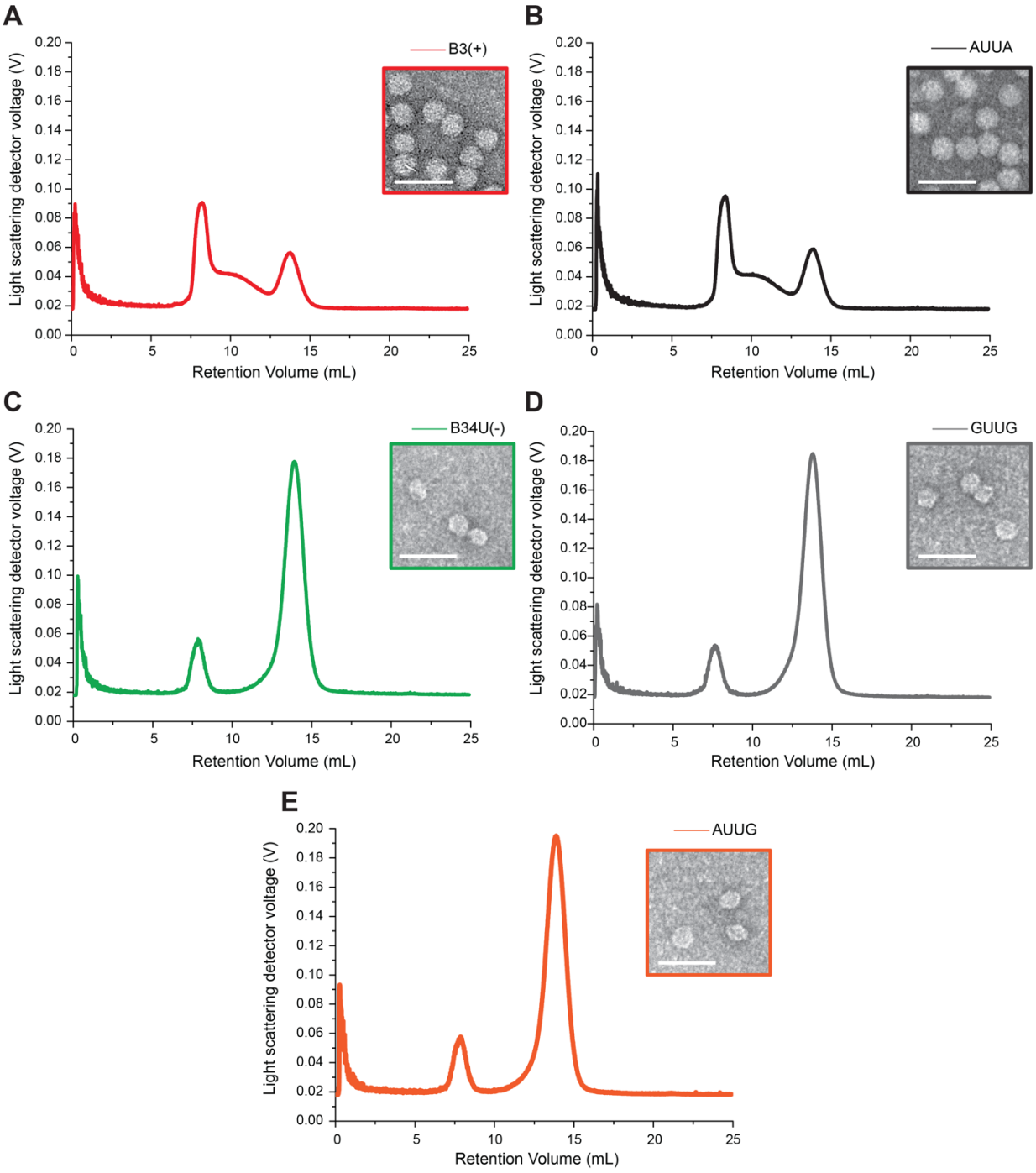


Figure 4.12: B3 Variant light scattering. Ensemble reassembly of 'good' and 'bad' B3 variant RNAs, analysed by LS (6μM CP), B3 (red), AUUA (black), B34U (green), GUUG (grey) and AUUG (orange). Areas under the curves calculated for the assembled VLP peaks (7.0-9.6 mL) B3 = 0.139, AUUA = 0.137, B34U = 0.085, GUUG = 0.059 and AUUG = 0.088.

4.2.3 Single molecule fluorescence correlation spectroscopy (smFCS)

An adapted smFCS assay was used to investigate the relative CP affinity of the loop variants as demonstrated in Figure 4.13. In this assay, purified CP was titrated into labelled B3 to form the trimeric capsomere, formation of this intermediate was judged by the R_h value. This species was monitored to ensure its stability at which point 100-fold molar excess of an unlabelled B3 variant was added. A variant that binds with similar affinity to the B3 aptamer is able to displace the labelled RNA which restores the R_h to that of CP-free RNA. Those variants that do not bind with similar affinity do not displace the bound RNA. These results are shown in groups of three, highlighting the differing behaviour of species with similar sequences (Figure 4.14 to 4.18).

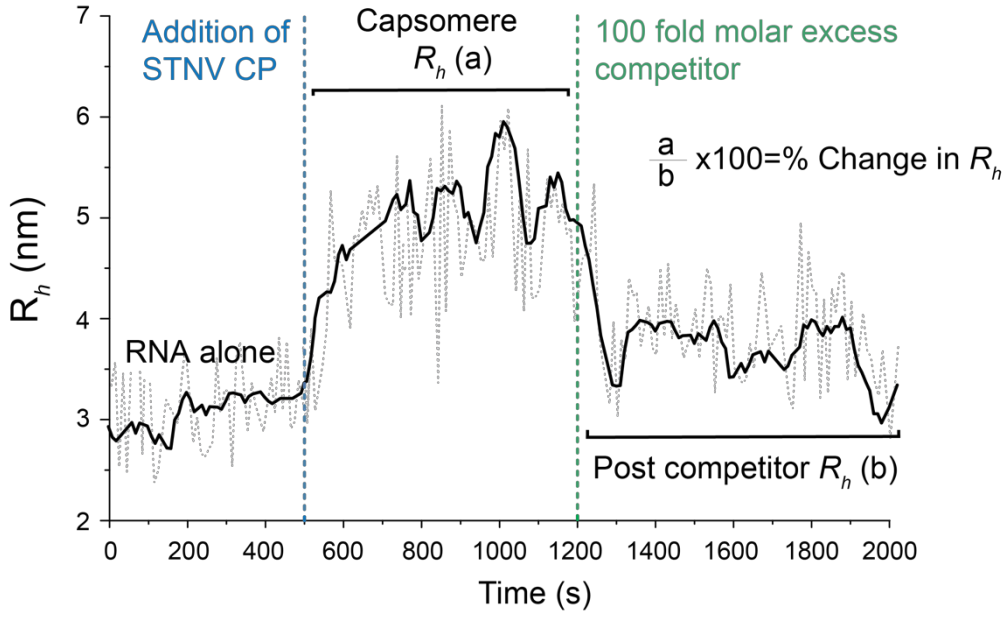


Figure 4.13: Cartoon of adapted single molecule fluorescence correlation spectroscopy (smFCS) assay. Purified STNV CP was titrated into labelled B3 and R_h monitored for the formation of capsomere (5 nm). This species was monitored to ensure stability. 100-fold molar excess of a given variant was added at the time indicated by the green dotted line. The R_h was monitored post addition of variant. Both capsomere and post variant R_h were used to calculate the percentage change in R_h as indicated in the figure.

Figure 4.14 shows single molecule displacement assay results for the AUUA, AUUU and UUUA variants. Two of these stem loops, AUUA and UUUA are able to displace the dye label which is seen in the reduction of the hydrodynamic radius to approximately 3.2 nm. Surprisingly in this set of results not all the adenine containing loops are able to displace the dye labelled B3 (ACAA). Previously these variants produced morphologically similar particles in both TEM and svAUC however, it appears that at the lower concentrations (nm as opposed to μM) used in single molecule these loops do not perform with similar ability. The AUUU variant is unable to displace the B3 (ACAA) which suggests that this variant does not have the requisite affinity for the CP to overcome the bound B3.

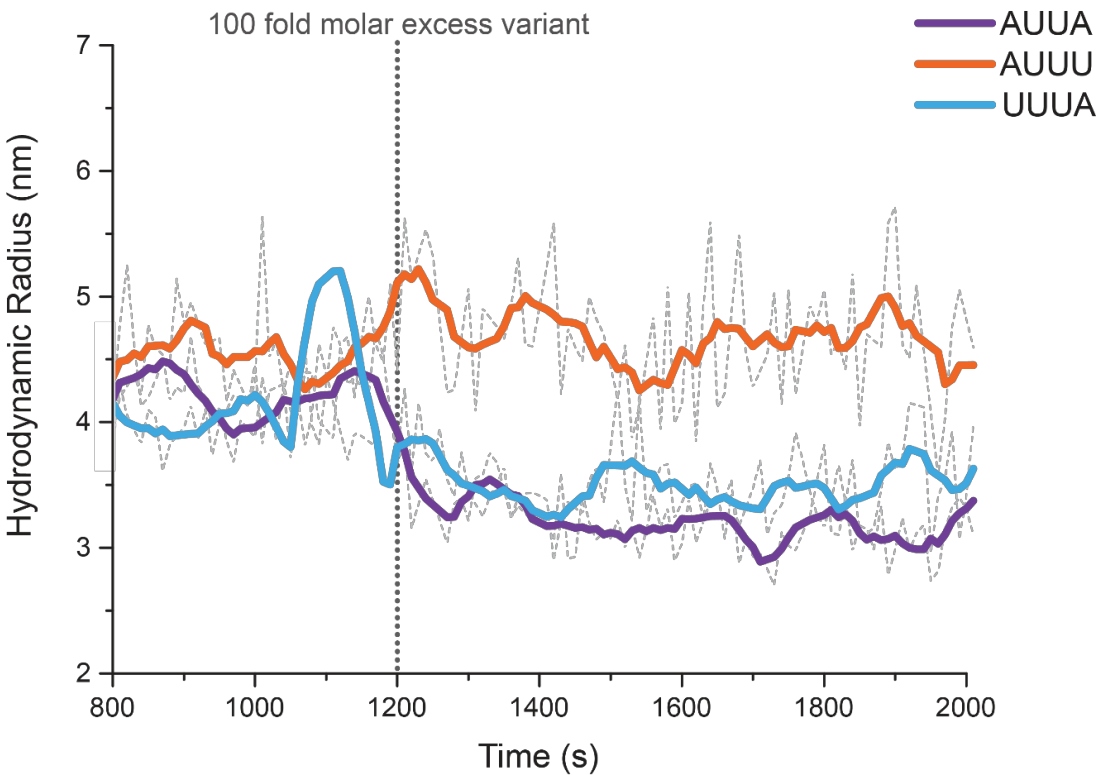


Figure 4.14: Outer variant capsomere displacement assays 1. Purified STNV CP was titrated into 5 nM Alexa Fluor 488 (AF488) labelled B3 to form a capsomere of ~5 nm R_h. This was equilibrated for 1200 s, at which point 100-fold molar excess of a B3 variant AUUA (purple), AUUU (orange) or UUUA (blue) was added and the resulting R_h change tracked for 800 s.

G containing variants GUUG, GUUU and UUUG are also unable to displace the dye label as shown in Figure 4.15. All three species retain their hydrodynamic radius at the expected size for a capsomere indicating that none of the variants in this group possess affinity for CP capable of interrupting the intermediate capsomere structure. This mirrors the TEM and svAUC results where these variants were unable to produce morphologically correct particles, or particles with a similar sedimentation coefficient to that of wild type or *E.coli* derived STNV. The behaviour of these variants in the single molecule displacement assay further suggests that guanosine substitutions at these outer locations in the stem loop adversely affect particle formation and affinity of a step loop for cognate CP

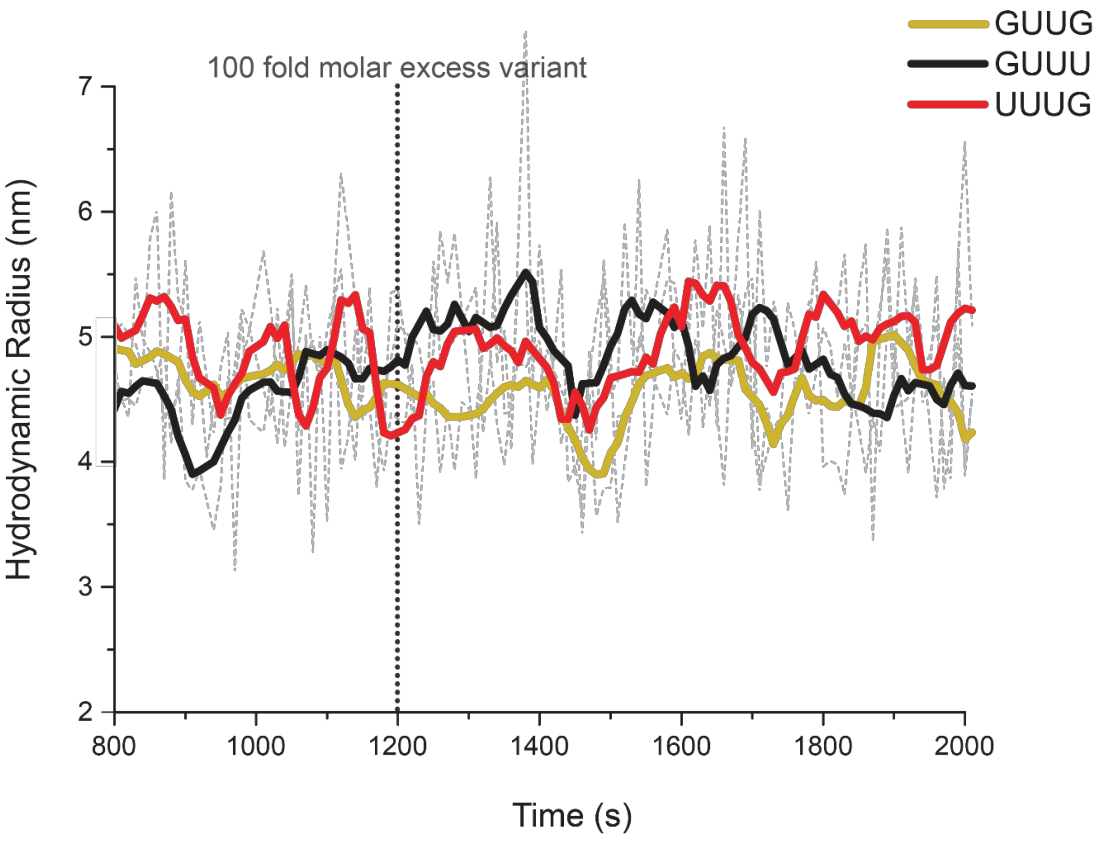


Figure 4.15: Outer variant capsomere displacement assays 2. Purified STNV CP was titrated into 5 nM AF488 labelled B3 to form a capsomere of ~5 nm R_h. This was equilibrated for 1200 s, at which point 100-fold molar excess of a B3 variant GUUG (gold), GUUU (black) or UUUG (red) was added and the resulting R_h change tracked for 800 s.

Interestingly, another variant stem loop is able to compete off the dye label. While the AUUG variant retains the hydrodynamic radius of the trimeric capsomere formed with STNV CP and labelled B3, the GUUA stem loop is able to displace the dye (Figure 4.16). When reassembled, this variant produced a range of particle shapes with both complete and partially formed capsids visible via TEM. However, at the lower concentrations of single molecule it seems that this stem loop has the required CP affinity to disrupt the interaction. The B3 positive control is also shown in Figure 4.16 and behaves as expected 100-fold molar excess of B3 is able to compete off the dye label.

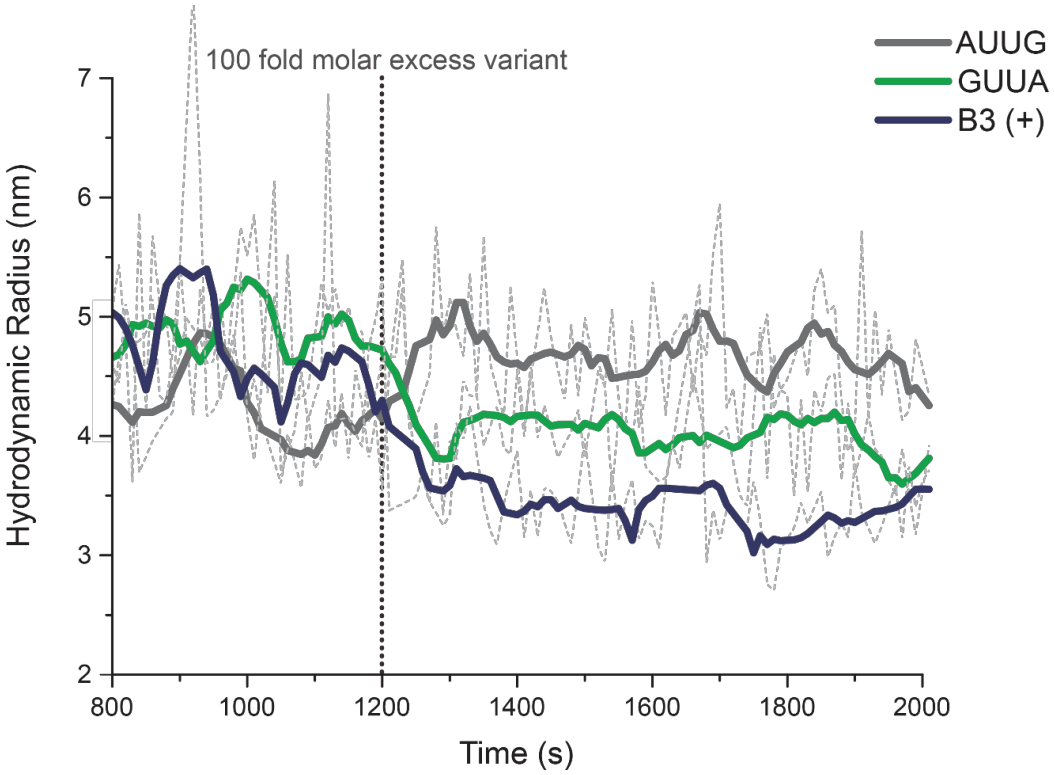


Figure 4.16: Outer variant capsomere displacement assays 3. Purified STNV CP was titrated into 5 nM AF488 labelled B3 to form a capsomere of ~5 nm R_h. This was equilibrated for 1200 s, at which point 100-fold molar excess of a B3 variant AUUG (grey), GUUA (green) or B3 (navy) was added and the resulting R_h change tracked for 800 s.

These results have also been plotted to show the percentage R_h change across all 'outer' variants. Using the data generated from the displacement assay, an average species size could be calculated for both trimeric capsomere and post-introduction of variant. This was then used to calculate the percentage change in hydrodynamic radius (Figure 4.17). When plotted in this way, these results show that the greatest change in hydrodynamic radius is seen with B3, the positive control. Unsurprisingly, this is closely followed by the AUUA stem loop which out of all the variants tested follows the AXXA motif exactly. The GUUA and UUUA variants also display quite dramatic changes in hydrodynamic radius although not as drastic as seen for the AUUA variant. All other outer variants show <10% change in hydrodynamic radius in either a positive or negative direction.

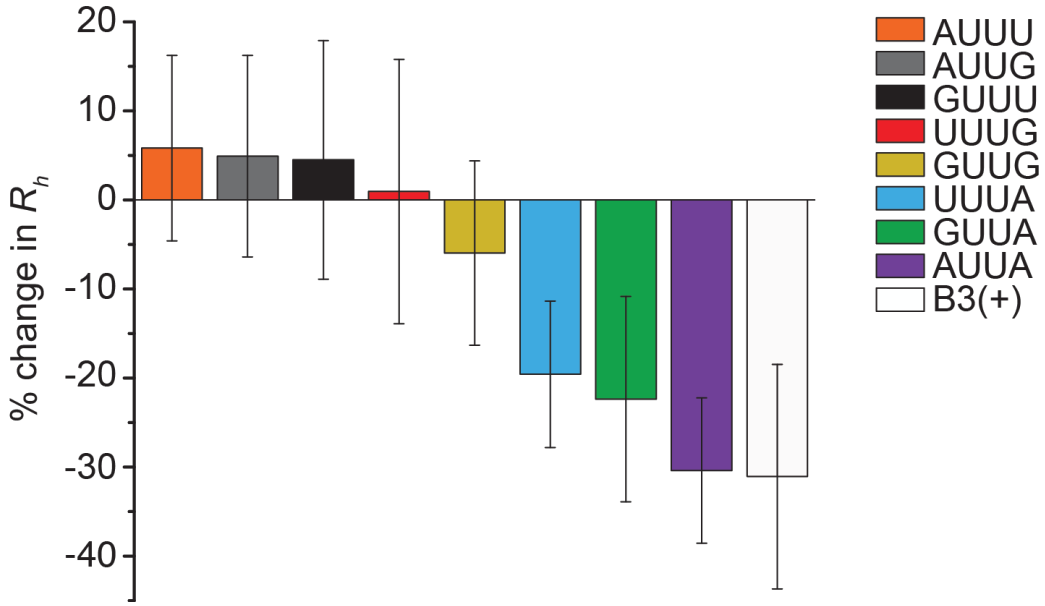


Figure 4.17: B3 variant capsomere displacement percent change in R_h. Percent change in R_h following addition of 100-fold molar excess of variant RNAs (colour-coded as inset) to a capsomere (R_h ~ 5 nm) formed with 1-nM AF488-labeled B3. Error bars indicate SEM.

The dye displacement assay was also repeated for the 5 inner variants previously used in the TEM and svAUC investigations (Figure 4.18). All the inner variants are able to compete off the dye labelled B3 and behave similarly to both the AUUA variant and B3 the positive control further suggesting that the inner nucleotides in the loop have no base specificity and do not form part of the CP recognition motif.

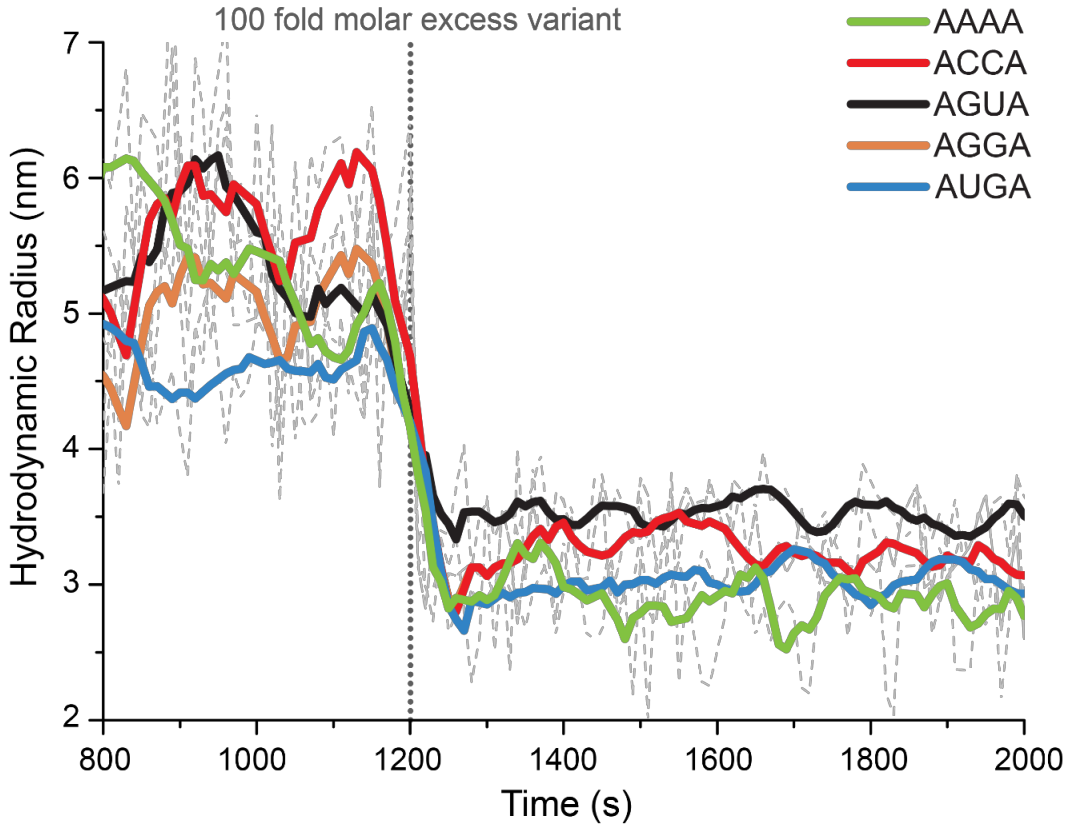


Figure 4.18: Inner variant capsomere displacement assays. Purified STNV CP was titrated into 5 nM AF488 labelled B3 to form a capsomere of ~5 nm R_h. This was equilibrated for 1200 s, at which point 100-fold molar excess of a B3 inner variant AAAA (green), ACCA (red), AGUA (black), AGGA (orange) or AUGA (blue) was added and the resulting R_h change tracked for 800 s.

4.3 Discussion

Analysis of all three STNV strains suggests that there are <30 AXXA presenting loops in each genome. Previous work has shown that purified CP, when titrated into labelled B3 results in the formation of a trimeric capsomere (R_h = 5nm) (Bunka *et al*., 2011). As the titrations continue, the R_h increases and plateaus at ~ 11.3 nm (R_h of a VLP). When this is repeated with a loop that presents a UUUU motif (B34U), the stem loop is capable of binding, but assembly does not continue to VLP. When the products of both these reassemblies are analysed by svAUC, different behaviour is observed. The B3 reassembly produces a species which sediments at ~ 42 S, with a minor peak at ~30 S. However, the B34U sample contains a significant amount of this secondary species and only small amounts of correctly sedimenting particles (Ford, 2012; Ford *et al.*, 2013). STNV-1 CP does not aggregate below 15 μM therefore the behaviours observed is due to the RNA:CP interactions (Patel *et al*., 2017). As the difference between B3 and B34U is so subtle, it stands that the AXXA motif is a key component in this interaction.

The stem loop variants created in this chapter aimed to identify the critical recognition features of PS3 by systematically altering the loop presented by the B3 stem. These stem loops were reassembled with purified STNV-1 CP and analysed by TEM, svAUC and LS. In each case it was evident that the stem loops behaved differently under identical reassembly conditions. As seen in the TEM images, there is wide variation in the ability of the stem loops to form complete particles. All inner variants display similar capsid architecture to WT STNV particles, additionally inner variant reassemblies sediment at a

similar position to PS3/B3 indicating that the particles formed are comparable. These results confirm previous assumptions that the inner two nucleotides do not form part of the STNV recognition motif. After svAUC analysis AUUA, AUUU and UUUA variants all have peaks at a similar position to PS3/B3 indicating that under these conditions they can form particles that sediment like wild type. All other variants fail to show this behaviour, with the majority of G containing variants displaying a peak at ~30 S, or in some cases a double peak profile with a major peak at approximately 30 S and a second, smaller peak approaching 40 S. This two-peak profile is seen for the B34U variant and suggests non-specific interactions which may be an artefact of high concentrations used for reassembly conditions (Ford *et al*, 2013). Integration of the svAUC data indicates that on average ~35% of the reassembled products have a sedimentation coefficient of 0-10 S (SEM =1.69) indicating that this fraction did not assemble into significant higher order structures. All variants not encompassing a G in an outer position had a higher percentage of sample within the 35-55 S designation, suggesting correctly formed particles. Those containing a G had a higher percentage of particles sedimenting at 20-35 S.

One ‘good’ variant (AUUA) and two ‘bad’ variants were also benchmarked using light scattering (LS) and compared to B3 and B34U as positive and negative controls respectively. In this experiment the AUUA variant behaved like B3, the positive control, with a large VLP peak at ~7-9.6 mL and a smaller CP peak at 12.5-15 mL. This similarity in behaviour suggests that both B3 and the AUUA variant contain the necessary motif to form *T*=1 particles. These

stem loops both share the AXXA motif suggesting that these outer adenines are components in the recognition motif.

An adapted smFCS assay was used to examine the relative importance of the loop residues for CP affinity. The AUUA stem loop was able to displace the dye label suggesting that it has a similar affinity for STNV CP as seen with B3/PS3. As this stem loop most closely follows the assumed preferred motif, this is not surprising and confirms that the AXXA motif is preferred for CP affinity. All variants with only guanine substitutions failed to displace the dye labelled B3 indicating no affinity for the STNV CP. The 'A' variants however, showed differences in their ability to displace the dye labelled B3. The AUUU variant fails, suggesting that this loop does not have the required affinity. Conversely, the UUUA variant does displace the dye labelled B3 which suggests that the 3′ adenine is perhaps more important for CP recognition. This is reflected in the behaviour of the GUUA variant which is also able to displace the dye further highlighting the potential significance of the 3′ adenine. An alternative interpretation of these results is that these two stem loops alter their structure in the context of CP binding to the 3′ adenine and present an AUUUAU/AGUUAU motif that is still recognised by the CP (Patel *et al*, 2017). Both the AUUU and UUUA variants formed correctly sedimenting particles when analysed by AUC however, under single molecule conditions their ability to displace the dye labelled B3 varies. This suggests that high concentration can drive the formation of correctly sedimenting particles with both species but at single molecule levels the mechanism is more discerning.

In addition to deconvolution of the STNV recognition motif recent work has also shown how other factors affect particle formation. Although the presented work demonstrates that it is the sequence specificity of the RNA:CP interactions that determine assembly ability, the historical suggestion is that electrostatics are the major component in this mechanism. Many plant viruses, which are typically used as model systems, have positively charged N-terminal arms implying that assembly occurs via charge neutralisation. The coat protein of STNV-1 contains 8 amino acids that carry positive charges at pH7. Mutation of side chains closest to the RNA seen in the B3 VLP structure (R8, R14 and K17) all produced coat proteins that expressed normally however their ability to bind single stem loops (B3) and the 5-PS fragment differs (Figure 4.19B). Only the R8A mutant was able to bind 1 nM B3 and assemble RNase resistant particles however, at an almost 10-fold increase in concentration, whereas the other variants fail to assemble. With the longer, 5 PS fragment both wild type CP and the R8A mutant show similar assembly behaviour (Figure 4.19C) which suggests that cooperativity between PSs can overcome the loss of affinity (Patel *et al*, 2015).

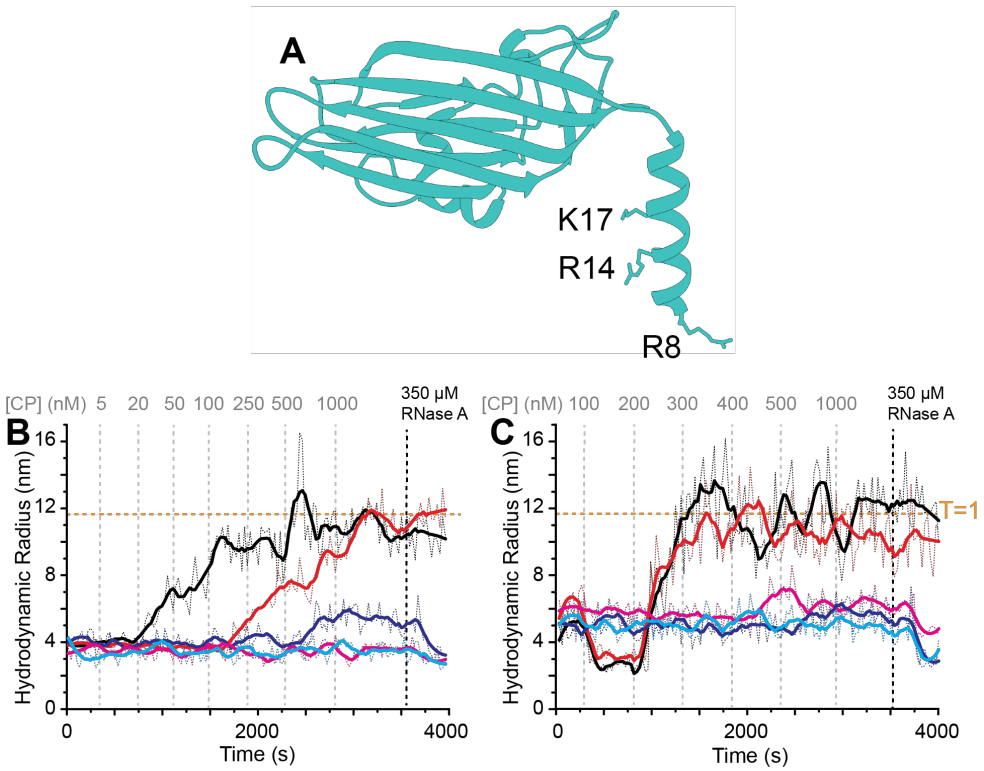


Figure 4.19: Roles of electrostatic interactions and PS co-operativity on assembly. (A) STNV CP with mutated residues highlighted (PDBID 4BCU). **(B)** Wild type CP was titrated into 1 nM B3 (black), R8A into 1 nM B3 (red), R8D into 1 nM B3 (blue), R14A/K17A into 1 nM B3 (cyan) and R14D/K17D into 1 nM B3 (magenta). **(C)** Wild type (black), R8A (red), R14A/K17A (cyan) and R14D/K17D (magenta) titrated into 10 nM PS 1-5. From Patel *et al*, 2017.

Previous work into the assembly mechanism of STNV has shown that in the 5 PS fragment, the spacing between the stem loops affects the formation of capsids as does folding propensity (Patel *et al*., 2015). Combined with the presented results which highlight the importance of sequence specificity in the loop, synthetic cassettes were produced incorporating essential aspects of the natural 127-mer. Base pairs were changed to G-C and existing G-C pairs were inverted until one fold, similar to that seen in the wild type fragment was seen. Spacing between the stem loops was maintained, however the sequence was again altered from the wild type. The sequence of the cassette is vastly

different from the natural sequence with approximately 80% being non-virally encoded. As a result of the sequence changes, the folding propensity of the cassette is stabilised with regards to wild type with all loops having a favourable folding energy (synthetic stabilised SP1-5). A similar cassette was also produced where the central highest affinity PS was given a less favourable folding energy than the wild type (unstable PS1-5). Both cassettes were able to assemble under single molecule conditions however, with differing coat protein concentration dependencies (Figure 4.20). Both show the initial drop in hydrodynamic radius, which is a hallmark of genomic assembly, followed by cooperative assembly to *T*=1. Interestingly, the synthetic stabilised cassette is able to assemble VLPs more rapidly than the wild type fragment suggesting that it is possible to extract critical features required for assembly from a viral genome and make them more efficient.

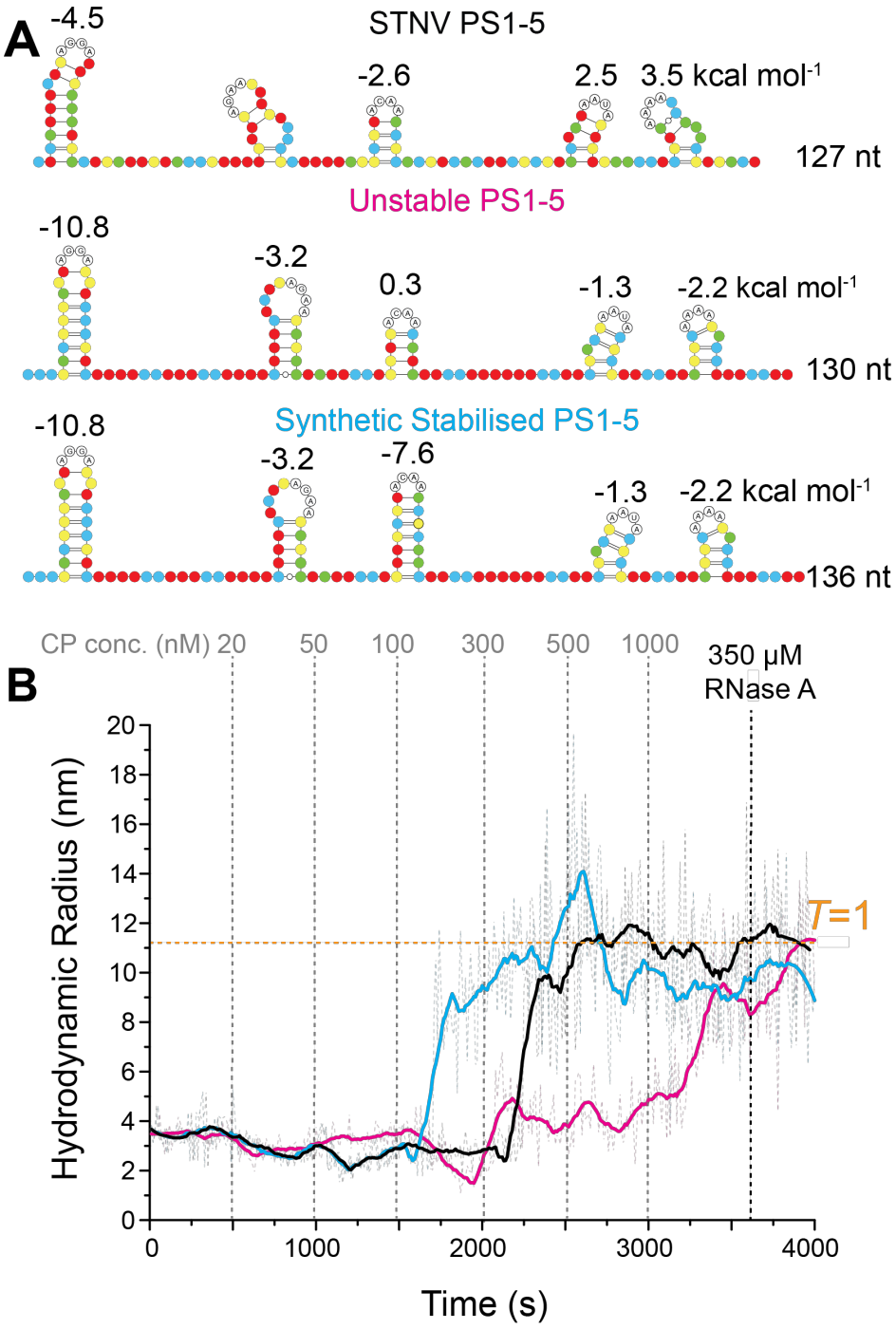


Figure 4.20: Assembly assays with synthetic 127-mer cassettes. (A) Secondary structure and folding energies of PS1-5 and synthetic counterparts. **(B)** STNV CP was titrated into 10 nM of each of PS1-5 (black), unstable PS1-5 (magenta) or synthetic stable PS3 (cyan). From Patel *et al*, 2017.

These synthetic cassettes were used to form chimeric genomes where the native 5 PS, on which these cassettes are based, were replaced with either the unstable or stable constructs (Figure 4.21). The unstable chimera fails to reach the R_h expected for correctly assembled particles and remains RNase sensitive whereas the stabilised chimera readily forms RNase resistant particles and also results in a higher yield. Direct head to head competition with the stabilised chimera and the wild type genome indicates that the chimera out-competes the wild type for coat protein binding (Patel *et al*., 2017).

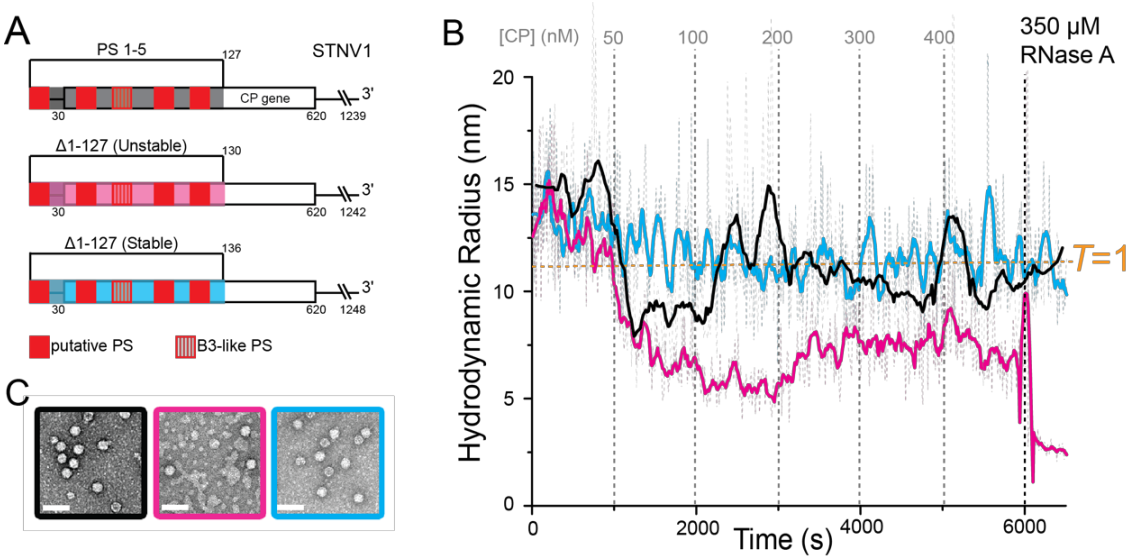


Figure 4.21: Assembly assays with genomic chimeras: **(A)** Schematics of the STNV-1 genome (black) and the modified variants, synthetic stabilised PS1-5+Δ1-127STNV-1 (cyan) and unstable PS1-5+Δ1-127STNV-1 (magenta). PSs are indicated by red boxes, with PS3 as a dashed red box. **(B)** STNV CP was titrated into 1 nM of STNV-1 (black), synthetic, stabilised PS1-5+Δ1-127STNV-1 (cyan) or unstable PS1-5+Δ1-127STNV-1 (magenta), and the resulting R_h was monitored using smFCS. The expected R_h of a recombinant T=1 particle is indicated in orange. **(D)** EM images of the products from **(B)** scale bar = 50 nm. From Patel *et al*, 2017.

The ability to improve the assembly capacity of viral RNA using non-native sequences has wide implications for viruses and VLPs and their use in a number of biological and medical contexts including vaccine development and VLP gene delivery. The practice of vaccination has a history that spans over 300 years (Plotkin, 2014) and is often seen as one of the most successful interventions in modern medicine. The earliest vaccine was developed by Edward Jenner against smallpox in 1798, followed by rabies (1885), typhoid (1896), cholera (1896) and plague (1897) (Plotkin and Plotkin, 2011). Throughout the 20th century and up to present day over 75 vaccines have been developed against previously devastating illnesses (Nabel, 2013).

Following on from the eradication of smallpox in 1980, attention turned to eradication of polio. A killed vaccine developed by Salk in 1953 (Inactivated Polio Vaccine, IPV) reduced the incidence of polio by 95% and upon introduction of the live Sabin vaccine (Oral Polio Vaccine, OPV) in the 1960s this dropped even further. In the 1950s there were between 1-10,000 cases of polio in the UK and by the end of the 1960s incidence was only 1-2 cases per year (Minor, 2015). The IPV is given via an injection which stimulates neutralising antibody production but does not confer intestinal immunity and therefore does not prevent faecal-oral transmission. Conversely, the OPV induces neutralising IgA antibodies in the intestinal tract (Pasetti *et al*., 2011), is easier to administer, and additionally is transmissible to secondary contacts. This however, is one of the major drawbacks, as vaccine strains can revert and cause Vaccine Derived Poliovirus (VDPV) outbreaks almost indistinguishable from wild type that can spread among populations with

insufficient immunity (Famulare *et al*., 2016). In 2017, VDPV paralysed more children than wild type polioviruses (Grassly, 2018) with a reported annual average of 76 cases between 2005 and 2013 (Bandyopadhyay *et al*., 2015).

By discerning the critical assembly features of a virus, it may become possible to produce synthetic assembly substrates for viral coat proteins. These assembled particles could display native antigenicity and be used for the purpose of vaccination without the chance of reversion to wild type potency as the assembly substrate would be replication incompetent. Such work in a model system for ss(+)RNA viruses paves the way for future experiments to be carried out featuring viruses of increased human disease significance.

[This page has been left intentionally blank]

Chapter 5: Investigating the packaging of genomic length STNV RNAs

[This page has been left intentionally blank]

5.1 Introduction

Following *in vitro* assembly of STNV investigations described in the previous Chapter, the work outlined here aims to investigate the *in vitro* assembly of genomic length substrates, define the mechanism of selectivity and understand the consequences for PS cooperativity. By competing genomic length templates of near identical size, with altered packaging signal characteristics, the hope was that the results could be used to interpret selectivity seen *in vivo* and inform the design of synthetic assembly substrates. Additionally, this Chapter will investigate the characteristics of assembly performed in two different ways. Titrations of defined concentration steps mimic single molecule investigations and may more closely reflect *in vivo* conditions where full CP quota is not available at the start of assembly. But does the *in vitro* assembly output suggest that this mechanism is more efficient than all components being readily available at the outset? For these investigations a series of genomic length assembly templates were designed, as illustrated in Figure 5.1.

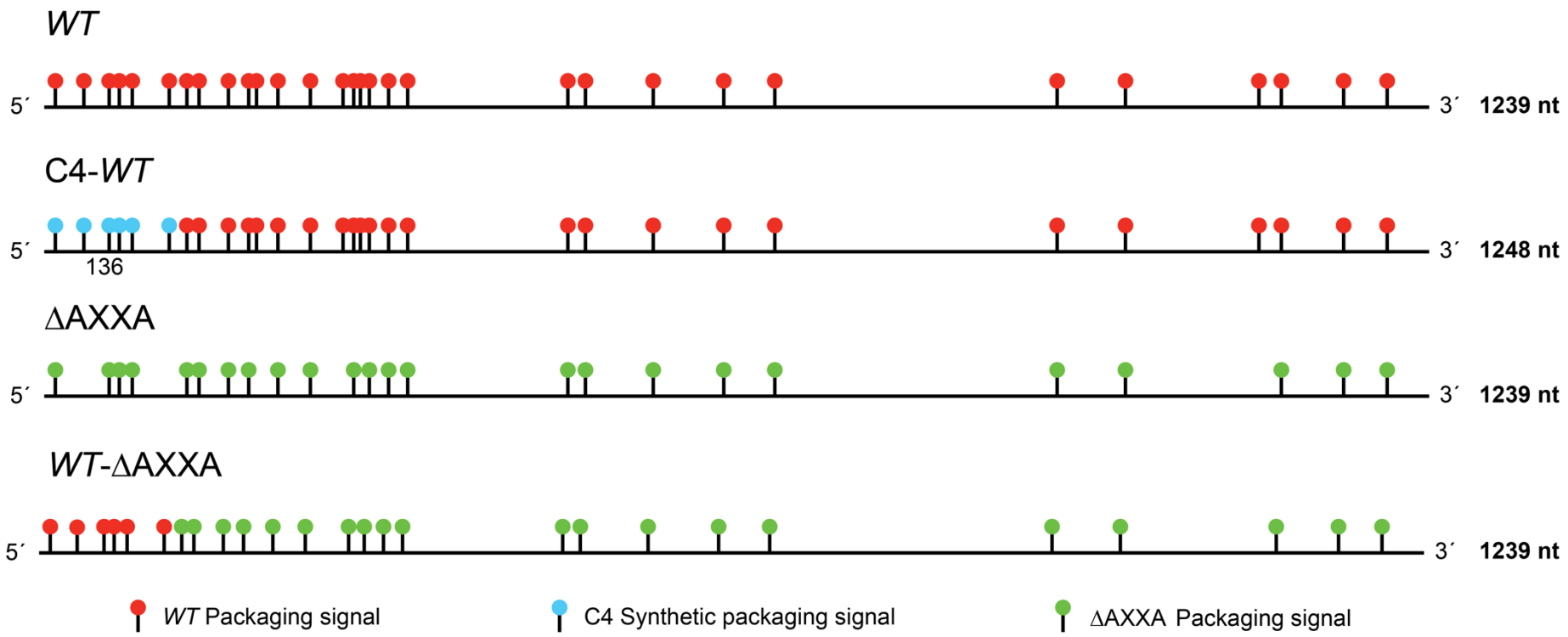


Figure 5.1: Schematic of STNV genomic variants. (A) Wild type STNV-1 genome including stem loops presenting the AXXA motif. **(B)** C4-WT substrate consisting of a modified 5′ 127 nt of the STNV-1 genome found to improve the packaging of STNV fused to 128-1239 nt of STNV-1. **(C)** ΔAXXA fragment comprised of STNV-1 genome with all AXXA presenting stem loops removed. B and C from (Patel *et al*., 2017). **(D)** 5′ 127 nt of the WT genome fused onto ΔAXXA 128 – 1239 nt creating a substrate with only 5 PSs at the 5′ end.

5.2 Results

5.2.1 Generation of transcription templates

WT, C4-WT and ΔAXXA pACYC1 plasmids were obtained from Dr. Nikesh Patel. GBlock® (IDT) encoding the STNV genome variant WT-ΔAXXA in Figure 5.1 was cloned into the BamHI and HindIII sites of a linearised pACYC1 vector using T4 DNA ligase (NEB) as in (Patel *et al*., 2017). WT-ΔAXXA plasmid was transformed into DH5α cells (NEB) and verified by sequencing.

5.2.2 Transcription and purification of genomic length variants

Templates generated in 5.2.1 were used in T7 transcription reactions with the addition of 5' amino-G-monophosphate (Amino-GMP). Amino-GMP allows for post-transcriptional labelling of the RNA with N-hydroxysuccinimide esters (NHS-ester) containing fluorescent dyes. The labelled genomic length transcripts were isolated and analysed by 1% (w/v) denaturing agarose gel (Figure 5.2).

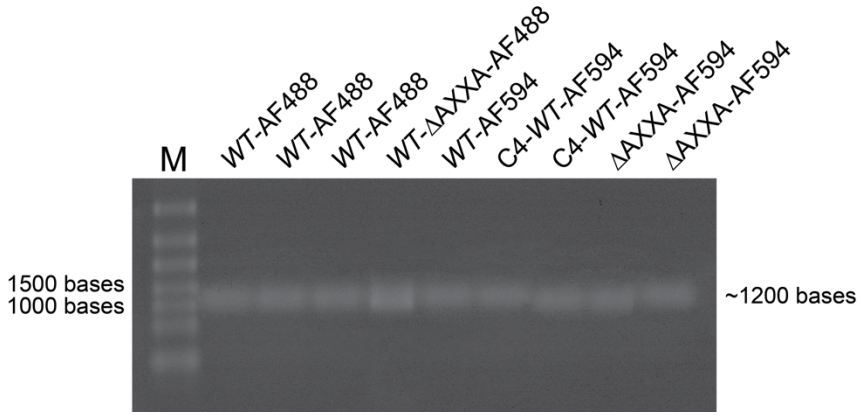


Figure 5.2: Denaturing agarose of genome length transcripts. Denaturing 1% (w/v) agarose of genomic length transcripts viewed using a 532 nm laser which excites both Alexa Fluor-488 and 594, with a Cy3 filter. M = RiboRuler high range RNA ladder (Thermo Fisher).

5.2.3 Reassembly of genomic length variants

5.2.3.1 Wild type STNV-1 Reassembly

Before investigating the assembly capacity of genomic length RNAs in competition for available CP, assembly in the absence of competition was explored. Two, differently labelled WT RNAs were assembled with enough CP to encapsidate both RNAs. In these reassemblies a liquid handling robot was used to add CP to RNA as outlined in Table 5.1. CP was added to the RNA in a series of defined titration steps like those used in published single molecule reassemblies with STNV (Patel, 2015) to a 1 (RNA) to 60 (CP) stoichiometry. Samples were stored at 4ºC for 16 hours to allow the assemblies to run to completion and reach a state of equilibrium.

As the reassemblies are performed at single molecule concentrations in the wells of a 96 well plate, the sample needs concentration before SEC-MALS analysis. Concentration, using Vivaspin20 columns (GE Healthcare) at 4ºC, reduced the sample volume from 19.2 mL to ~500 µL. Post assembly interval and concentration time were kept constant for all samples. The cut-off of the vivaspin20 columns is 50 kDa and therefore free CP and RNA should be removed from the sample leaving only species above this threshold. The sample was then loaded onto a pre-equilibrated TSK column (TOSOH), with associated UV absorbance and light scattering. Use of both Unicorn (GE Healthcare, 2014) and Astra (Wyatt Technology, 2016) software allows for the measurement of absorbance (260 and 280 nm) and light scattering. Fluorescence measurements at 510 nm and 617 nm were taken from the size

exclusion fractions using a nanodrop 3300 fluorospectrometer. The sample was reconstituted from the eluted fractions, concentrated using vivaspin20 columns to ~500 µL and RNase treated to determine if the RNA was accessible. Concentration times were again constant for all samples. Samples were again analysed by SEC-MALS and fluorescence measurements taken of fractions. Absorbance traces from before and after RNase treatment show the sample forms two peaks that elute at slightly different volumes and overlap extensively (Figure 5.3), a common feature across all experiments. Treatment with RNase reduces yields of both peaks but has a larger effect on peak 2. This suggests the peaks comprise two sets of species; correctly formed particles (peak 1), and particles which are enclosed but partially susceptible to RNase (peak 2).

Table 5.1:Experimental conditions for titrations.

Reassembly (RNA:CP)	[RNA] nM	[CP]/step nM	# Steps	Final [CP] nM
1:60	4	40	6	240
2:60	4	20	6	120
2:60 one shot	4	120	1	120

Before RNase treatment peak 2 has higher absorbance readings than peak 1, however, after RNase treatment, this is reversed. Absorbance at both 260 and 280 nm has dropped for both peaks. Yet, peak 2 is 22.8% of the initial height before RNase treatment, whereas peak 1 has only dropped to 77.7% of the initial height, indicating that a significant proportion of the second peak was accessible to the RNase and therefore, incorrectly formed. Yields of peaks 1 and 2 were calculated using unicorn software for the RNase treated samples. Peak 1 accounted for 52.86% of the total, whereas peak 2 accounted for 47.14%. The yield of peak 1 is not as high as could be expected from *in vivo*

assembly, which is assumed to be very efficient, suggesting that experimental conditions are not fully reflective of those *in vivo*. This experiment uses genome length RNA and CP titrations to mimic rising CP concentration. However, perhaps in nature PSs across the genome are not all presented at once, but rather fold as they are transcribed reducing the likelihood of CP binding at undesirable locations and genomes becoming trapped in non-productive conformations.

Light scattering (LS) as measured by detector voltage, from before and after RNase treatment also varies. Before treatment there is a lot of light scattering in peak 2 suggestive of large species, whereas there is relatively little scattering in peak 1 (Figure 5.3A). After RNase treatment, scattering in both Peaks is reduced significantly suggesting that larger species, which scattered the light were RNase susceptible and have been reduced in size (Figure 5.3). Negative stain TEM images from before and after treatment show particles in both peaks however, in peak 2 particles appear aggregated and before RNase treatment many are malformed (Figure 5.3B and C).

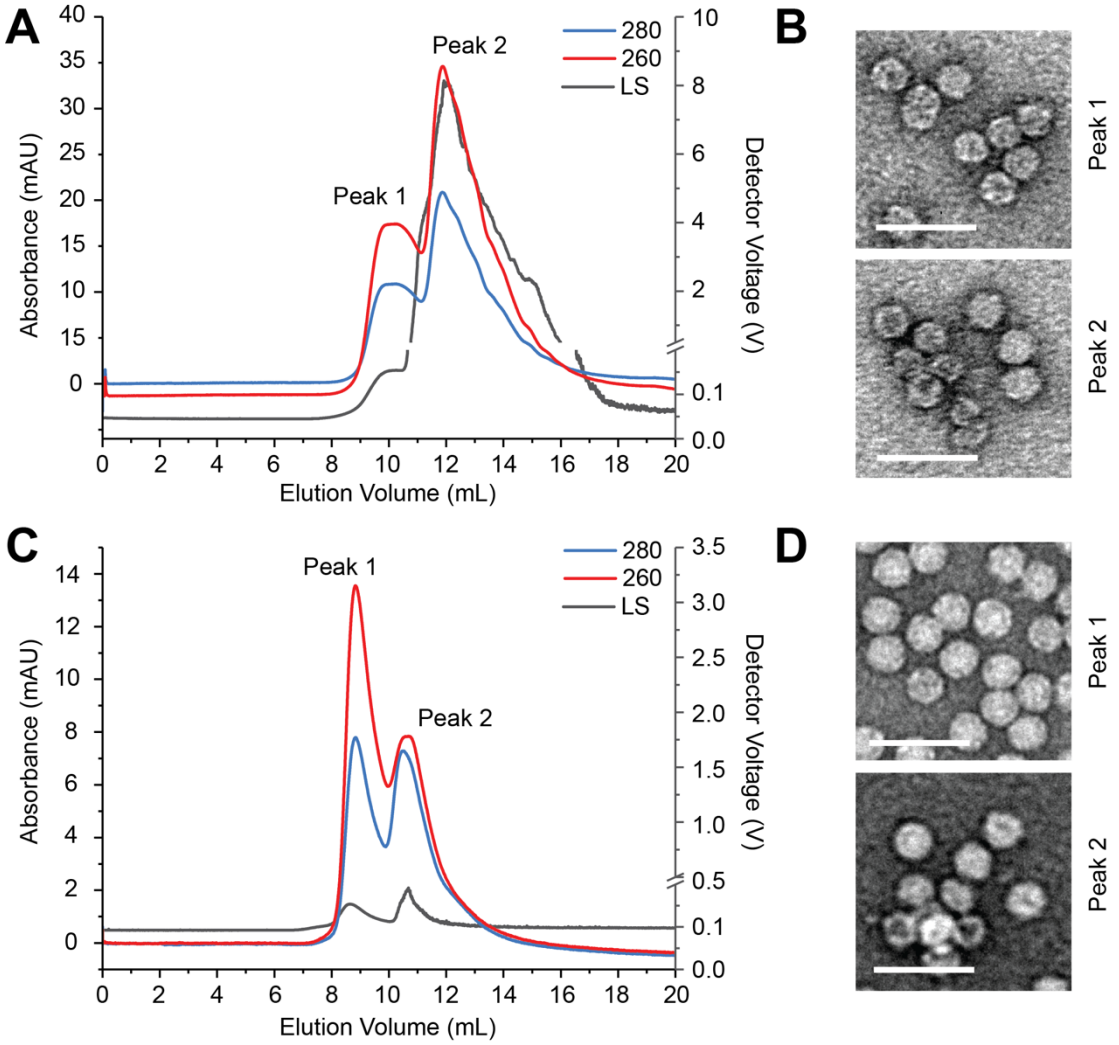


Figure 5.3: Assembly of genomic length RNAs. (A) SEC-MALS of WT-AF488 vs WT-AF594 1:60 (RNA:CP) assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-AF488 vs WT-AF594 competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C Peaks 1 and 2.

SEC-MALS with integrated LS is used to calculate the diffusion coefficient of the sample which is used in turn to calculate the hydrodynamic radius for peaks 1 and 2. Both before and after RNase treatment peak 1 has an R_h of 9.19 and 9.08 nm respectively which is consistent with the STNV particle diameter of ~17 nm (Table 5.2). Peak 2 however, contains larger species both

before (48.96 nm) and after (61.06 nm) RNase treatment. This suggests that some of the large aggregates seen in negative stain TEM images are RNase protected. Additionally, fluorescence of the peaks was measured both before and after treatment with RNase (Table 5.2). Before treatment there is a bias towards the Alexa Fluor labelled 488 RNA in both peaks 1 and 2. However, after treatment both fluorophores are equally represented suggesting they are packaged equally under these conditions. The fluorescence intensity ratio of this non-competitive reassembly was taken to represent 1:1 packaging and was used to normalise the competitive reassembly fluorescence readings in the rest of this chapter as a normalised 617 nm/510 nm ratio. Raw data for all fluorescence readings is shown in Appendix F.

Table 5.2: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	9.19 ± 0.16	9.08 ± 0.18	48.96 ± 2.40	61.06 ± 1.65
λ510 nm	8.9	0.9	14.8	2.8
λ617 nm	7.1	0.9	5.6	2.8

5.2.3.2 *In vitro* competition reassemblies investigating cooperativity between PS

When two, differently labelled RNAs are assembled in a 2:60 (RNA:CP) ratio, i.e. only enough CP to package half the RNA, they are in competition. In these conditions different behaviour is seen in comparison to when there is sufficient CP to theoretically encapsidate all RNA. Variants in Figure 5.1 were created to probe different aspects of PS mediated assembly for STNV as outlined in Table 5.3. Investigations were carried out in the same format as in 5.2.3.1,

with the exception that reassembly was performed in a 2:60 RNA:CP ratio as opposed to 1:60.

Table 5.3: Research questions.

Competitor 1	Competitor 2	Investigating
WT	WT	Are both dye labelled RNAs packaged similarly?
WT	C4-WT	Does the synthetic stabilised 5´ cassette from (Patel et al., 2017) improve packaging due to better initiation?
WT	ΔAXXA	Is RNA with AXXA loops removed packaged as efficiently as WT? i.e. are PSs required for assembly?
WT-ΔAXXA	ΔAXXA	Is the 5′ 127 of WT STNV sufficient to package the full genome or are other PSs throughout the genome also required?

Size exclusion chromatography of two WT RNAs, in competition for CP also produces two peaks both before and after the addition of RNase (Figure 5.4). As seen previously, before RNase treatment peak 2 is higher than peak 1. However, after RNase treatment both peaks are a similar height. There is a drop in absorbance in peak 2, 34.4% of original height, whereas the drop in peak 1 is only 9.1%. The yields also differ, with peak 1 accounting for only 32.46% of the total whereas peak 2 accounts for 67.54% after RNase treatment. This suggests that there are more malformed particles in peak 2

than correctly formed particles in peak 1. If both RNAs are competing equally for an insufficient amount of CP, then assembly could be biased towards unfinished/aggregated particles. Initiation can occur on all RNAs but cannot run to completion. This is supported by the light scattering measurements which indicate that even after RNase treatment peak 2 scatters more light than peak 1, which differs from a 1:60 reassembly.

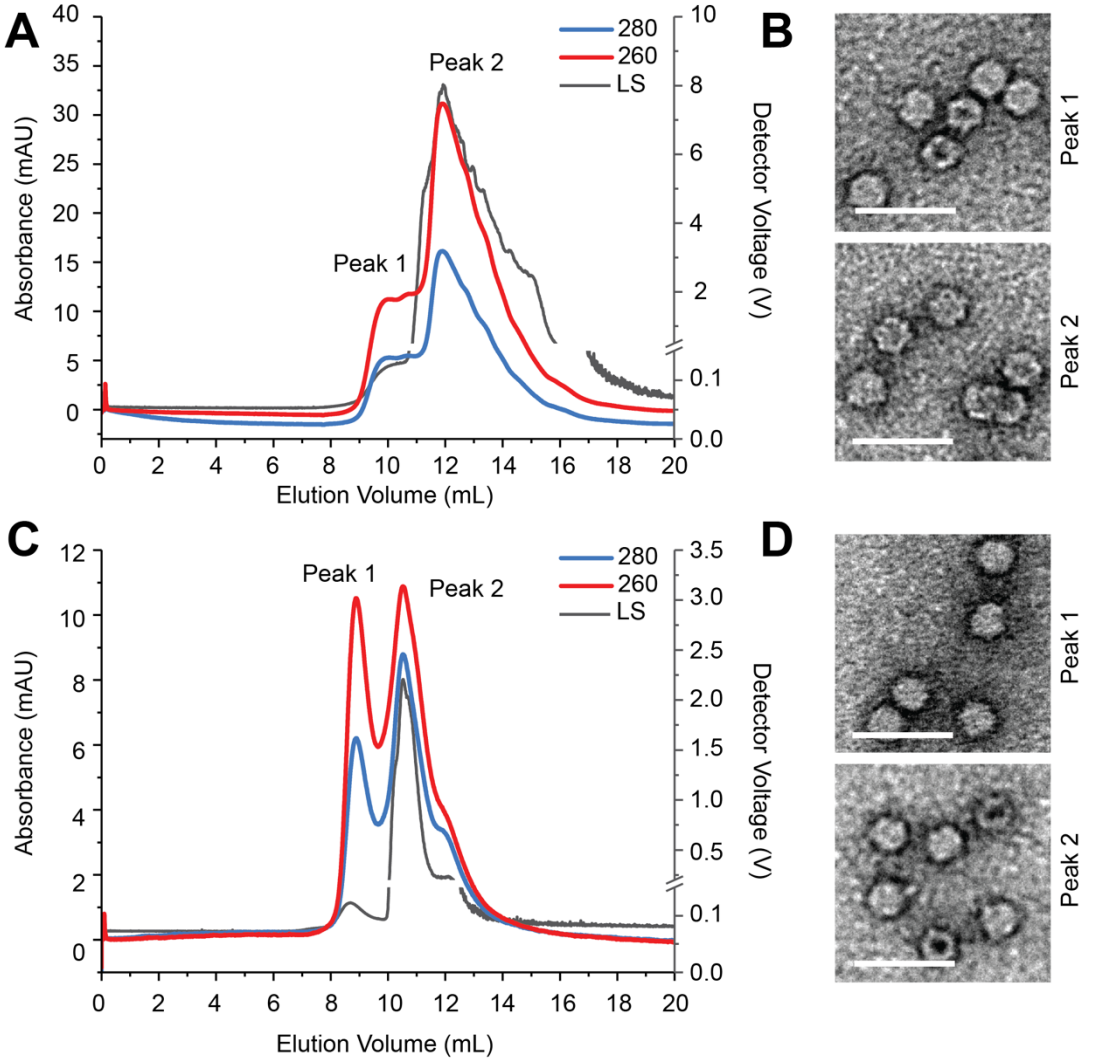


Figure 5.4: Assembly of genomic length RNAs. (A) SEC-MALS of WT-AF488 vs WT-AF594 in a 2:60 (RNA:CP) competition assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-AF488 vs WT-AF594 competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C Peaks 1 and 2.

Both before and after RNase treatment peak 1 has R_h measurements consistent with the STNV particle diameter (8.72 and 8.53 nm respectively). peak 2 again contains larger species with R_h measurements of 48.30 nm before RNase treatment and 59.08 nm after (Table 5.4). Normalised fluorescence measurements suggests that before RNase treatment there is a bias in peak 1 to the AF488 labelled genome and in peak 2 to the AF594 labelled genome, however after RNase treatment 617/510 ratio is 1, suggesting equal packaging of RNAs under these conditions.

Table 5.4: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	8.72 ± 0.16	8.53 ± 0.17	48.30 ± 2.36	59.08 ± 1.32
λ617/510 nm	0.819	1.0	1.74	1.0

In Patel *et al*, 2017, the cassette C4 was shown to trigger assembly of VLPs at a lower CP concentration than the equivalent region in the WT genome, suggesting it was a better assembly substrate. The C4-WT variant is formed from the C4 cassette fused to the 3′ 128-1239 nt of WT STNV-1. This competition aims to investigate if the C4 cassette can improve genome packaging. When competed with WT and analysed by SEC, two peaks are again seen both before and after the addition of RNase (Figure 5.5). Peak 2 is larger before RNase addition and drops by 75% after the addition of RNase. peak 1 in contrast drops 28.6%. Calculated yields indicate that peak 1 accounts for 44.04% of the sample, whereas peak 2 accounts for 55.96%. These values are similar to those seen in the 1:1 assembly which suggests there is little competition between the RNAs. As in the previous competition

peak 2 scatters more light after RNase treatment and negative stain TEM images show aggregated particles.

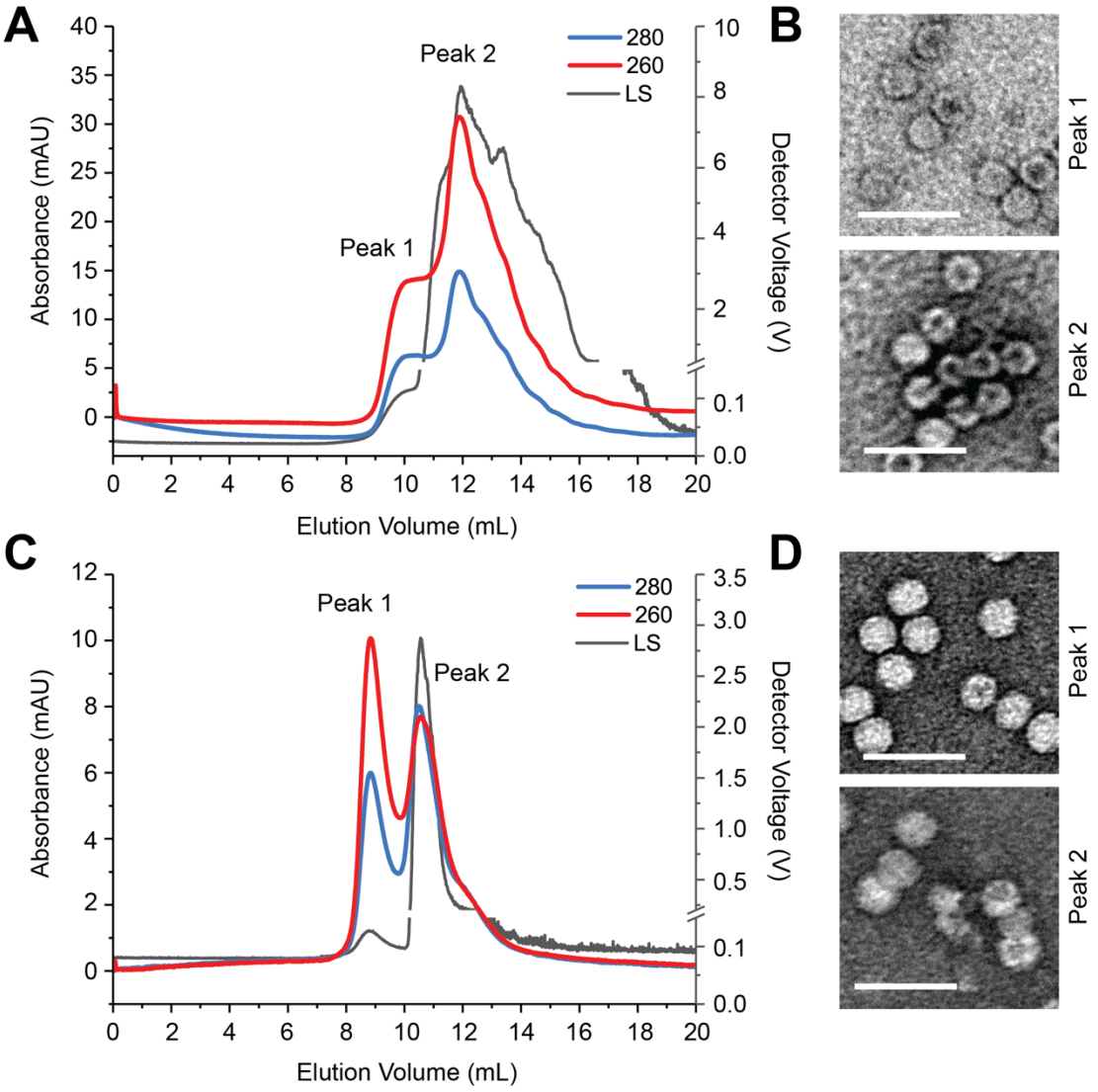


Figure 5.5: Assembly of genomic length RNAs. (A) SEC-MALS of WT-AF488 vs C4-WT-AF594 competition assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-AF488 vs C4-WT-AF594 competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C Peaks 1 and 2.

Peak 1 has R_h measurements consistent with STNV, both before and after RNase treatment (9.01 nm and 8.52 nm respectively) whereas peak 2 contains larger particles with R_h measurements of 49.00 nm before RNase treatment and 87.92 nm after. Normalised fluorescence measurements before and after treatment with RNase indicates that Alexa Fluor 594 is the dominant fluorophore (Table 5.5). This suggests that the C4-WT has been packaged preferentially over the WT genome under these conditions.

Table 5.5: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	9.01 ± 0.17	8.52 ± 0.18	49.00 ± 1.94	87.92 ± 1.97
λ617/510 nm	3.578	22	6.607	8.071

The competition between a WT genome and ΔAXXA, a construct with no AXXA motifs, aims to test the importance of the AXXA loops to packaging, i.e. are they required for assembly. Single molecule investigations with this construct show aggregation and an increase in R_h instead of an R_h collapse (Patel *et al*., 2017). Size exclusion of the WT vs ΔAXXA competition shows two peaks before and after the addition of RNase, where peak 1 has higher absorbance readings after RNAse treatment than peak 2 (Figure 5.6). After treatment the height of peak 1 drops by ~20%, while peak 2 drops ~77%. Peak 1 accounts for 50.15% of the sample and peak 2 accounts for 49.85%. This percentage yield breakdown is reminiscent of the 1:60 RNA:CP reassembly which suggests that there was little competition between RNAs under these conditions.

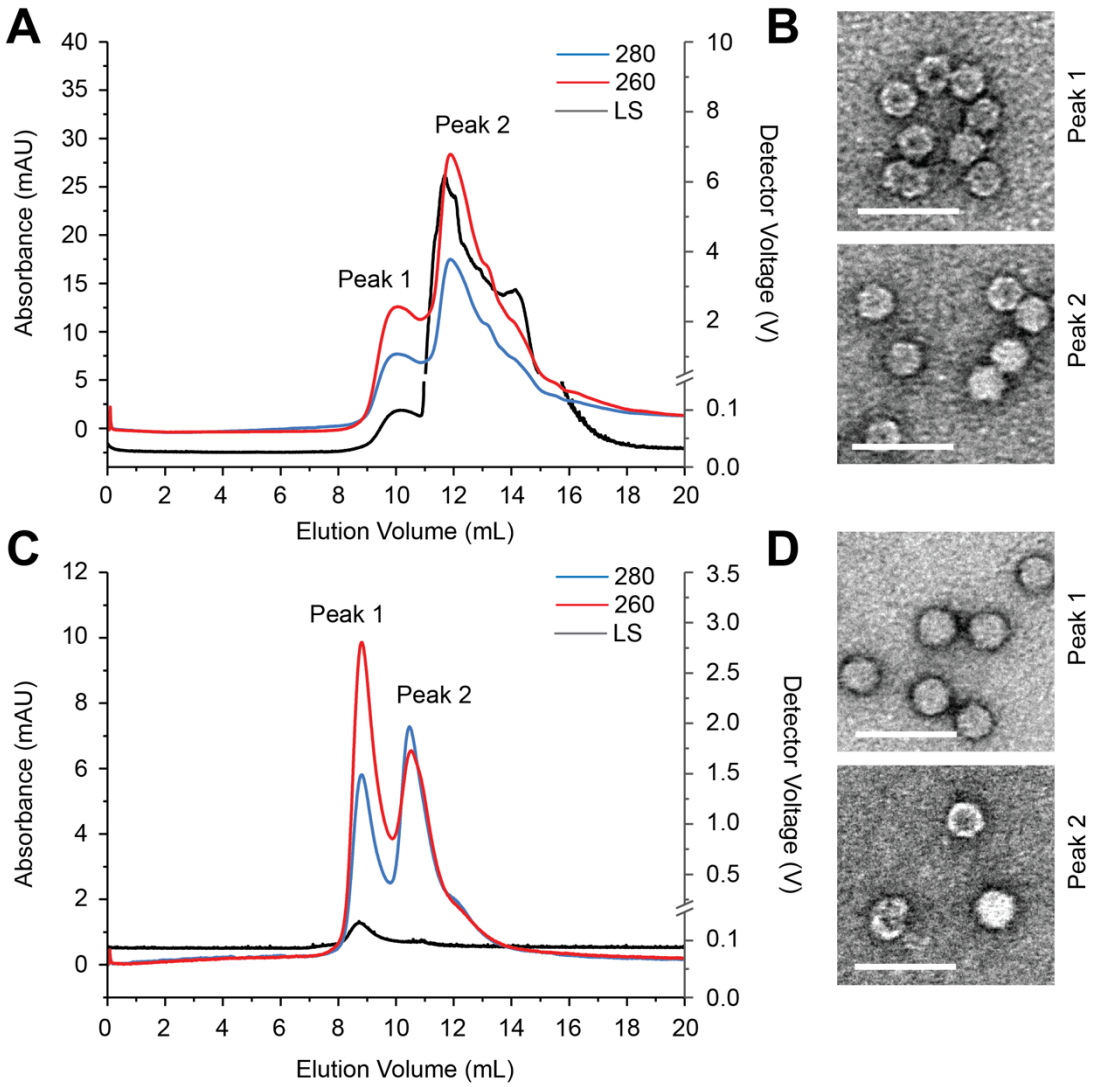


Figure 5.6: Assembly of genomic length RNAs. (A) SEC-MALS of WT-AF488 vs ΔAXXA-AF594 competition assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-AF488 vs ΔAXXA-AF594 competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C Peaks 1 and 2.

Hydrodynamic radius measurements for peak 1 both before (8.85 nm) and after (8.59 nm) RNase treatment are consistent with the expected size for STNV. Peak 2 contains larger particles with R_h of 197.25 after RNase treatment (Table 5.6). The ΔAXXA construct has been shown to form large aggregates in single molecule (Patel *et al*., 2017) which could account for the

large R_h in peak 2. Normalised fluorescence measurements show that Alexa Fluor 488 is the dominant fluorophore in both peaks before and after RNase addition suggesting that under these conditions the WT RNA is packaged preferentially over ΔAXXA.

Table 5.6: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	8.85 ± 0.16	8.59 ± 0.18	48.36 ± 1.76	197.25 ± 16.89
λ617/510 nm	0.262	0.285	1.866	0.24

The 5′ 127 nt of STNV contains the 5 highest affinity PS. The competition of WT-ΔAXXA and ΔAXXA aims to investigate to what extent the 5′ portion orchestrates the assembly process and the involvement of other PSs in efficient assembly. SEC analysis showed 2 peaks both before and after RNase treatment (Figure 5.7). Peak 2 is larger before RNase addition but sees a reduction in peak height of 80% after RNase. Peak 1 in contrast drops 30%. Calculated yields indicate that peak 1 accounts for 38.3% of the sample, whereas peak 2 accounts for 61.7%. These results indicate that there are more malformed particles in peak 2 than correctly formed particles in peak 1. This suggests that in this instance assembly is biased towards malformed particles. Single molecule experiments show that the ΔAXXA construct forms aggregates (Patel *et al*., 2017), which could account for this shift in yield. Light scattering also indicates large particles in peak 2 both before and after RNase treatment.

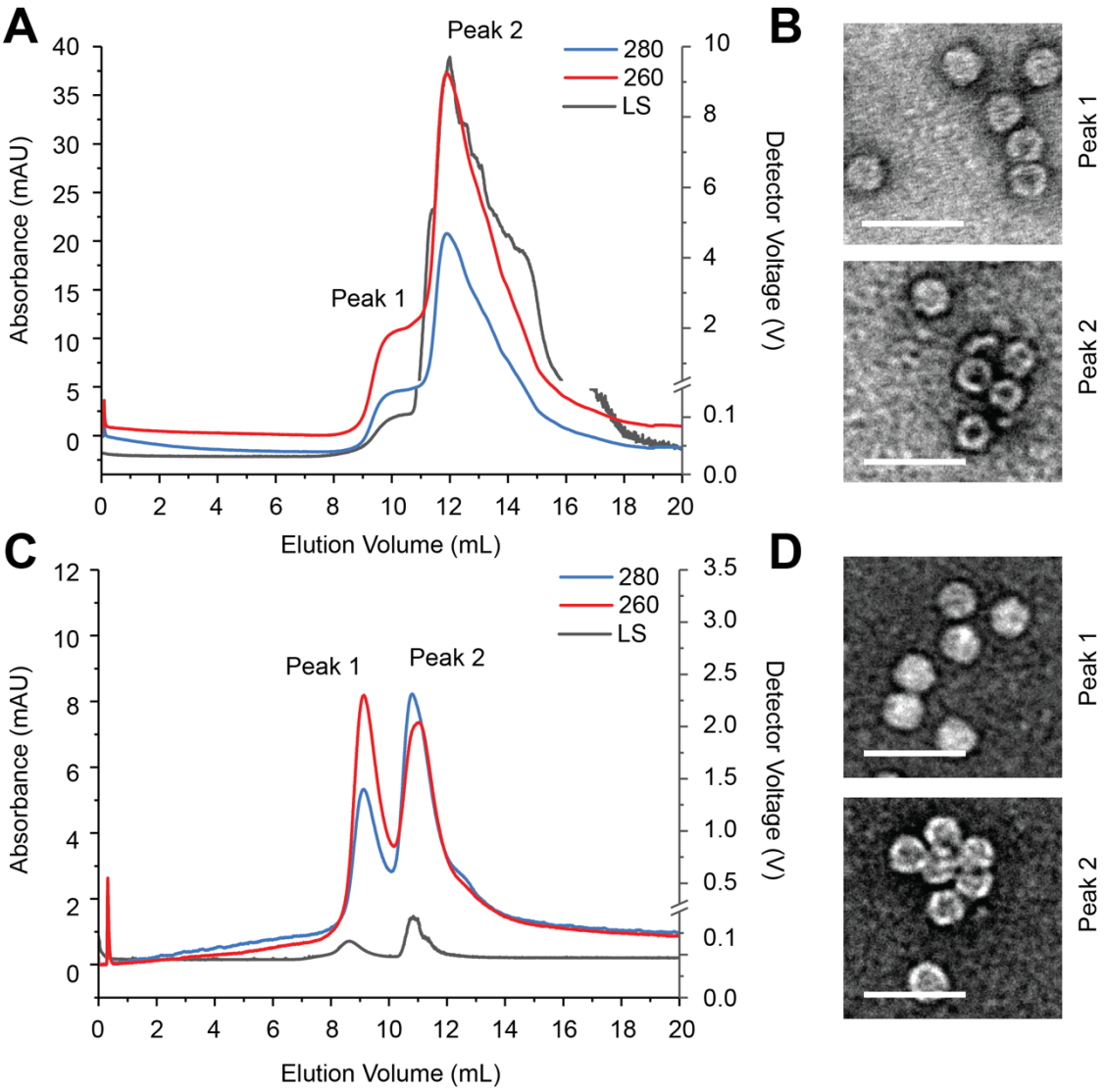


Figure 5.7: Assembly of genomic length RNAs. (A) SEC-MALS of WT-ΔAXXA-AF488 vs ΔAXXA-AF594 competition assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-ΔAXXA-AF488 vs ΔAXXA-AF594 competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C Peaks 1 and 2.

Peak 1 has R_h measurements consistent with STNV, both before and after RNase treatment (8.99 nm and 8.48 nm respectively) whereas peak 2 contains larger particles with R_h measurements of 48.85 nm before RNase treatment and 246.70 nm after (Table 5.7). Due to the known aggregation of ΔAXXA the increase in R_h in peak 2 is not unexpected. Normalised fluorescence readings show that Alexa Fluor 488 is the dominant fluorophore in both peaks before

and after RNase addition suggesting that under these conditions the WT-ΔAXXA RNA is packaged preferentially over ΔAXXA. Taken as a whole the results of this competition suggest that even though WT-ΔAXXA is the superior substrate, the bias towards malformed particles indicates that other PSs outside the 5′ are involved in efficient packaging.

Table 5.7: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	8.99 ± 0.17	8.48 ± 0.18	48.85 ± 2.97	246.70 ± 4.54
λ617/510 nm	0.613	0.494	1.585	0.803

5.2.3.3 Competition reassembly with one-shot addition of CP

In vitro reassembly is usually performed by titrating CP into the sample at defined concentration steps. This is thought to be more reflective of the increase in CP concentration during viral infection *in vivo*. Adding all the CP in at the same time and analysing the products allows us to see what effect this has on assembly under these experimental conditions. The competition was performed with the WT and C4-WT RNAs in a 2:60 (RNA:CP). SEC of RNase untreated sample shows two peaks as for the previous competitions. However, in this instance peak 1 is less defined (Figure 5.8A). Negative stain TEM images also show malformed particles in both sampled areas (Figure 5.8B). After RNase treatment the peak morphology becomes vastly different to the previous results (Figure 5.8). Peak 2 is larger before and after RNase addition and sees a reduction in peak height of 65% after treatment. The reduction in peak 1 after treatment is 54%. Calculated yields indicate that peak 1 accounts for 24.93% of the sample, whereas peak 2 accounts for 75.07%. Light

scattering before and after RNase treatment indicates larger species in the sample and negative stain TEM images show malformed and aggregated particles.

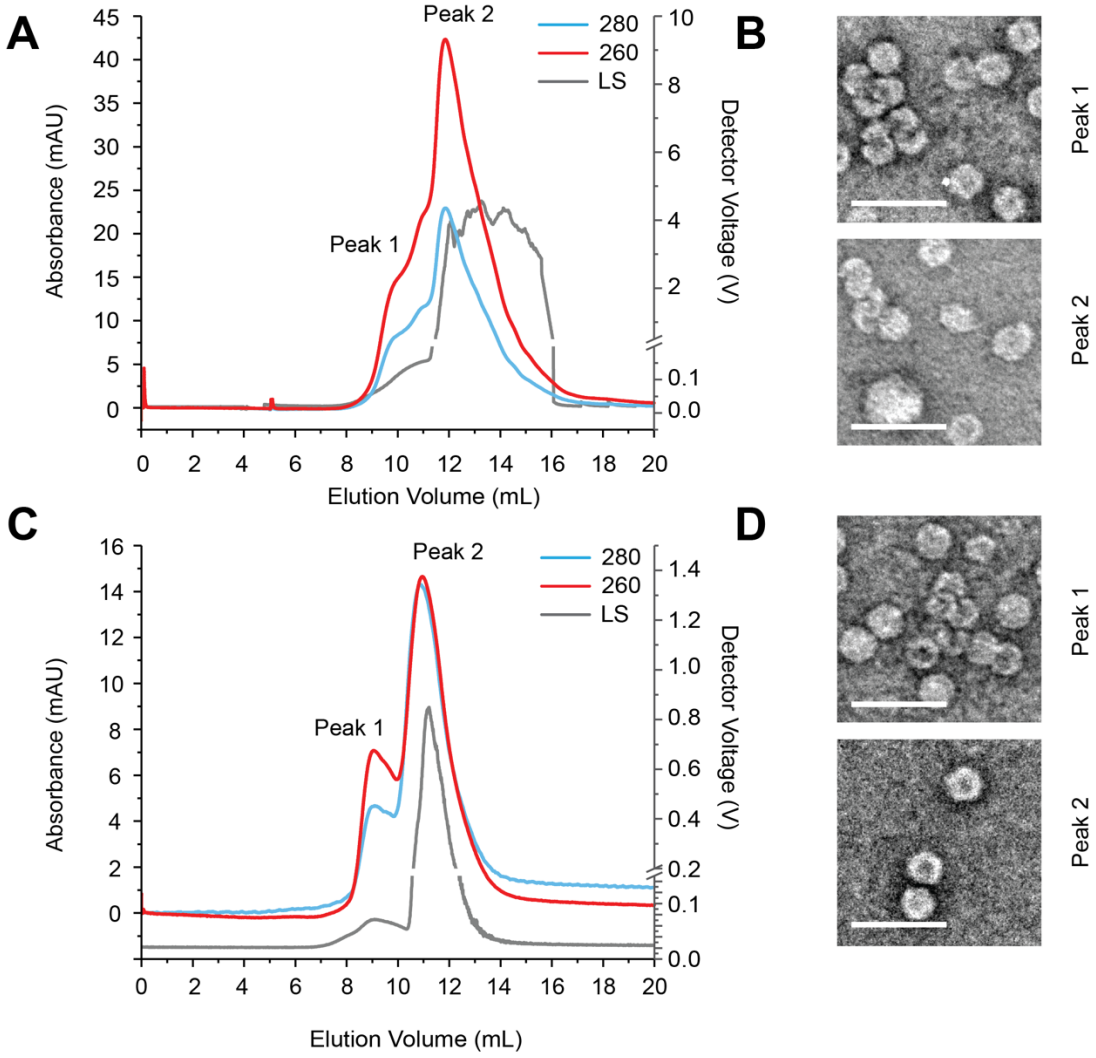


Figure 5.8: One shot assembly of genomic length RNAs. (A) SEC-MALS of WT-AF488 vs C4-WT-AF594 one shot competition assembly measuring absorbance at 260 (Blue) and 280 nm (Red) accompanied by light scattering data (LS, Grey). **(B)** Negative stain TEM images of fractions from A peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000X. **(C)** SEC-MALS of WT-AF488 vs C4-WT-AF594 one shot competition assembly after incubation with RNase A. Legend as in A. **(D)** Negative stain TEM images of fractions from C peaks 1 and 2.

In this instance, the R_h measurements for peak 1 both before and after RNase are slightly larger than seen in the previous competition experiments (Table

5.8). Before RNase treatment R_h is 11.07 nm and after treatment 10.13 nm. This is larger than the average peak 1 R_h before and after RNase of 8.95 nm and 8.64 nm respectively. Normalised fluorescence measurements before and after treatment with RNase indicate that Alexa Fluor 594 is the dominant fluorophore. This suggests that as in the previous WT vs C4-WT competition, C4-WT has been packaged preferentially under these conditions. However, under these conditions assembly is not as efficient in the production of correctly formed particles.

Table 5.8: Hydrodynamic radii and fluorescence measurements.

	Peak 1	Peak 1 post RNase	Peak 2	Peak 2 post RNase
R_h (nm)	11.07 ± 0.16	10.13 ± 0.21	36.06 ± 1.33	114.14 ± 1.62
λ617/510 nm	2.998	5.967	5.640	1.610

5.3 Discussion

Table 5.9: Summary of Peak Yields after RNase treatment.

RNAs and RNA:CP ratio	Peak 1 % yield	Peak 2 % yield
WT *vs* WT assembly (1:60)	52.86	47.14
WT vs *WT* (2:60)	32.46	67.54
WT vs C4-WT (2:60)	44.04	55.94
WT vs ΔAXXA (2:60)	50.15	49.85
WT-ΔAXXA vs ΔAXXA (2:60)	38.38	61.62
WT vs C4-WT (2:60) one shot	24.93	75.07

When two wild *type* genomes, with different labels, are reassembled with enough CP to package both genomes and treated with RNase, two peaks are observed in SEC-MALS (Figure 5.3C). Peak 1 contains particles with an R_h similar to WT STNV while those in Peak 2 are larger. Fluorescence measurements after RNase treatment show an almost equal representation of both fluorophores in both peaks. This suggests that in a non-competitive reassembly under these conditions both substrates are packaged equally. Peak 1 and 2 account for 52.86% and 47.14% of the sample respectively. As *in vivo* assembly is thought to be very efficient, the yields are perhaps lower than expected. This suggests that the experimental conditions are not wholly reflective of those *in vivo*. The mechanism of assembly could be a co-transcriptional process where PSs form on nascent chains and are therefore not available all at once. Despite this, the non-competitive assembly does give a benchmark for WT assembly under these conditions, and almost 50:50 split between peak 1 and peak 2. This allows for inferences to be made about the variant RNAs as assembly substrates. Fluorescence measurements made in this assembly can also be used to normalise the fluorescence measurements

of competitive reassemblies as in the absence of competition packaging should be equal.

When identical substrates are in competition for available CP there is an exquisite difference in behaviour. As seen in the non-competitive reassembly, competition between wild type genomes with different fluorescent labels produces two peaks after the addition of RNase. Yet, these peaks are different. Peak 1 accounts for 32.46% of the sample and peak 2, 67.54%. Normalised fluorescence readings show an equal representation of both fluorophores in after RNase treatment suggesting an equal packaging of assembly substrates under these conditions. The shift in morphology to a dominant peak 2 could be due to both substrates having equal ability to assemble with the CP and form particles. As both constructs are able to do this, assembly begins equally on both substrates, and a larger proportion of particles are unable to complete the assembly process faithfully. The WT vs WT competitive reassembly also puts the other results in context. Both substrates are equally matched in their ability to bind CP. However, what happens when one is better substrate than the other?

When competition occurs with substrates of differing characteristics more complex behaviour is seen. Previously, a synthetic cassette was identified which could improve the packaging of the STNV-1 genome (Patel *et al.*, 2017). When this chimeric genome (C4-WT) is used in competition with the WT, treated with RNase and analysed by SEC-MALS it differs subtly from the WT-WT competition. Peak 1 accounts for 44.04% and peak 2, 55.94% suggesting

a shift back towards correctly formed VLPs. Normalised fluorescence measurements for both peaks have a higher value for AF594 labelled species, suggesting that in this instance the balance of assembly is offset in the favour of the C4-WT which is able to out-compete the wild type. In this case, after RNase treatment normalised results suggest that C4-WT vs WT packaging is 22:1. Under these conditions, a synthetic stabilised 5′ 127 nt is able to shift assembly towards correctly formed particles highlighting the importance of the 5′ end of the genome.

When the wild type genome is in competition with ΔAXXA, two peaks are again seen after incubation with RNase. Peak 1 accounts for 50.15% of the sample and peak 2, 49.85%. This behaviour is reminiscent of the WT vs C4-WT competition and the non-competitive reassembly, suggesting an unequal balance in substrate assembly competency. The ΔAXXA construct represents a genome without packaging signals, where all the identified AXXA containing loops have been altered to UXXU. If these stem loops were of no importance there should be equal representation of both fluorophores in the final peaks as seen for the WT-WT competitive assembly. However, the normalised fluorescence measurements indicate that AF488 is the dominant species in both peaks suggesting that the WT genome is packaged preferentially over ΔAXXA.

The WT-ΔAXXA construct is comprised of the 5′ 127 nt of the wild type STNV-1 genome fused to 128 – 1239 nt from the ΔAXXA fragment creating a substrate with PSs only at the 5′ end. When this construct is in competition

with ΔAXXA, which contains no PSs, peak morphology shifts again to a dominant peak 2. Peaks 1 accounts 38.38% of the sample and peak 2, 61.62%. As ΔAXXA is known to aggregate in single molecule larger species are not unexpected however the presence of five WT PSs were not able to restore peaks to WT yields. Normalised fluorescence measurements indicate that the dominant fluorophore in peak 1 is AF488, representing WT-ΔAXXA, whereas the measurements for peak 2 show a more equal distribution of AF488 and AF594. This suggests that the presence of the 5′ 127 nt of the STNV-1 does make the WT-ΔAXXA construct more successful than the ΔAXXA construct but does not restore the selectivity to WT levels.

For all the reassemblies in this chapter, the species with the smaller hydrodynamic radius as measured by light scattering exits the column prior to the larger species. This is more obvious after RNase treatment where the sample is more clearly divided into two peaks. As this is SEC, separation in this manner seems counter intuitive and could indicate that the column matrix is interacting in some way with the reassembled sample. The larger particles in the second peak could be aggregated VLPs, incomplete or multiple capsid-like structures sharing one or more RNAs. The RNA, which could be exposed in these particles could interact with the matrix, retarding these structures within the column. Increased salt in the buffer used for SEC could prevent this interaction with the matrix however, reassembled particles are sensitive to salt concentration and could dissociate under increased salt conditions.

The previous competitions are all the result of titrating of CP into RNA. This method is thought to be more representative of *in vivo* processes where CP accumulates during replication thus allowing assembly to proceed along a defined pathway(s). To investigate the effect of titration on the assembly of VLPs the WT vs C4-WT reassembly was repeated with one single CP addition. Again, analysis with SEC-MALS shows two peaks in the sample. On this occasion the second peak yield is significantly larger than the first (75.07% and 24.93% respectively) which suggests a larger proportion of incomplete particles. Normalised fluorescence measurements show that once again C4-WT is the dominant species however, unlike the titrated reassembly where C4-WT outcompeted WT 22:1, in the absence of titration the ratio is only 5.96:1.

By comparing the absorbance measurements for both the titration and one-shot WT vs C4-WT competition the differences become more apparent. As seen in Figure 5.9, the 260 and 280 nm readings for the titrated competition (solid red and solid blue respectively) show two peaks with the first being larger than the second. In contrast, the one-shot sample (280 nm dashed blue, 260 nm dashed red) has a significantly larger second peak, which is also broader than in the titration. Additionally, the first peak has a shallow tail when compared to the titration peak, which is symmetrical in shape, suggesting the sample contains particles of a wider range of sizes than seen previously.

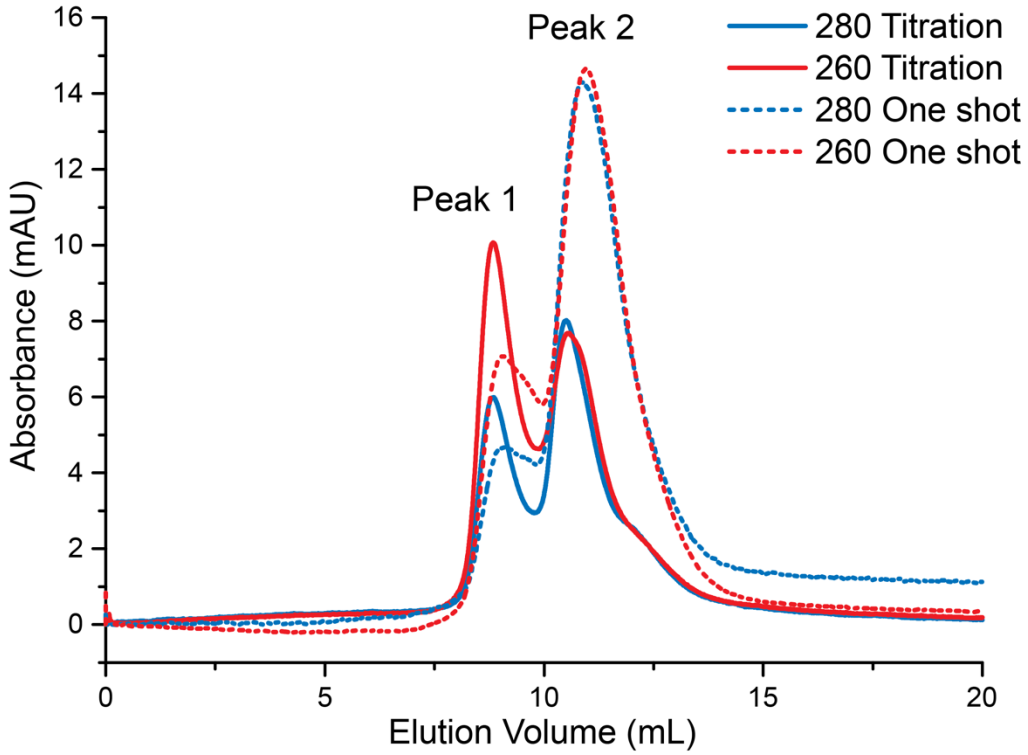


Figure 5.9: Comparison of titrated and one-shot competition reassemblies after RNase treatment. SEC-MALS of WT-*AF488* vs C4-WT-AF594 competition performed as both a titration (solid lines) and a one-shot procedure (dashed lines) measuring absorbance at 260 (Blue) and 280 nm (Red).

Taken as a whole, these results suggest that packaging signal stem loops are important for genome packaging and therefore viral assembly. Constructs containing AXXA stem loops are packaged preferentially over their non AXXA containing counterparts. If the assembly mechanism were solely due to electrostatics there would be no selectivity and both fluorophores would be equally represented in the sample. While the 5′ 127 nt of STNV-1 contains what is thought to be the highest affinity PS (PS3) it is not enough to restore packaging to WT levels under these conditions. This suggests that other PSs distributed throughout the genome may be important in this mechanism. Additionally, these results show that titration of CP in controlled concentration

steps results in a higher proportion of correctly formed VLPs. This suggests that the mechanism of regulated protein concentration may be more representative of that seen in *in vivo* assembly, where CP binds to PSs of varying affinity in a specific order leading to nucleation at specific sites. This in turn directs assembly down a small number of distinct assembly pathways and regulates the assembly process.

Chapter 6: Production and characterisation of Alphavirus core proteins in *E.coli*

[This page has been left intentionally blank]

6.1 Introduction

Unlike other viruses, for which an assembly mechanism regulated by packaging signals is a relatively new concept, packaging signals are widely accepted in the alphavirus literature (Frolova *et al.*, 1997; White *et al.*, 1998; Weiss, Geigenmüller-Gnirke and Schlesinger, 1994b). However, unlike the PSs discussed thus far, they are quite large. For example, a PS was identified in Sindbis virus (SINV) encompassing nucleotides 948-1226, making it 178 nt in length. Deletions in this region alter SINV packaging efficiency. However, constructs lacking this region entirely could still be packaged with lower efficiency suggesting that other areas/sequences may influence this process (Strauss and Strauss, 1994). This is consistent with multiple dispersed PS sites, as seen in STNV and other viruses studied in the Stockley Lab.

A more recent investigation (Kim *et al*., 2011) using Venezuelan equine encephalitis virus (VEEV) identified a region between 856 and 1150 nt of the genome that is required for packaging of infectious particles. It also found that the location of this region within a construct was typically independent of function. The structure of the PS identified is predicted to fold into multiple stem loops each displaying a GGG motif. A motif which is shared across PSs identified in different alphaviruses. An artificial PS with a non-WT nt sequence, but still presenting a number of GGG loops, was able to package to levels comparable to WT as measured by viral titre (Kim *et al*., 2011).

Alphavirus packaging investigations, and the identification of packaging signals, have typically involved using sequence analysis of naturally occurring defective interfering RNAs to identify areas of conservation (Weiss *et al*., 1980; Levis *et al*., 1986; Bredenbeek *et al.*, 1993). In some examples the capsid binding ability of these sequences has been examined *in vitro* (Weiss *et al*., 1989; Weiss, Geigenmüller-Gnirke and Schlesinger, 1994a) or packaging of genomic fragments containing these sequences is investigated with structural genes provided in *trans* (Kim *et al*., 2011). These methods are unlike those used to identify PSs in other ssRNA viruses (Bunka *et al*., 2011; Shakeel *et al*., 2017; Rolfsson *et al*., 2016; Stewart e*t al*., 2016) and even pre-genomic RNA of some DNA viruses (Patel *et al*., 2017). These investigations use SELEX to determine sequences that bind to the interior surface of CPs, which are compared to genomic sequences to identify putative PSs. This approach requires the production and/or purification of viral capsid/core proteins and this was the aim of the work described in this Chapter.

6.2 Results

6.2.1 Expression of Alphavirus core proteins

6.2.1.1 Production of Alphavirus core protein plasmids

Plasmids containing codon optimised (Integrated DNA Technologies, 2018.) sequences of selected alphavirus core proteins (Table 6.1) were obtained from Dr. Chris Kirkham in the laboratory.

Table 6.1: Alphavirus core proteins and Genbank accession numbers of core protein genes.

Sequence Name	Genbank Accession number	Vector	Name
Chikungunya virus	ADJ93861.1	pET-22b(+)	pET-22b(+)-CHIKV
Semliki Forest virus	AAM64227.1	pET-22b(+)	pET-22b(+)-SFV
Ross River virus	ABB53381.1	pET-22b(+)	pET-22b(+)-RRV
Sindbis virus	NP_740673.1	pET-22b(+)	pET-22b(+)-SINV
Western equine encephalitis virus	P13897.1	pET-22b(+)	pET-22b(+)-WEEV

6.2.1.2 Alphavirus core protein expression in *E.coli*

The plasmids containing the core protein genes were used to transform BL21-CodonPlus(DE3)-RIL cells. These cells contain a pACYC-based, ColE1-compatible plasmid containing genes encoding tRNAs present in low amounts in *E.coli*. This plasmid also carries chloramphenicol resistance, therefore cells transformed with pET22(+)CHIKV must be grown in the presence of both chloramphenicol and carbenicillin (Agilent Technologies, 2015).

Single carbenicillin and chloramphenicol resistant colonies were used to inoculate 10 mL starter cultures. These were used for inoculation of 6x500 mL flasks of media, Luria-Bertani (LB) or Terrific broth (TB). The six flasks were used to test expression conditions for CHIKV Cp (Table 6.2). When the optical density at 600 nm (OD_{600}) was ~0.6 for LB or ~2.5 for TB (~16 h) flasks 3 and 4 were heat shocked at 42˚C for 30 minutes then cooled to expression temperature for 20 minutes. 1 mL was removed from each flask (pre-induction sample) before induction with IPTG (1 mM final concentration). The flasks were incubated with shaking for approximately 16 h at the temperatures indicated (Table 6.2). Later, a further 1 mL was removed from each flask. Cells from pre and post induction were centrifuged and resuspended in Laemmli sample buffer (Thermo Fisher) before SDS-PAGE analysis.

Table 6.2: Table of CHIKV Cp expression conditions

Flask #	Media	Heat shock	Expression temp
1	LB	No	37˚C
2	LB	No	28˚C
3	LB	Yes, 42˚C	37˚C
4	LB	Yes, 42˚C	28˚C
5	TB	No	37˚C
6	TB	No	28˚C

Of all the conditions tested, only TB incubated at 37ºC produced a band of the expected size for CHIKV Cp (CHIKV Cp = 29417.94 Da) (Figure 6.1A and Appendix E). Following this, an expression time course was carried out to determine the optimal time for protein harvesting and to assess protein degradation. Single colonies of BL21-CodonPlus(DE3)-RIL transformed with pET22b(+)-CHIKV were used to inoculate a 10 mL LB overnight starter culture. This was added to 500 mL TB (100 µg/mL carbenicillin, 34 µg/mL

chloramphenicol) and once OD_{600} reached ~2.5 the sample was induced with 1mM IPTG. An aliquot was removed from the flask periodically for 24 hours and stored as a cell pellet at -20˚C until samples were analysed by SDS-PAGE. CHIKV core protein expression is seen after induction with IPTG (Figure 6.1B) with peak levels seen 6 hours post induction. After this time protein levels did not rise perceptibly, and at 24 hours levels of expressed protein seemed to drop. An expression time of 6 hours post induction was used for all future alphavirus core protein expression.

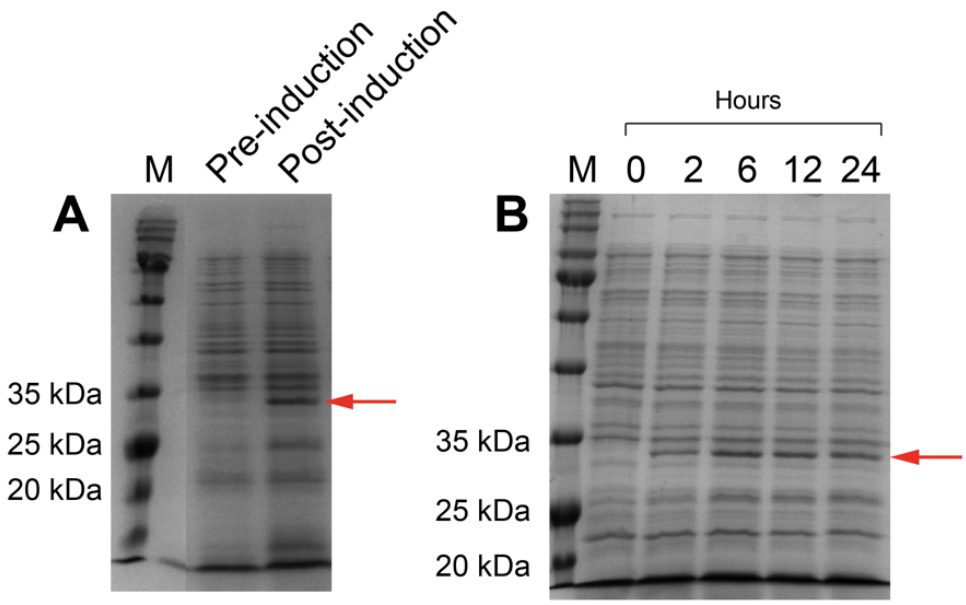


Figure 6.1: CHIKV core protein expression and time course. (A) SDS-PAGE of CHIKV Cp expression in BL21-CodonPlus(DE3)-RIL cells, at 37ºC in TB. Expected size of CHIKV Cp ~ 30 kDa. **(B)** Expression time course of CHIKV. M = BLUeye protein ladder (Geneflow). For full version of A see appendix E.

In addition to BL21-CodonPlus(DE3)-RIL cells, pET22(+)-SFV and pET22(+)-RRV were transformed into Rosetta (DE3) competent cells. These cells contain a chloramphenicol resistant plasmid (pRARE) encoding a wider range of low copy tRNAs than found in BL21-CodonPlus(DE3)-RIL (Novagen, 2018).

Single carbenicillin and chloramphenicol resistant colonies were used to inoculate 10 mL LB starter cultures. These were used to inoculate 500 mL flasks of LB. When the OD_{600} was ~0.6 samples were heat shocked at 42˚C for 30 minutes then cooled to 28°C for 20 minutes. 1 mL was removed as a pre-induction sample before induction with IPTG (1 mM final concentration). Flasks were then incubated with shaking for approximately 6 hours at 28°C. Previous expression of RRV core proteins by Tellinghuisen (1999) was performed at 28°C in LB (Tellinghuisen *et al*., 1999). Following a 6-hour incubation, samples were removed and analysed by SDS-PAGE (Figure 6.2). As seen in Figures 6.2A and B, both proteins are expressed to a higher level in the Rosetta cells in LB.

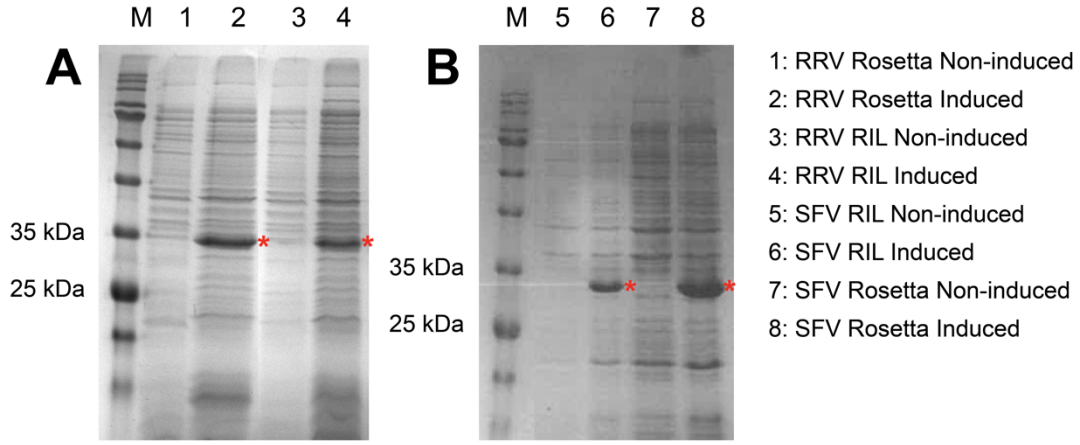


Figure 6.2: Expression of RRV Cp and SFV Cp in Rosetta and RIL cells. (A) SDS-PAGE of RRV Cp expression in Rosetta and RIL cells at 28°C. **(B)** SDS-PAGE of SFV Cp expression in Rosetta and RIL cells at 28°C. M = BLUeye ladder (Geneflow). Stars indicate induced Cp expression.

Expression of both SINV and WEEV core proteins was unsuccessful in both BL21-CodonPlus(DE3)-RIL cells and Rosetta cells (Appendix E). Trials were also carried out in LB and TB at a range of different temperatures without appreciable protein production (Appendix E). Optical density of the cultures was monitored to investigate bacterial growth. Increasing OD_{600}

measurements indicate that cells do reach the stationary phase. However, induction with IPTG at this point did not result in the expression of core proteins (Appendix E). This behaviour suggests that expressed protein is toxic to the cells.

Plasmids were transformed into Lemo21 cells, suitable for difficult/toxic proteins, and plated out in the presence and absence of IPTG. In the absence of IPTG colonies grew for both samples however, for both constructs induction of protein expression affected bacterial growth (Figure 6.3). Expression in Lemo21 cells is tuneable and can be modulated by the inclusion of rhamnose in the media. In these cells T7 lysozyme expression, the inhibitor of the T7 promoter, is under the control of a rhamnose promoter. For both SINV and WEEV, inclusion of rhamnose resulted in the growth of colonies, however, upon inclusion of rhamnose into the media, there is no protein expression (Appendix E).

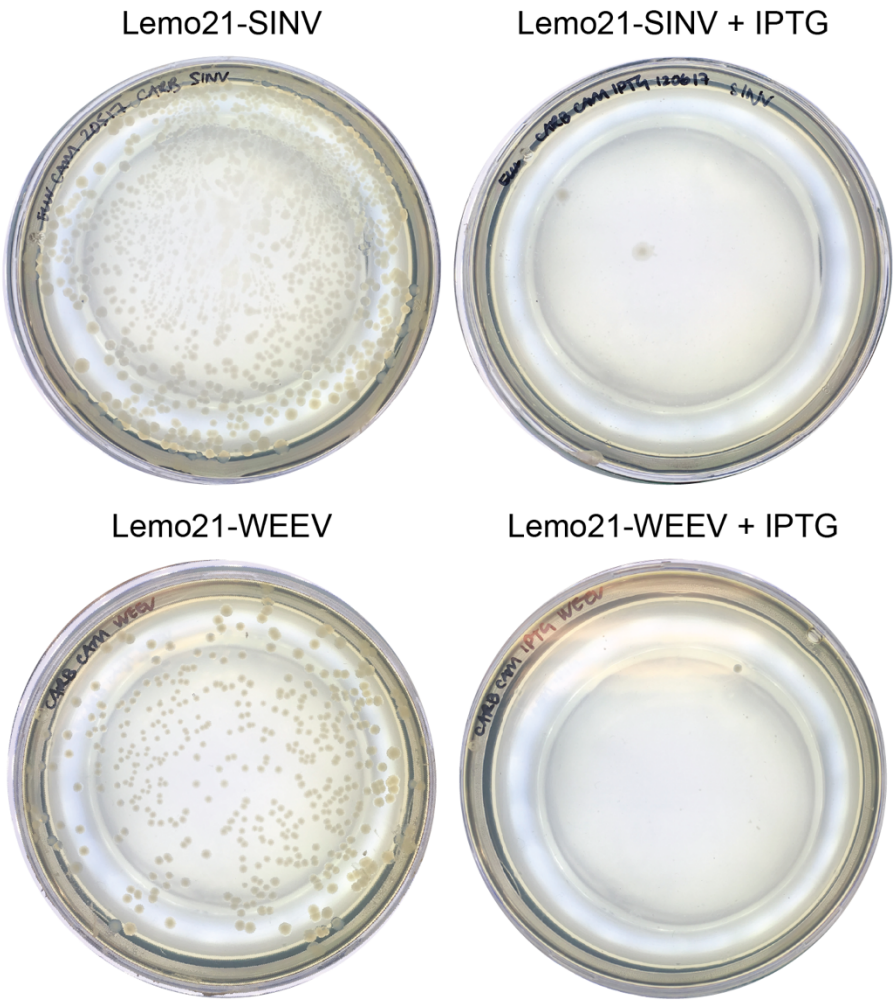


Figure 6.3: Toxicity of SINV and WEEV core proteins. Lemo21 expression of SINV Cp and WEEV Cp in the presence and absence of IPTG.

6.2.2 Core protein purification

6.2.3.1 CHIKV core protein purification

TB (100 µg/mL carbenicillin, 34 µg/mL chloramphenicol) was inoculated with overnight starter cultures of BL21-CodonPlus(DE3)-RIL-pET22b(+)-CHIKV and incubated at 37˚C, 220 rpm. Once OD_{600} reached ~2.5 the flask was heat shocked to 42˚C for 30 minutes and cooled to 28˚C for 20 minutes. Cells were induced with 1mM IPTG and incubated at 37˚C, 180 rpm for a further 6 hours. The contents of the flask were centrifuged to pellet the *E.coli* expressing

CHIKV core protein and resuspended in alphavirus native buffer. The resuspended pellet was incubated with 1 mg of lysozyme (Sigma) and 1 EDTA-Free protease inhibitor cocktail tablet (Thermo Fisher) for 30 minutes prior to sonication. The cell lysate was then incubated with 1 mg of DNase I (Sigma) at 4˚C for 1 hour, and then centrifuged to remove insoluble components. When samples from the stages of the trial purification were analysed by SDS-PAGE however, it appears that the CHIKV core protein remains in the insoluble fraction and is pelleted out with the cellular debris after centrifugation (Figure 6.4). Core protein is visible in large quantities from the earliest stages of purification however this amount is significantly decreased in the supernatant sample and as such measures to solubilise the protein were undertaken.

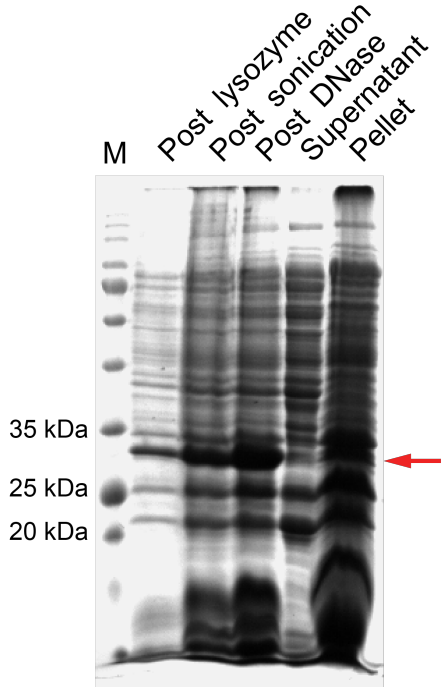


Figure 6.4: CHIKV core protein purification. SDS-PAGE of the stages in CHIKV Cp purification. Induced protein is visible in all stages of purification (red arrow) however, after ultracentrifugation the induced protein remains in the pellet indicating that it is insoluble.

6.2.3.2 CHIKV core protein solubilisation

The Astbury Centre at the University of Leeds includes a Molecular Interactions Centre for investigating molecular interactions including the aggregation of proteins. To probe the potential solubility of the CHIKV core protein a solubility assay was devised using a modified Optim unfolding and aggregation screening library (Table 6.3). CHIKV core protein was purified and pelleted out of solution via ultracentrifugation. This pellet was then resuspended in alphavirus native buffer and dispensed into 180 µL aliquots. Each component (20 µL) was added to the appropriate tube, mixed gently and incubated at room temperature for 30 minutes. The samples were then centrifuged for 10 minutes at high speed to separate the soluble and insoluble components and the supernatant from each sample was analysed by SDS-PAGE. Some of the conditions used in the screen were successful in solubilising the CHIKV core protein as seen in Figure 6.5 by the presence of a band in lanes 5 and 17, and to a lesser extent, lanes 2 and 16. The samples incubated with 5% sarkocyl and 1 M magnesium chloride were the most successful in solubilising the CHIKV core protein. Large bands at ~30 kDa are observed in both lanes when compared to water which acted as the negative control (lane 1). As sarkocyl brought many proteins into solution, $MgCl_2$ was used in future experiments to solubilise CHIKV Cp.

Table 6.3: Screening components used for CHIKV Cp solubilisation.

#		#	
1	Water (-ve)	10	50 mM EDTA
2	5% SDS	11	100 mM DTT
3	5% CTAB	12	1 M Na_2CO_3
4	5% Deoxycholate	13	1 M Na_2HPO_4
5	5% Sarkocyl	14	2 M Na_2HPO_4
6	10% Tween 20	15	1 M Na_2SO_4
7	10% Triton X-100	16	4 M NaCl
8	Igepal	17	1 M $MgCl_2$
9	8 M Urea		

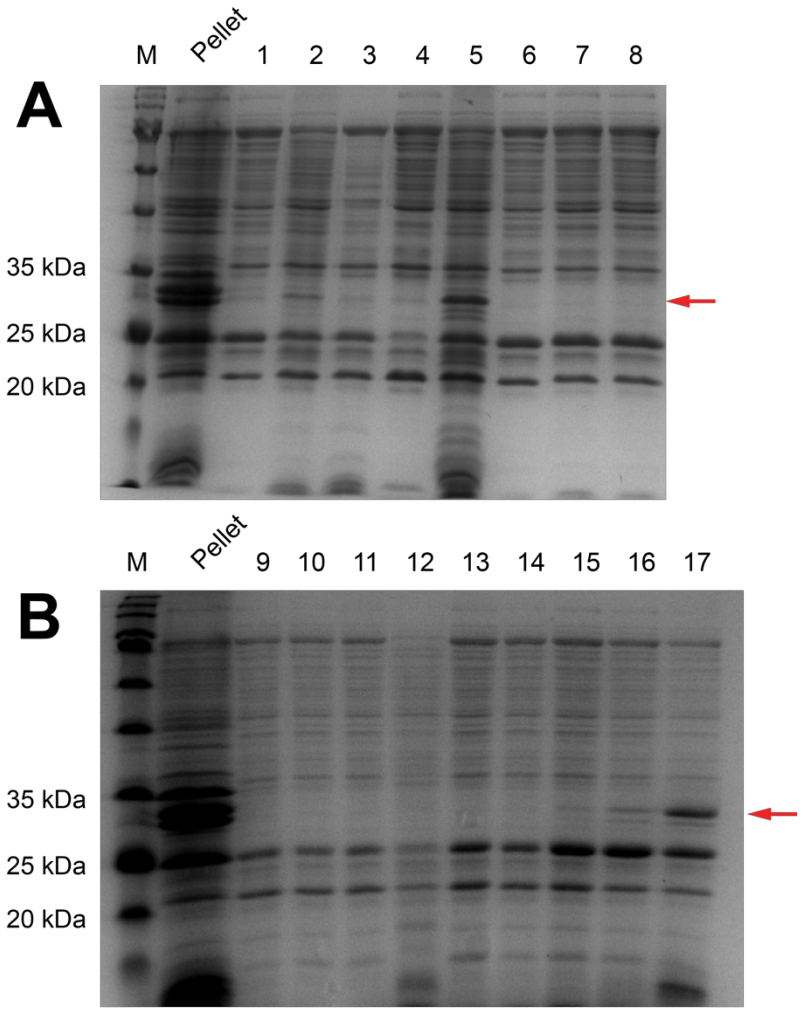


Figure 6.5: CHIKV core protein solubility screening. (A) SDS-PAGE of CHIKV Cp pellets after incubation with components 1-8 from Table 6.3. **(B)** SDS-PAGE of CHIKV Cp pellets after incubation with components 9-17 from Table 6.3. Protein brought back into solution is seen at ~30 kDa and indicated with the red arrow. Pellet alone is run for comparison. M = BLUeye protein ladder (Geneflow).

6.2.3.3 Magnesium chloride solubilisation gradient

To determine the amount of $MgCl_2$ needed to bring the CHIKV core protein out of solution the solubilisation assay was repeated with differing concentrations of $MgCl_2$. As shown in Figure 6.6, 25-100 mM, and to some extent 12.5 mM were able to bring the protein into the supernatant. 50 mM $MgCl_2$ was incorporated into a solubilisation buffer used to bring the protein out of the pellet after ultracentrifugation.

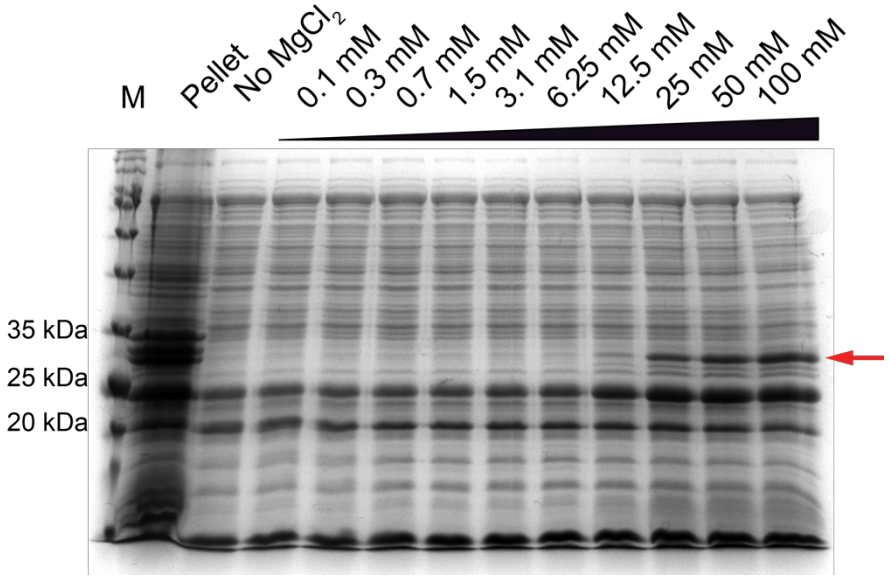


Figure 6.6: The effect of $MgCl_2$ concentration on the solubilisation of CHIKV core protein. CHIKV Cp pellets were incubated with increasing concentrations of $MgCl_2$, centrifuged and the resulting supernatants were analysed by SDS-PAGE. CHIKV core protein is seen at ~ 30 kDa and is indicated by the red arrow. Magnesium concentrations of above 25 mM are able to bring the core protein into solution, with some effect also seen at 12.5 mM. M = BLUeye protein ladder (Geneflow).

6.2.3.4 SFV and RRV core protein purification

LB (100 µg/mL carbenicillin, 34 µg/mL chloramphenicol) was inoculated with an overnight starter culture of Rosetta-pET22b(+)SFV or Rosetta-pET22b(+)RRV and incubated at 37˚C, 220 rpm. Once OD_{600} reached ~0.6 the flask was heat shocked and cooled before induction with 1mM IPTG and incubated at 28˚C, 180 rpm for 6 h. The contents of the flask were centrifuged

to pellet the *E.coli* expressing core protein and resuspended in alphavirus native buffer. The resuspended pellet was incubated with 1 mg of lysozyme (Sigma) and 1 EDTA-Free protease inhibitor cocktail tablet (Thermo Fisher), for 30 minutes, prior to sonication. The cell lysate was then incubated with 1 mg of DNase I (Sigma) at 4˚C for 1 hour and ultracentrifuged to remove insoluble components. Samples from purification stages were analysed by SDS-PAGE (Figure 6.7) which indicates that like the CHIKV core protein, the SFV and RRV core protein remains in the insoluble fraction. Solubilisation buffer devised for CHIKV Cp purification was therefore used in the purification of SFV and RRV Cps.

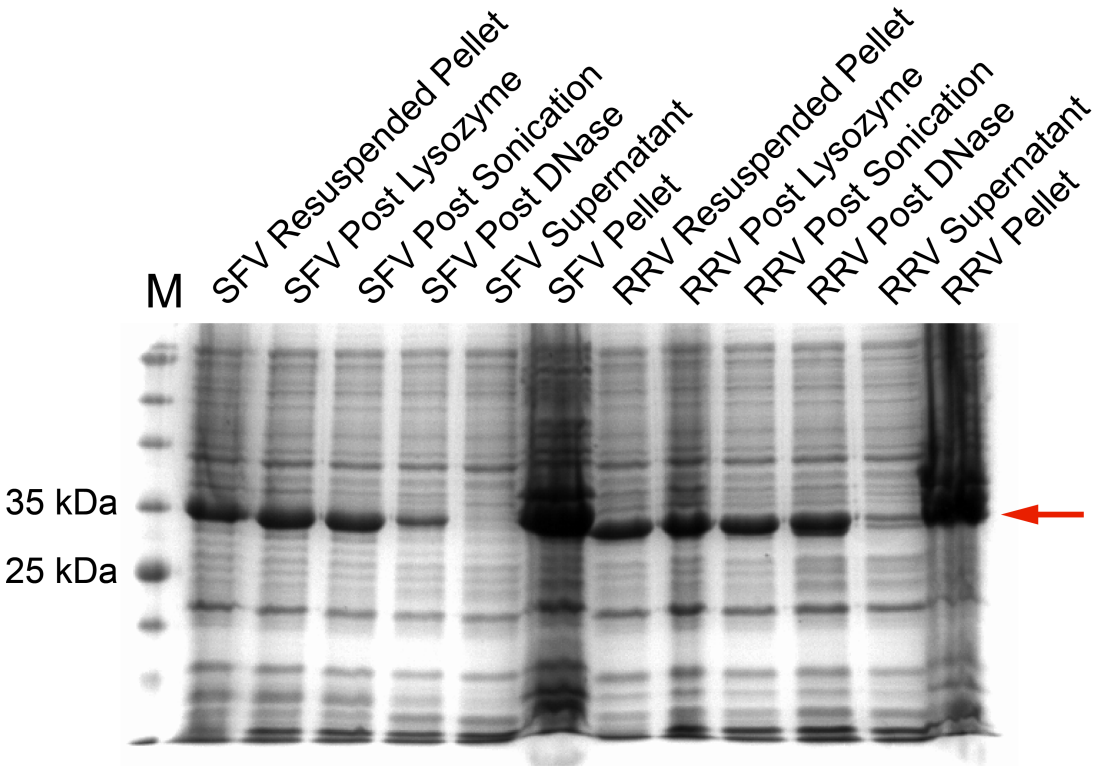


Figure 6.7: SFV and RRV core protein purification. SDS-PAGE of the stages in SFV and RRV Cp purification. Induced protein is visible in all stages of purification however, like CHIKV Cp the induced protein remains in the pellet after ultracentrifugation. M = BLUeye protein ladder (Geneflow).

6.2.3.5 Purification of core proteins by ion exchange chromatography

After ultracentrifugation, cell pellets were resuspended in solubilisation buffer (SB) and rolled at 4˚C for one hour. The sample was centrifuged at 5000 x *g* for 10 minutes to separate soluble and insoluble components before loading on to a pre-equilibrated 5 mL SP-FF column using the sample pump. Flow through was collected in 50 mL fractions and the column washed with 25 mL of SB before bound protein was eluted with a gradient of alphavirus buffer B. 2 mL Fractions were collected and those with absorbance at 260 and 280 nm were analysed by SDS-PAGE. This method is common to CHIKV, SFV and RRV Cp purification and shown in Figures 6.8-6.10.

SDS-PAGE of CHIKV Cp ion exchange shows that more than one protein is eluted from the column, suggesting that they have a similar isoelectric point. CHIKV Cp is visible at approximately 30 kDa consistent with images from throughout the purification process. This band is however, accompanied by other bands including a large one at ~25 kDa (Figure 6.8B). SFV and RRV Cp ion exchange fractions show that again more than one protein is eluted from the column (Figures 6.9B and 6.10B). RRV Cp and SFV Cp are visible at ~30 kDa, as seen during the purification process. However, unlike CHIKV Cp, both SFV and RRV Cp are in separate fractions from the co-eluting proteins and can therefore be separated by ion exchange chromatography.

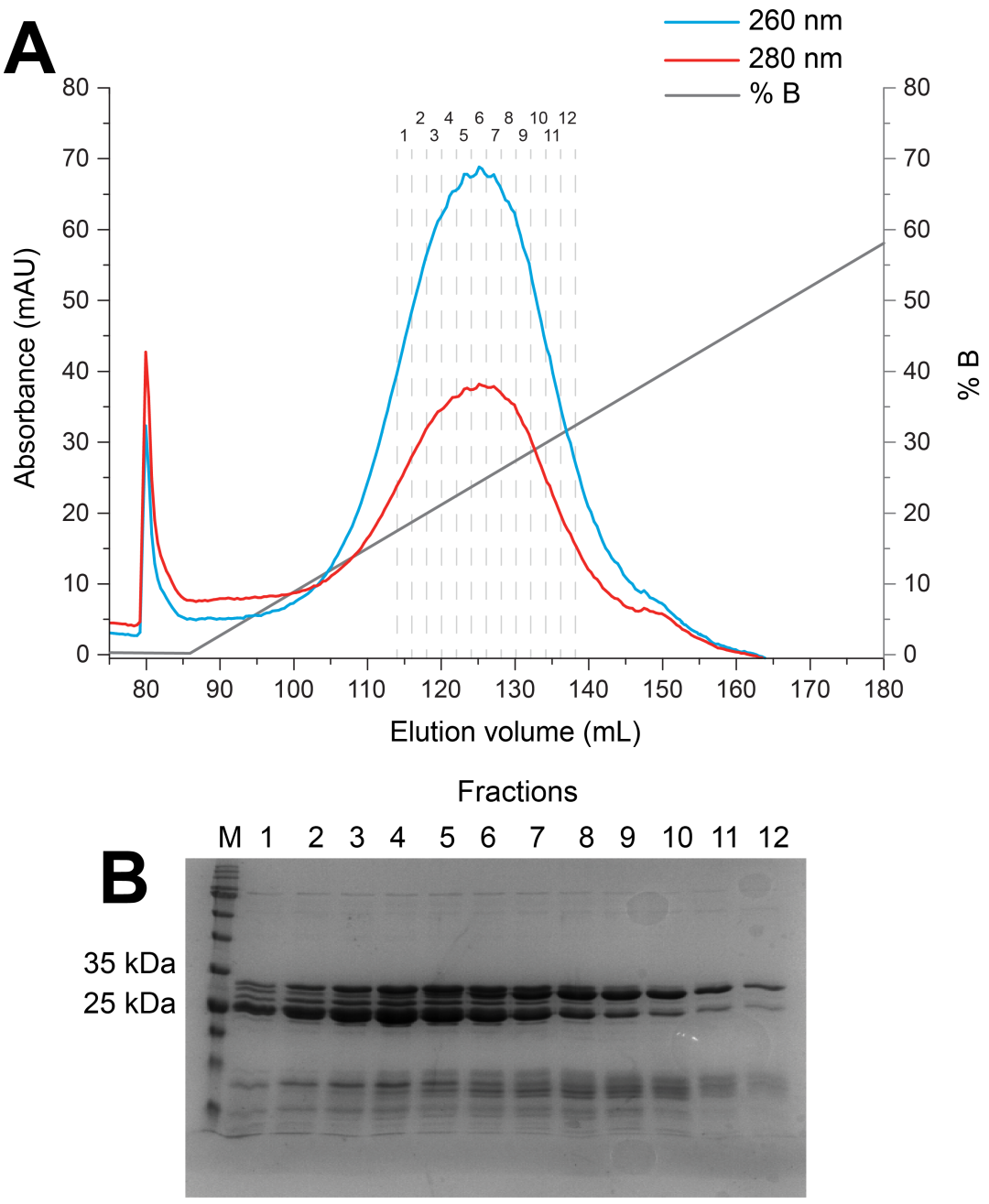


Figure 6.8: Purification of CHIKV core protein. (A) Chromatogram showing absorbance at 260 and 280 nm of of Alphavirus B gradient. **(B)** SDS-PAGE of fractions with absorbance at 260 and 280 nm. M = BLUeye protein ladder (Geneflow).

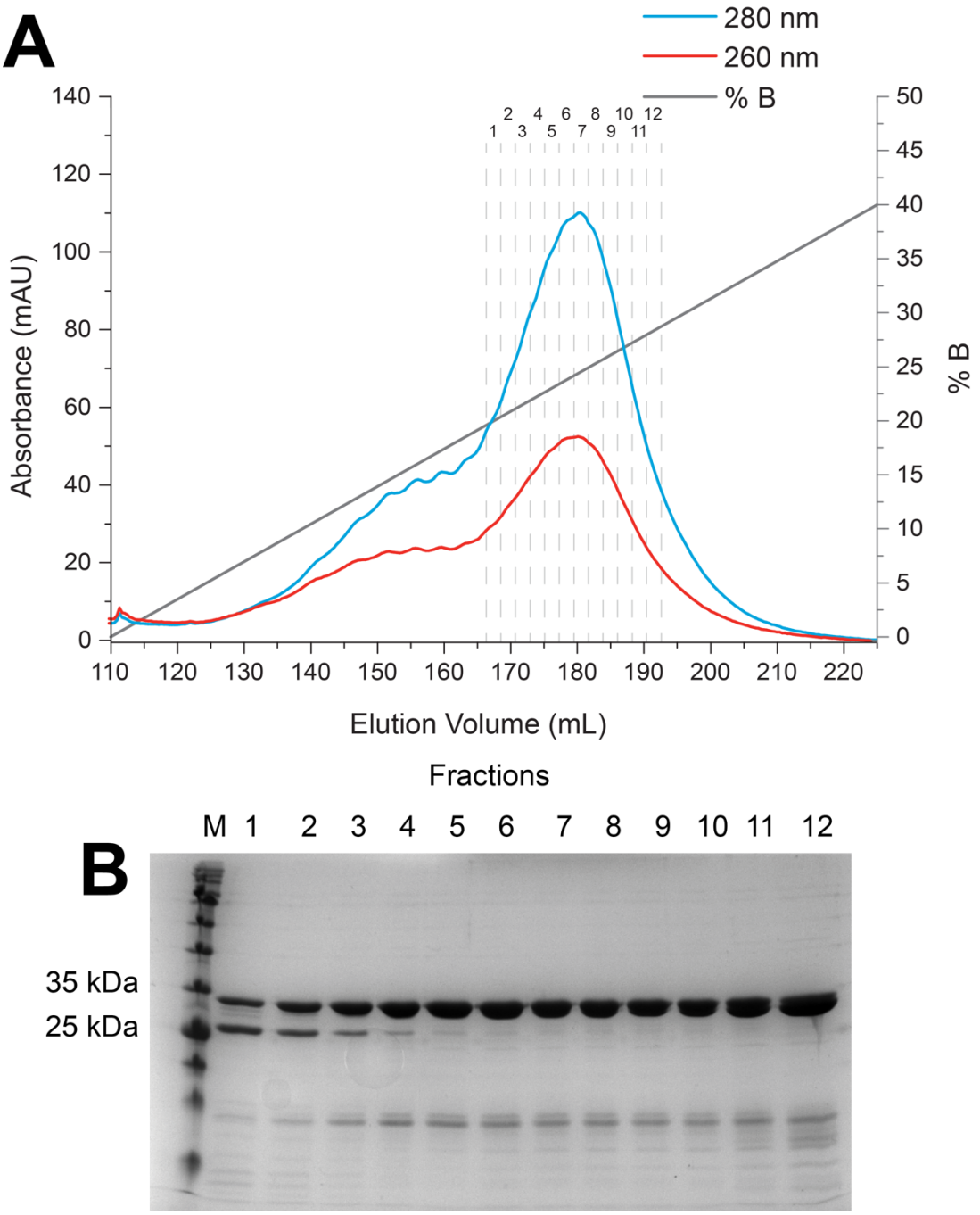


Figure 6.9: Purification of SFV core protein. (A) Chromatogram showing absorbance at 260 and 280 nm of gradient of of Alphavirus B gradient. **(B)** SDS-PAGE of fractions with absorbance at 260 and 280 nm. M = BLUeue protein ladder (Geneflow).

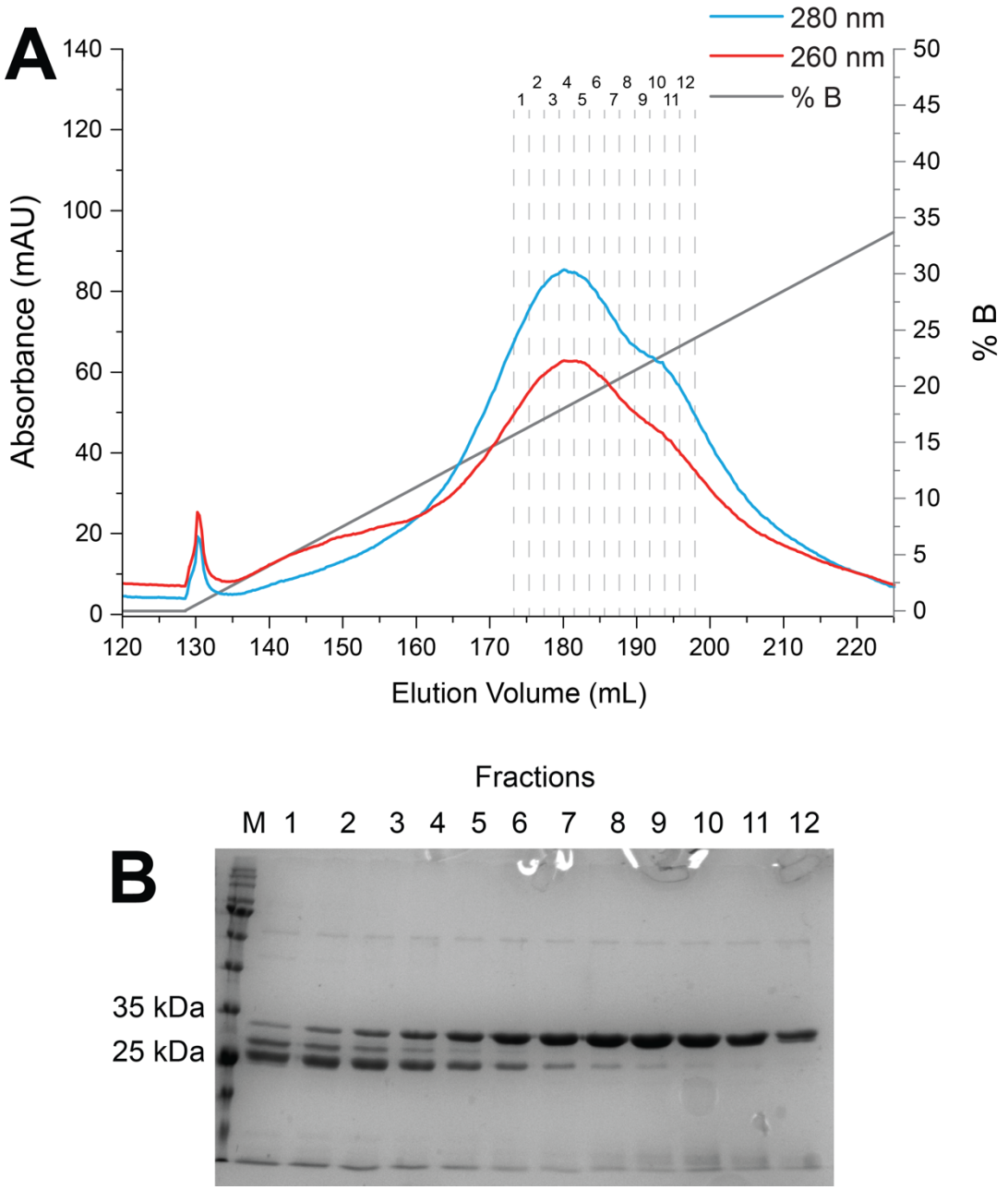


Figure 6.10: Purification of RRV core protein. (A) Chromatogram showing absorbance at 260 and 280 nm of Alphavirus B gradient. **(B)** SDS-PAGE of fractions with absorbance at 260 and 280 nm. M = BLUeye protein ladder (Geneflow).

6.2.3 Core protein characterisation

6.2.3.1 Mass spec analysis of proteins purified by ion exchange

As multiple proteins were purified during CHIKV Cp ion exchange, trypsin digest mass spectrometry was carried out to identify the co-eluting proteins. Dialysed protein was concentrated using Vivaspin20 centrifugal concentrators (GE Healthcare), buffer exchanged into ammonium acetate and lyophilized. The sample was then analysed by in house trypsin digest mass spectrometry at the University of Leeds. MS/MS data was processed with PEAKS studio (Bioinformatic Solutions Inc, Waterloo, Ontario, Canada) and searched against the *E.coli* entry in the Uniprot database (release 2016_12) and all Swissprot CHIKV sequences. The quality of a peptide spectrum match is given a linear discriminative function (LDF) score by PEAKS software. This score is based on the match between peaks in the spectrum and fragment ions, in addition to similarity to the database peptides and the *de novo* sequencing peptide. For ease of human interpretation, the LDF score is converted to a P-value and subsequently to -10lgP' (-10*log(P-value)). A higher -10lgP value indicates a more significant match, a -10lgP of 20 = 1% P-value. Figure 6.11 shows the peptide spectrum score distribution of the PEAKS search for trypsin digest mass spectrometry of the CHIKV and co-eluate sample. For this data set a threshold of ~24 has been set, any score greater than this is considered of relatively high confidence due to the low presence of decoy matches beyond this point.

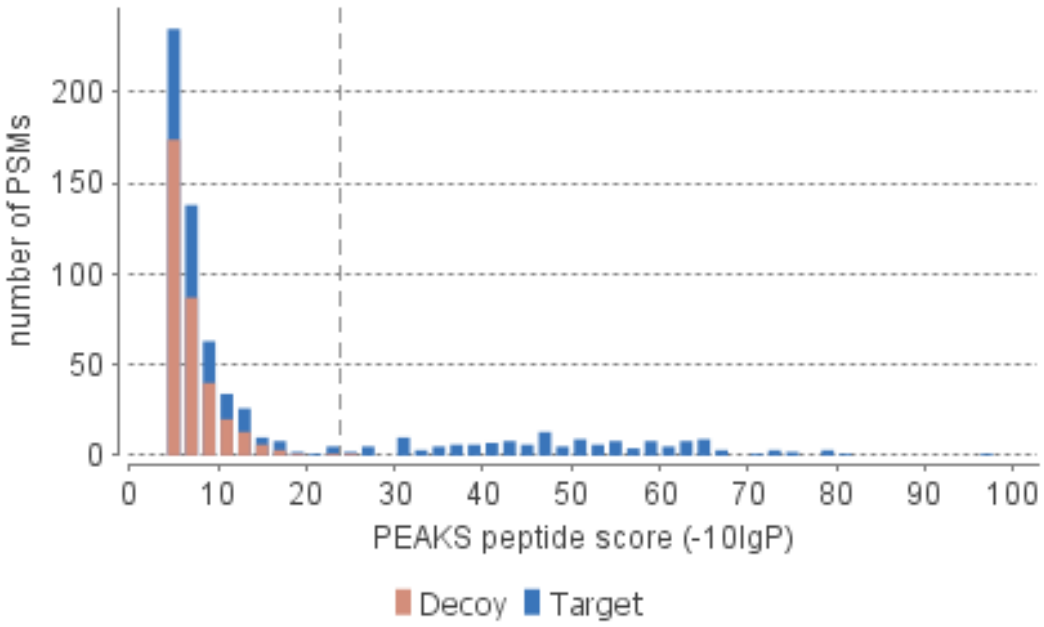


Figure 6.11: PEAKS DB search result for CHIKV Cp and co-eluate sample. Screenshot of the PEAKS DB result for the CHIKV and co-eluate sample submitted for trypsin digest mass spectrometry. PEAKS peptide score, which indicates the significance of the match between the peptide and the database, is shown on the X axis and the number of peptide spectrum matches on the Y.

105 matches were made to proteins in this bespoke database with varying -10lgP values from 300.45 to 24.12 (Appendix E). The identified proteins can be grouped into 11 distinct categories with multiple 'hits' for each. The proteins identified with the highest confidence, match to the structural polypeptide of chikungunya virus, more specifically the core protein coding region. The other hits from trypsin digest mass spec are all proteins carried over from expression in *E.coli* including 50S ribosomal proteins L16, L18, L35 and 30S ribosomal proteins S4 and S14. The co-elution of alphavirus core proteins and ribosomal subunits is not surprising as previous experiments have indicated that there is a ribosome binding site in the alphavirus core protein and this activity is linked to viral core disassembly (Würkner and Wengler, 1992).

The ProtParam tool from ExPASy allows for the calculation of a theoretical isoelectric point (pI) from a given amino acid sequence. The theoretical pI of the CHIKV core protein is 10.29. Table 6.4 shows the theoretical pIs of the identified proteins as predicted by ExPasy protparam (Gasteiger *et al*., 2005). As the proteins co-elute, this indicates that they have similar pIs, this allows some of the potential candidates to be discarded. Both the S4 and L18 have pIs close to that of CHIKV core protein, 10.05 and 10.42 respectively. In addition, mass determination mass spec carried out without solubilising the CHIKV protein, gives the co-eluting protein a mass of 23415.62 Da (Figure 6.12). These results taken in combination suggest that the co-eluting protein is the S4 subunit of a 30S ribosome.

Table 6.4: Calculated masses of identified species from mass spectrometry

Description	ProtParam pI	ProtParam Mass
50S, L16	11.22	15281.20
30S, S4	10.05	23469.09
DsbG	8.32	27494.99
50S, L18	10.42	12769.63
Acyl Carrier protein	3.98	8639.52
Pyruvate dehydrogenase E1	5.46	99668.49
30S, S14	11.16	11580.48
50S, L35	11.78	7728.93

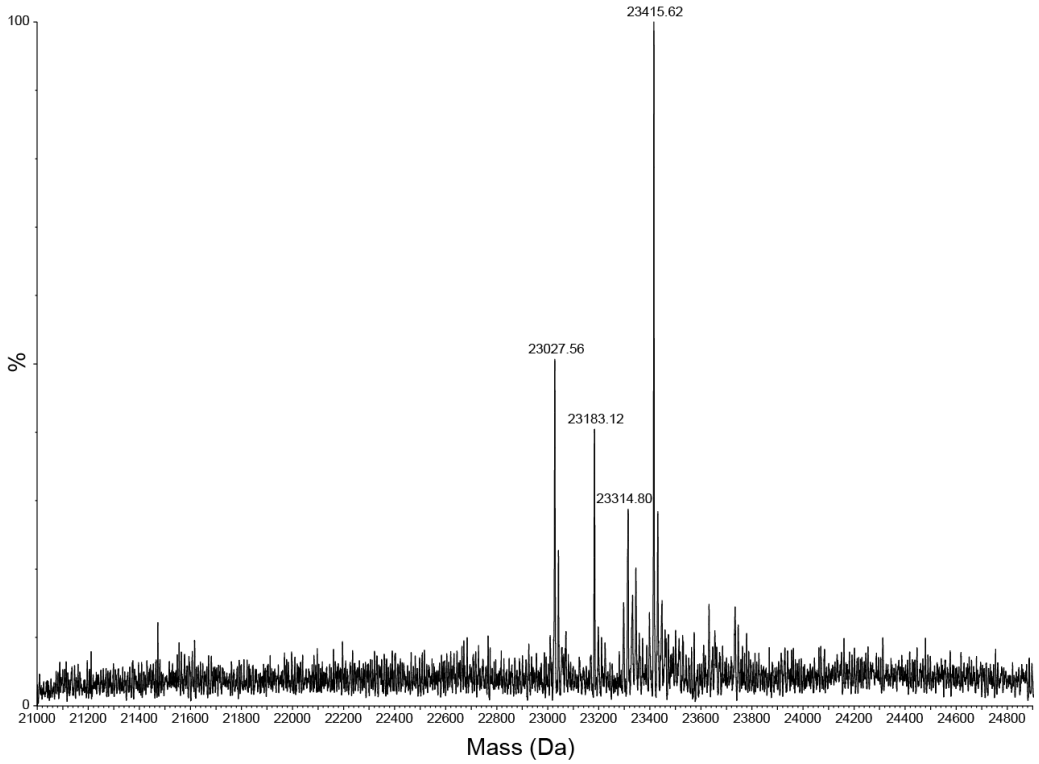


Figure 6.12: Mass spectrometry trace of CHIKV Cp co-eluate. Mass spec of the protein which co-eultes with CHIKV core protein. The most abundant mass is seen at 23415.62, with all detected masses between 23027.56 and 23415.62 Da. In conjuction with the results from trypsin digest mass spec (Appendix E), the observed masses indicate that 30S, S4 is co-eluting with CHIKV core protein.

6.2.3.2 Chemical cleavage of S4 with hydroxylamine

As both CHIKV core protein and the S4 subunit have similar isoelectric points, 10.29 and 10.05 respectively, they are difficult to separate by means of charge and being similar in size ~30 kDa and ~23 kDa they are also difficult to separate by size exclusion. ExPASy Peptide cutter was used to find an enzyme that would discriminate between the two proteins and only cleave the S4 subunit (Gasteiger *et al*., 2005). A search of the database revealed that hydroxylamine, Factor Xa, Caspase 1 and Caspase 4 cleave the 30S S4 sequence but not the CHIKV Cp sequence. As hydroxylamine cleavage of the S4 subunit has already been characterised (Changchien and Craven, 1986) this method of chemical cleavage was chosen.

Chemical hydroxylamine cleavage occurs between asparagine-glycine bonds of which the 30S S4 subunit has one at positions 124-125 (Changchien and Craven, 1986) by the mechanism shown in Figure 6.13 Cleavage of this bond should result in fragments which are 126 and 80 residues in length and therefore could be separated from CHIKV core proteins on the basis of size. Cleavage with hydroxylamine requires incubation at 45°C for 4 hours in lysis buffer containing hydroxylamine and guanidine-HCL, before quenching with 2% (v/v) TFA. The sample was desalted using a NAP-25 column and fractions analysed by SDS-PAGE (Figure 6.14). Although some co-eluate cleavage seems to have occurred within the sample there still remains a significant amount after a 4-hour incubation.

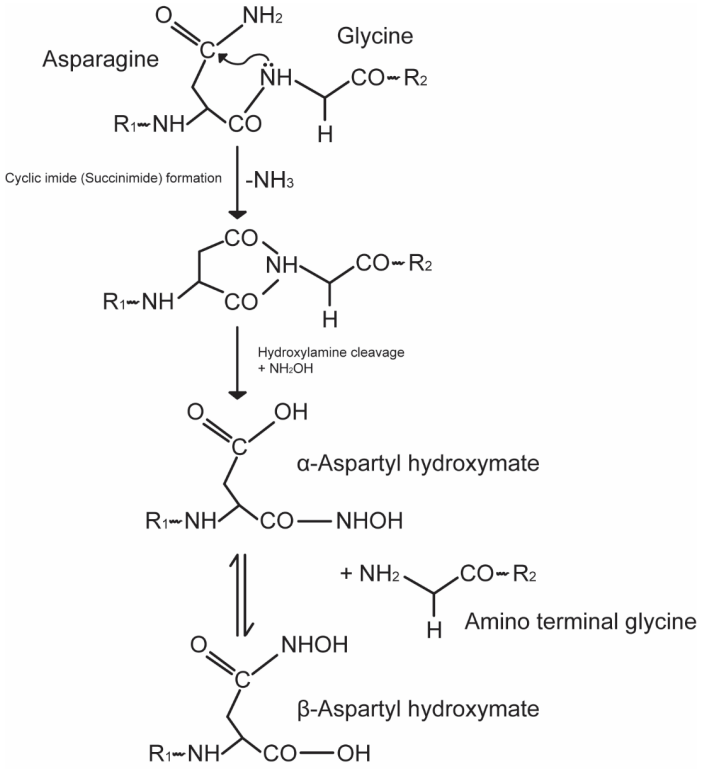


Figure 6.13: Mechanism of hydroxylamine cleavage. Chemical hydroxylamine cleavage occurs between asparagine-glycine bonds. The 30S S4 subunit has one at positions 124-125 and cleavage at this position will result in 2 fragments of 126 and 80 residues in length (Changchien and Craven 1986).

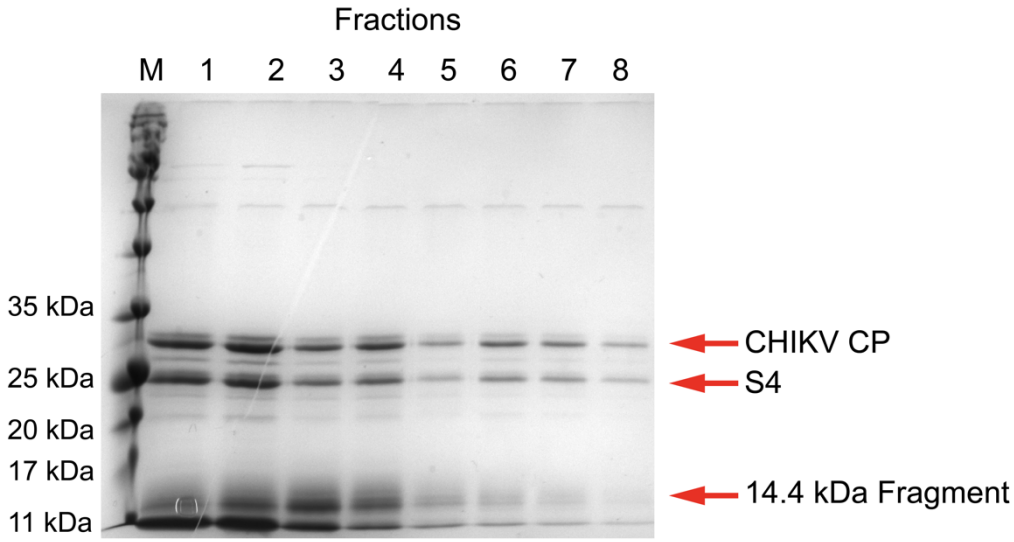


Figure 6.14: S4 chemical cleavage. SDS-PAGE of fractions from S4 chemical cleavage. Cleavage occurs between the asparagine-glycine bond and produces fragments of approximately 14.4 kDa and 9.82 kDa.

6.2.3.3 Mass determination of SFV and RRV core proteins

Purified SFV and RRV core proteins were concentrated using a centricon with a 10k MWCO (Merck), dialysed into 50 mM ammonium acetate and submitted for mass determination mass spectrometry at the University of Leeds mass spectrometry facility. Mass spec results indicate that the purified SFV core protein has a mass of 29775.32 Da (Figure 6.15A) which is consistent with the calculated mass of 29777.51 Da. The purified protein is also stable at 4˚C for at least 96 hours. Mass spec results show the purified RRV core protein has a mass of 30799.41 Da (Figure 6.15B) which is consistent with the calculated mass of 30521.40 Da (Gasteiger *et al*., 2005). However, as previous mass determination attempts had shown cleavage of the protein, investigations into the stability of RRV Cp were carried out.

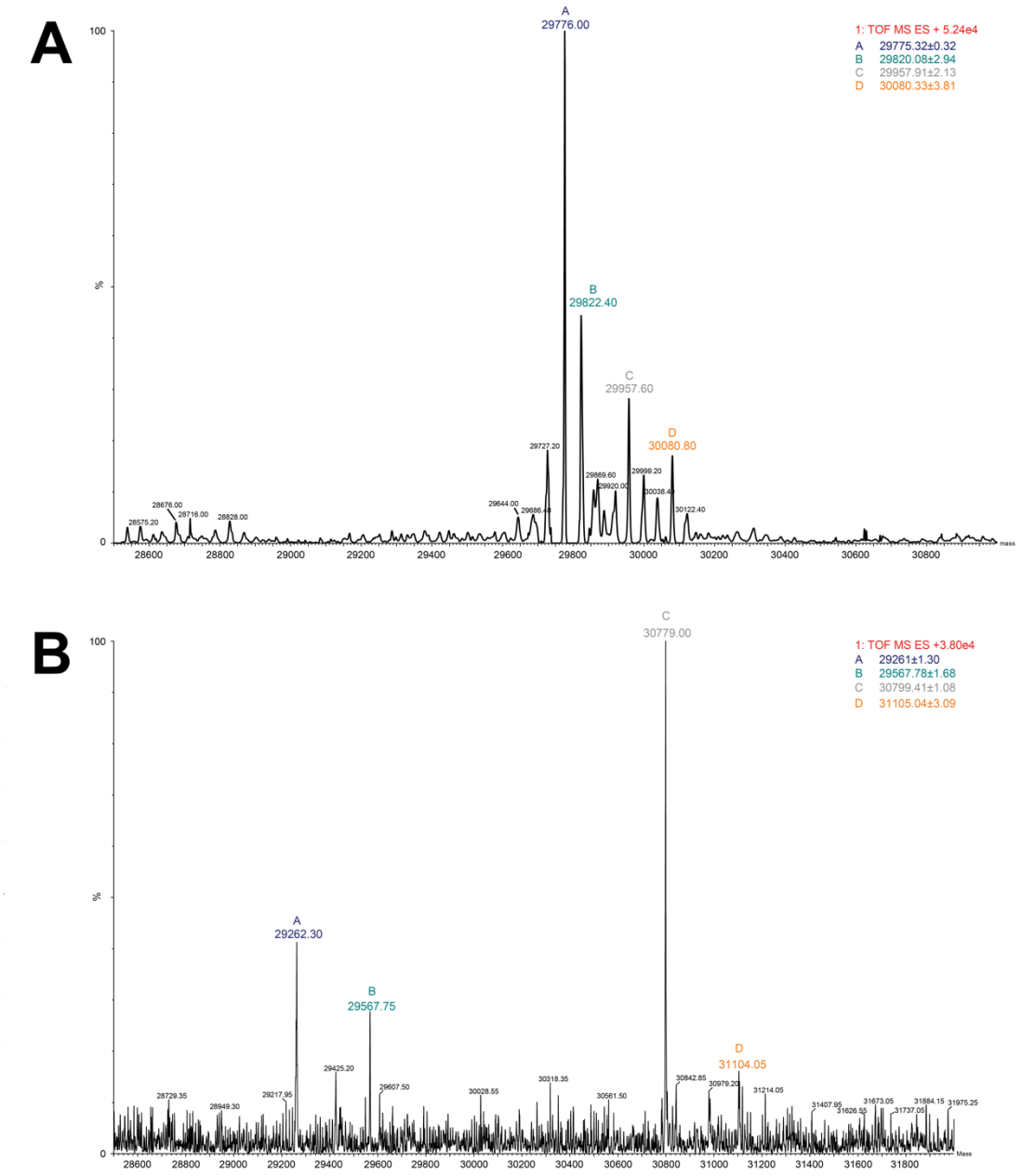


Figure 6.15: Mass spectrometry of SFV and RRV core proteins. (A) Mass spec of SFV core protein after storage at 4°C for 96 hours. The most abundant mass is seen at 29775.32 Da which is consistent with the calculated mass of 29777.51 Da. **(B)** Mass determination Mass spec of RRV core protein. The most abundant mass is seen at 30799.41 Da which is consistent with the calculated mass of 30521.40 Da.

Mass spectrometry (MS) results indicate that the RRV core protein undergoes proteolytic cleavage readily after purification. Unlike SFV core protein, which after 4 days at 4°C remains at the expected mass, RRV Cp has undergone proteolytic cleavage in the same time frame. Freshly purified RRV protein was analysed by mass determination mass spectrometry and was found to be comparable to the expected size with a mass of 30799.41 Da (Figure 6.16A). This is also true at 24 hours post purification after storage at 4°C (30797.87 Da). However, after 96 hours at 4°C the protein has cleaved and products at 23876.79, 24519.05, 24824.38 and 26130.40 kDa are present in the sample (Figure 6.16C).

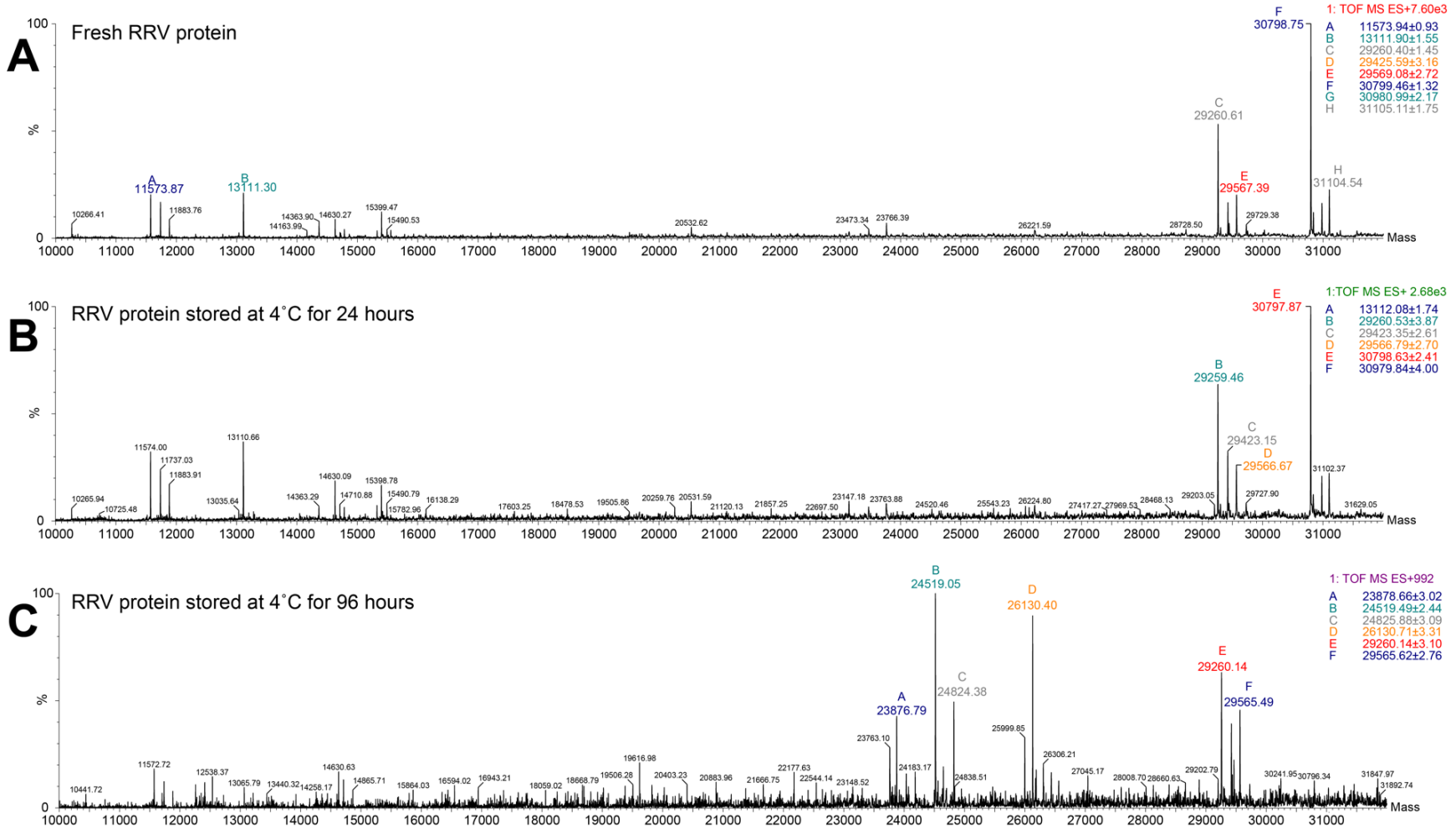


Figure 6.16: Mass spec of RRV Cp degradation. Mass determination Mass spec of RRV core protein **(A)** freshly purified, **(B)** stored at 4°C for 24 hours and **(C)** stored at 4°C for 96 hours. Calculated mass of RRV core protein is 30521.40 Da.

6.2.3.4 Sedimentation Velocity Analytical Ultracentrifugation (svAUC) of SFV core protein

Purified SFV core protein was buffer exchanged into alphavirus native buffer and analysed by svAUC to determine the state of the protein in solution. As explained in Chapter 2, svAUC allows for the determination of the sedimentation coefficient of a sample which can give insights into protein conformation and quaternary structure. The sample was placed into a 2-sector cell and centrifuged at 655,200 x *g* while scanning at 280 nm. The data was analysed using Sedfit and is shown in Figure 6.17. The calculated molecular mass is ~28.1 kDa which is consistent with the protein being a monomer in solution at 10 µM concentration in the absence of nucleic acid.

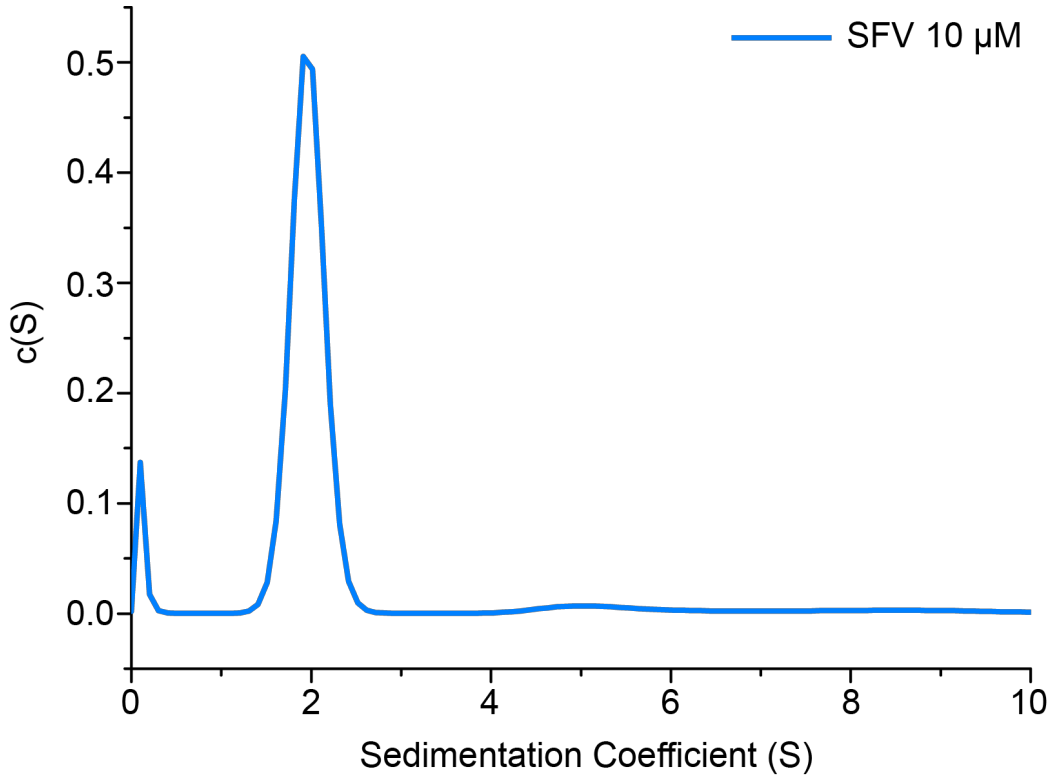


Figure 6.17: Sedimentation properties of recombinant SFV core protein. Analysis of 10 µM purified SFV protein by svAUC indicates that at this concentration the protein is monomeric with a calculated mass of 28.1 kDa. c(S) is a continuous distribution from the Lamm equation model (Schuck *et al*. 2015).

6.2.3.5 Circular Dichroism (CD) of recombinant SFV core protein

To ensure that the recombinant core protein had a similar structure to the wild type viral core protein it was analysed by CD. Purified SFV Cp was buffer exchanged into phosphate buffer and diluted to 0.02 mg/mL for analysis. CD was carried out using a Chirascan from 340-180 nm, at 22°C. Molar ellipticity of SFV Cp was calculated using experimental data (Appendix E) as outlined in materials and methods. The SFV Cp CD spectrum is most similar to that of a disordered/irregular structure with a drop in molar ellipticity at ~200 nm (Figure 6.18) (Kelly *et al*., 2005). The data was analysed using Dichroweb (Whitmore and Wallace, 2004) and CDSSTR reference set 4 (Sreerama and Woody, 2000).

Analysis with the CDSSTR reference set assigns secondary structure split into 6 categories where both α-helix and β-sheet are divided into regular ($_R$) and distorted ($_D$). Dichroweb calculated secondary structure fractions are shown in Table 6.5. Deconvolution of this data indicates a structural content of 8% α-helix, 34% β-sheet, 23% turns and 34% random coil. The solved crystal structure for SFV Cp is shown in Figure 6.19 along with the sequence chain view from PDB (PDBID: 1VCP, (Choi *et al*., 1997)). The protein consists of two β-barrel subdomains (1 and 2) comprised of six and eight β-strands respectively connected by a 10-residue linker. The β-barrel domains are organised in a 'Greek key' fold as seen in other chymotrypsin-like serine proteases (Sharma *et al*., 2016).

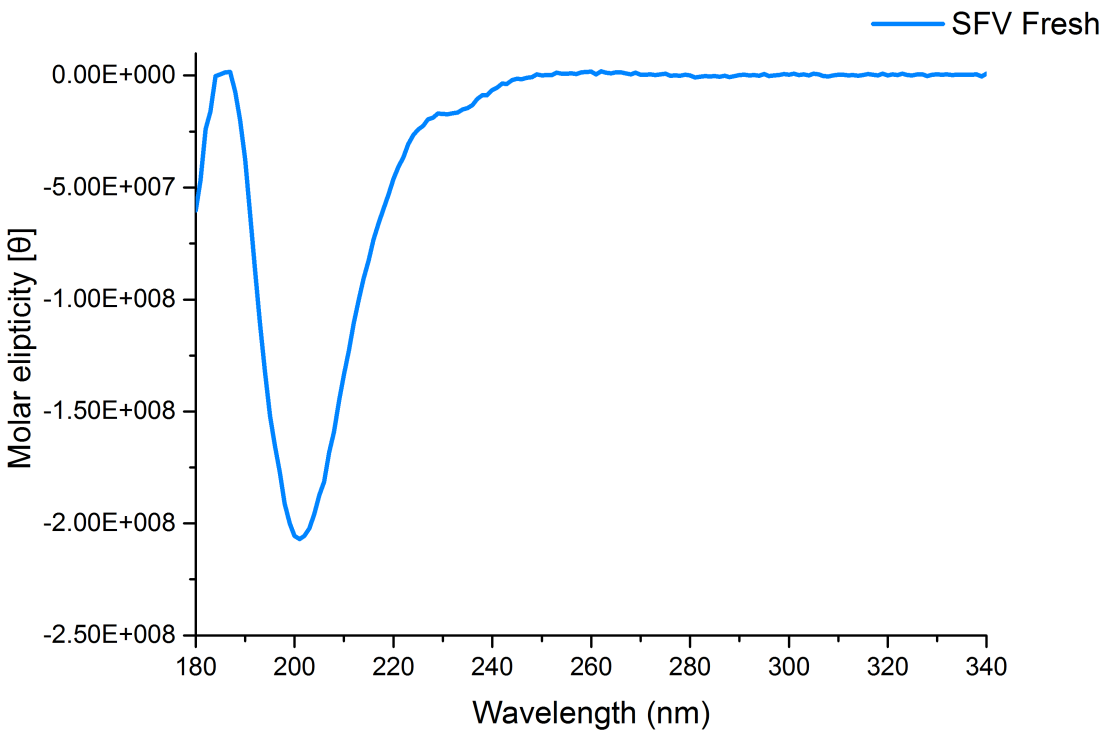


Figure 6.18: CD spectra results for SFV Cp. CD spectrum of 0.02 mg/mL purified SFV protein in 10 mM phosphate buffer (pH7.5) plotted as molar elipticity. Spectra were recorded using a Chirascan CD spectrometer at wavelengths between 180-340 nm with 1 nm steps and a buffer blank collected.

Table 6.5: Dichroweb secondary structure fractions for SFV Cp

α_R	α_D	β_R	β_D	Turns	Unordered	Total
0.01	0.07	0.22	0.12	0.23	0.34	0.99

The initial 118 amino acids are not seen in the crystal structure 1VCP and are thought to be disordered or cleaved before crystallisation (Choi *et al*., 1997). Of the available secondary structural data (aa 118-267), 6.7% is α-helix, 49.6% is β-sheet, 16.1% turns and 27.5% unordered including (S bend). The differences in percentages could be due to the inclusion of the N-terminal in CD analysis or folding of the protein on assembly.

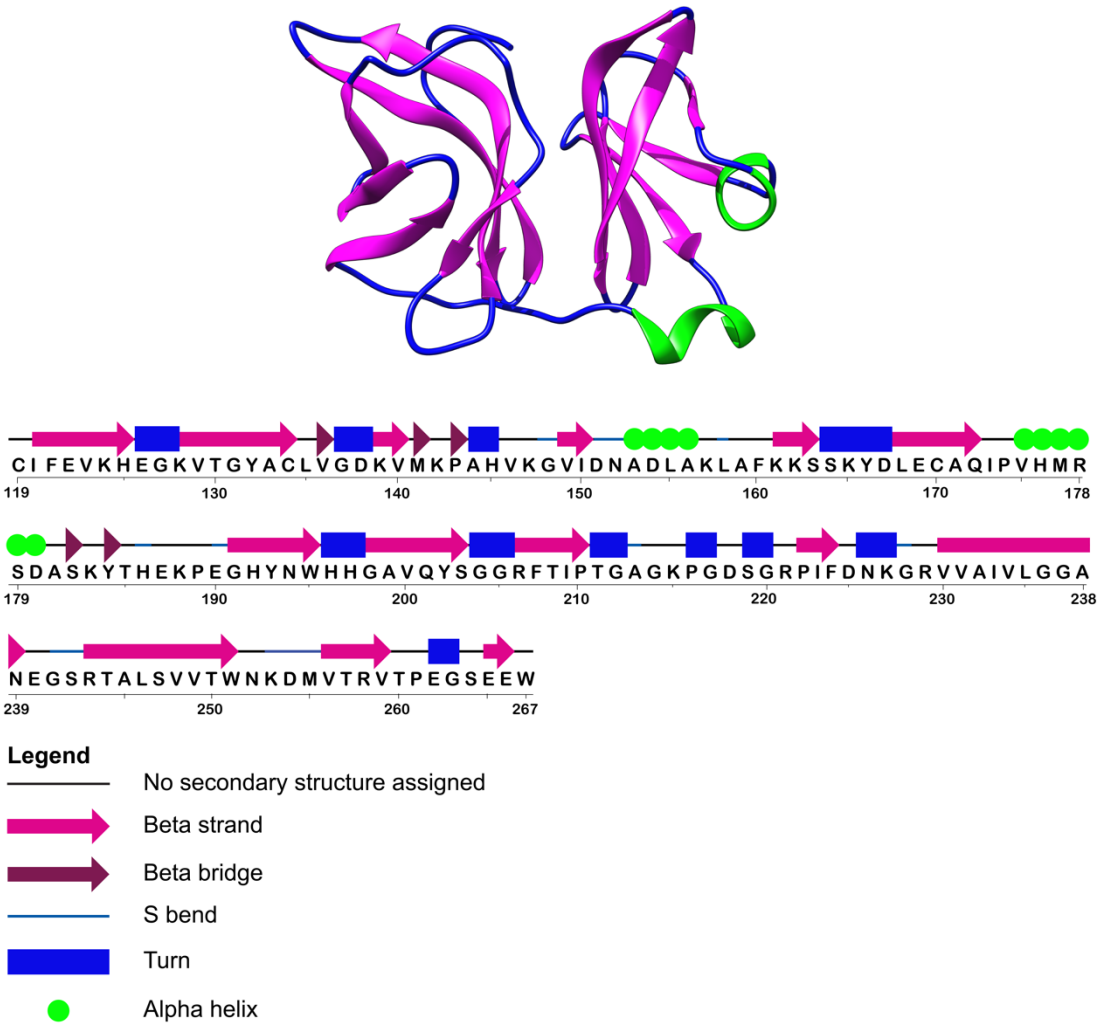


Figure 6.19: Crystal structure of SFV Cp. Structure of SFV Cp (PDBID: 1VCP) with secondary structure coloured (α-helix, green, β-sheet, magenta and unstructured, blue). Structure was produced using ChimeraX (Goddard *et al*., 2018). Sequence chain view was adapted from PDB. Both structure and chain view from (Choi *et al*., 1997)

6.2.4 Functional assays of recombinant alphavirus Cp

6.2.4.1 Assembly of core like particles with non-specific DNA

In 1999 Tellinghuisen *et al* investigated the assembly of alphavirus nucleocapsid cores using truncated SINV (aa 19-264) and RRV proteins. Core proteins were mixed with non-specific DNA in a roughly 1:1 molar ratio and incubated at 30°C for 30 minutes. Core like particles (CLPs) were

observed under negative stain TEM, suggesting that purified protein could encapsidate non-native DNA. Purified SFV and RRV core proteins from *E.coli* were used in a similar reassembly investigation to assess the assembly capacity with non-specific DNA.

Protein and DNA (DAR82, Materials and Methods) were incubated at 30°C separately for 30 minutes, before combining followed by a further 30-minute incubation at 30°C. At this point samples were removed for visualisation using negative stain TEM and the remaining mixture was incubated with high salt for an additional 30 minutes. As shown in Figure 6.20A and E, purified SFV and RRV core proteins incubated with non-specific DNA form CLPs of approximately 40 nm in diameter, consistent with 410 Å particles seen by Tellinghuisen *et al* (Tellinghuisen *et al*., 1999). Upon incubation with 1M NaCl (Figure 6.20C and G) CLPs are not visible with only small aggregates present in the sample, much smaller than the size expected for assembled particles. In the absence of DNA, no higher order structures are visible (Figure 6.20D and H). This suggests that assembly happens only in the presence of the nucleic acid as SFV Cp (10 μM), is monomeric when analysed by svAUC. This investigation however, was performed at high concentrations (μM range) which is not necessarily reflective of protein and nucleic concentrations of a *bona fide* assembly process. Assembly of nucleocapsid like particles was therefore subsequently investigated on a single molecule level.

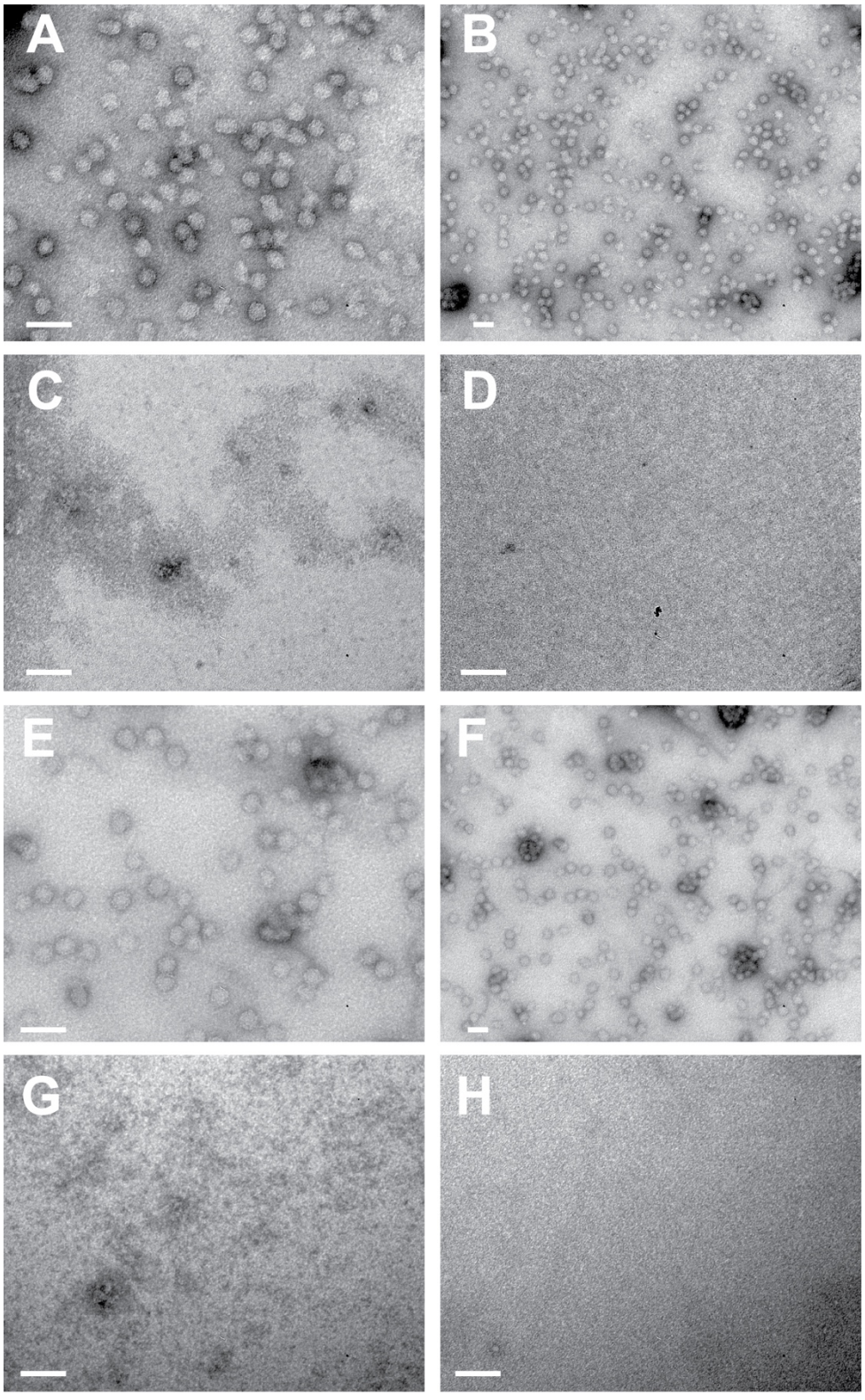


Figure 6.20: Reassembly of recombinant alphavirus core proteins with non-specific DNA. Negative stain TEM images of SFV Cp (A-D) and RRV Cp (E-H) reassemblies with non-specific DNA. (**A and E)** Cp reassembled with non-specific DNA at 30,000X magnification. **(B and F)** Cp reassembled with non-specific DNA at 12,000X magnification. **(C and G)** Cp reassembled with non-specific DNA incubated with 1M salt at 30,000X magnification. **(D and H)** Cp alone at 30,000X magnification. Markers = 100 nm.

6.2.4.2 Single molecule investigations of SFV reassembly

Assembly of alphavirus CLPs by both Wengler (Wengler *et al*., 1982) and Tellinghuisen (Tellinghuisen *et al*., 1999) indicate that nucleic acid is a crucial component of the nucleocapsid assembly mechanism. However, CLPs have been formed with both DNA and RNA suggesting that assembly is due to charge neutralization and there is no specificity (Mendes and Kuhn, 2018). As seen in the previous section a non-specific DNA fragment can induce the formation of particles using the Tellinghuisen reassembly method. However, when purified SFV core protein is titrated into 1 nM of the same fluorescently labelled, non-specific, DNA fragment (Appendix E) there is no increase in hydrodynamic radius suggestive of particle assembly (Figure 6.21). This result suggests that particle formation seen in the Tellinghuisen reassembly method is driven by high concentrations however, at single molecule concentrations (nM), perhaps more reflective of in vivo conditions, there is selectivity in packaging.

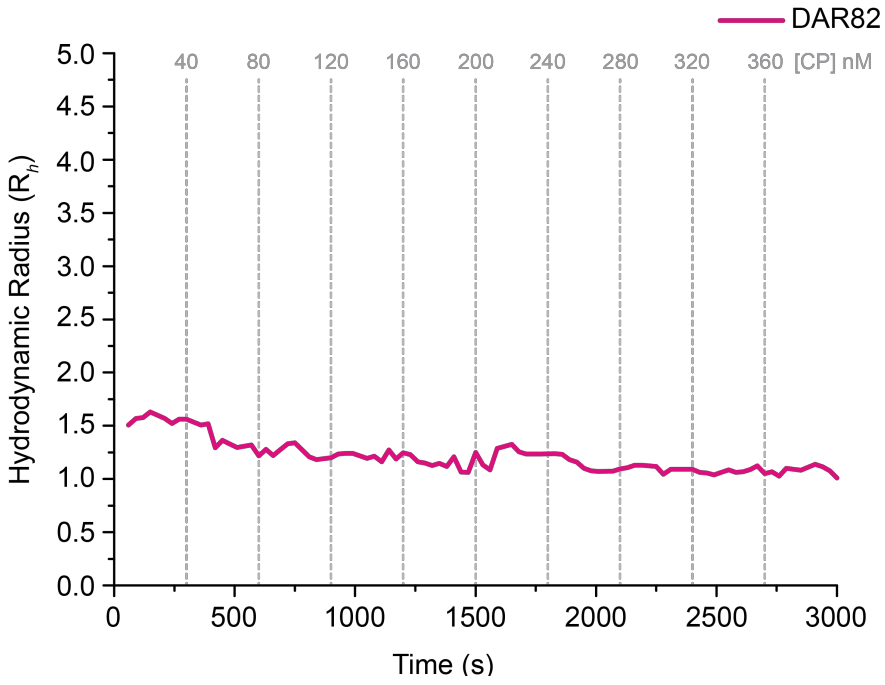


Figure 6.21: Single molecule reassembly of purified SFV with non-specific DNA. Purified SFV core protein was titrated into 1 nM Alexa Fluor 488 labelled DAR82 in 20 nM steps and the resulting R_h was monitored using smFCS.

6.3 Discussion

Previous investigations into alphavirus assembly mechanisms and packaging signals have exploited the use of bacterial protein production. Full length WEEV and RRV Cp, alongside truncated SINV Cp have been used in *in vitro* reassembly assays to produce CLPs with structures similar to those seen in native cores (Tellinghuisen *et al*., 1999; Mukhopadhyay *et al*., 2002). The aim of this chapter was the production of alphavirus core proteins for use in assembly investigations and packaging signal identification. Five alphavirus species were selected ranging from well-known human pathogens (CHIKV), potential bioweapons (WEEV) and prototypic alphaviruses (SINV and SFV) from the most populous antigenic serocomplexes.

The generation of alphavirus Cp gene containing plasmids was successful for all five chosen targets, although differences between individual proteins did not become apparent until expression trials were undertaken. Three of the five core proteins (CHIKV Cp, SFV Cp and RRV Cp) were all expressed, and bands were seen of the correct size after induction with IPTG. However, two of the five proteins, SINV Cp and WEEV Cp, did not express in either Rosetta or RIL cells at varying temperature and media conditions. Cell density (OD_{600}) of the samples in both cases does increase and reaches the stationary phase nonetheless, induction at this point did not result in protein expression. Upon transformation in Lemo-21 cells, which have tuneable expression due to the presence of T7 lysozyme under the control of a Rhamnose promoter, it became apparent that the proteins were potentially toxic. In the presence of IPTG, for both SINV Cp and WEEV Cp, no colonies were observed whereas

in the absence of IPTG colonies are present. This behaviour suggests that induction of protein production is the cause. Inclusion of rhamnose in the agar plates was able to rescue colony formation however, inclusion of rhamnose in the media was not sufficient to result in protein production.

Previous investigations have identified the presence of a ribosome binding site, on alphavirus core proteins thought to be involved in the nucleocapsid uncoating process which could explain toxicity in cells (Würkner and Wengler, 1992). If, when expressed, the proteins are binding to ribosomes this could interrupt their natural function and cause cell death. To further this, purification of CHIKV Cp using ion exchange chromatography resulted in multiple proteins eluting from the column at similar concentrations of buffer B. Trypsin digest mass spectrometry and mass determination mass spectrometry identified the co-eluate as the S4 subunit of a 30S ribosome. Co-elution is also seen in the purification of both SFV Cp and RRV Cp however, though protein peaks overlap they are distinguishable and therefore purifiable. Cleavage of the S4 subunit for the CHIKV samples was attempted however, due to the harsh, potentially denaturing, nature of the protocol (16.3 M Hydroxylamine, 7.5 M guanidine-HCL, incubation at 45°C for 4 hours and quenching with 2% (v/v) TFA) and the inefficiency of the cleavage, purification of CHIKV Cp was not attempted on a large scale from *E.coli*.

As previously mentioned purification of RRV Cp was successful, as the co-eluate was separated into different fractions from the protein of interest. The purified protein was reassembled with a non-specific DNA fragment, DAR82,

as per the method devised by Tellinghuisen (Tellinghuisen *et al*., 1999). Upon incubation with the DNA, particles of the correct size are seen, whereas there are no higher order structures in the Cp-only control suggesting that, as seen previously, the presence of nucleic acid is required for the formation of particles. It was observed however, that the RRV Cp readily cleaved when stored at 4°C for over 24 hours which suggests that more optimal buffer/storage conditions could be devised.

The most comprehensive characterisation of an alphavirus in this chapter was performed on SFV Cp. Like RRV Cp, SFV Cp was separated from the co-eluate during ion exchange chromatography and due to its relative stability at 4°C (>96 hours) more extensive investigations could be performed. Sedimentation velocity AUC of the protein at 10 μM indicates that it is monomeric at that concentration with an estimated mass of 28.1 kDa. Purified SFV Cp was also reassembled with the non-specific DNA DAR82, where particles of the expected size are seen after incubation. No particles are seen in the Cp only control which, combined with the svAUC results, suggest that formation of particles is dependent on the presence of nucleic acid. This reassembly however, is performed at concentrations not expected to be achieved *in vivo*. When reassembly is taken to the single molecule level, and purified SFV Cp titrated into 1 nM Alexa Fluor labelled DAR82 no change in hydrodynamic radius is detected. This suggests that, while high concentration can be used to induce particle formation without selectivity, at single molecule levels this process is more discriminating and complex.

Overall the expression of alphavirus core proteins in *E.coli* has yielded some interesting results. Reassembly of RRV and SFV core proteins at high concentration indicates that previous investigations into non-cognate packaging were correct however, the single molecule reassembly of SFV Cp highlights the non-native conditions used in previous assembly investigations. The ability to produce and purify two recombinant alphavirus core proteins paves the way for more in-depth analysis of the assembly process and the identification/confirmation of PSs in the alphavirus genus.

[This page has been left intentionally blank]

Chapter 7: Production of Alphavirus core proteins in SF9 insect cells

[This page has been left intentionally blank]

7.1 Introduction

Group I of the Baltimore classification system, double stranded DNA (dsDNA) viruses, contains among other human and animal viruses the family *Baculoviridae*. They have circular, supercoiled genomes within a capsid displaying helical symmetry and packaged within a lipoprotein envelope (King and Possee, 1992). The natural hosts for this family of viruses are insects and some crustaceans with over 600 identified isolates infecting over 100 species (Invitrogen, 2002; Chambers *et al*., 2018).

Baculoviruses used for expression purposes differ from wild type because they lack the polyhedrin gene which is expressed in the last stage of the viral life cycle. It is replaced by the gene that needs to be expressed. The polyhedrin protein forms polyhedral inclusion bodies (PIBs) which protect the viral particles as cell lysis progresses. This gene, along with another very late gene, p10, is not essential for viral particle production *in vitro*. As expression of these genes occurs after the production and maturation of virus particles, any cytotoxicity of the expressed protein should not affect replication. But, polyhedrin and p10 have strong promoters and so infected cells can produce large quantities of these two proteins (Invitrogen, 2002).

Baculoviruses were first used for the production of proteins by Smith *et al* (1983) and Pennock *et al* (1984). Both these groups used Autographa californica multicapsid nucleopolyhedrovirus (AcMNPV) as an expression vector and produced human β-interferon and *E.coli* β-galactosidase, respectively (King and Possee, 1992).

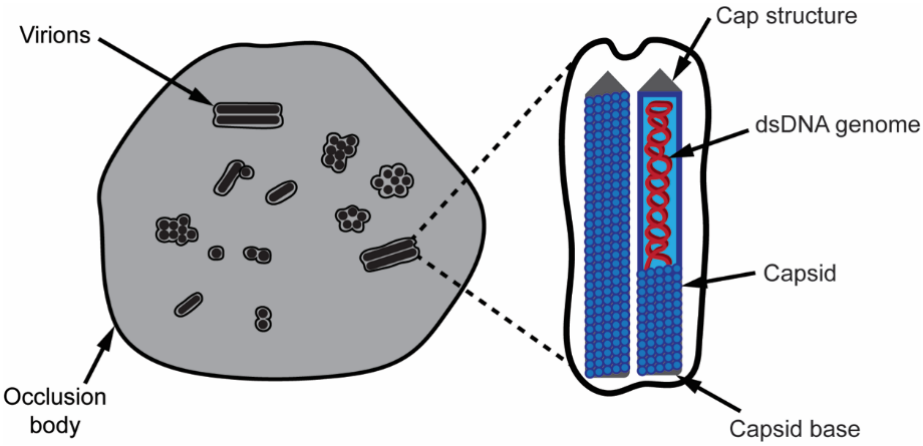


Figure 7.1: AcMNPV polyhedra. Schematic of AcMNPV polyhedra with occluded baculovirus particles.

The Bac-to-Bac® system is a commercially available kit from Invitrogen used for the generation of recombinant baculoviruses for protein expression in insect cells. A schematic of the expression system is shown in Figure 7.2. The gene of interest is cloned into a pFastBac™ vector, in this instance pFastBac1™ (referred to from this point on as FastBac1), under the control of AcMNPV polyhedrin (PH) promoter. The multiple cloning site of this vector is flanked by the left and right components of the bacterial transposon Tn7 forming a mini-Tn7 element. The plasmid containing the foreign gene is then transformed into DH10Bac™ (referred to as DH10Bac) cells housing a mini-*att*Tn7 target site on a bacmid that can be propagated in both *E.coli* and insect cells. In addition, a helper plasmid provides transposition proteins in trans and confers tetracycline resistance. Transposition occurs between the Tn7 sites of bacmid and pFastBac1, creating a recombinant bacmid containing the foreign gene. Transposition is selected for via colour dominant selection as the foreign gene is inserted within the LacZα gene. Recombinant bacmids are then purified and transfected into SF9 insect cells for the production of proteins.

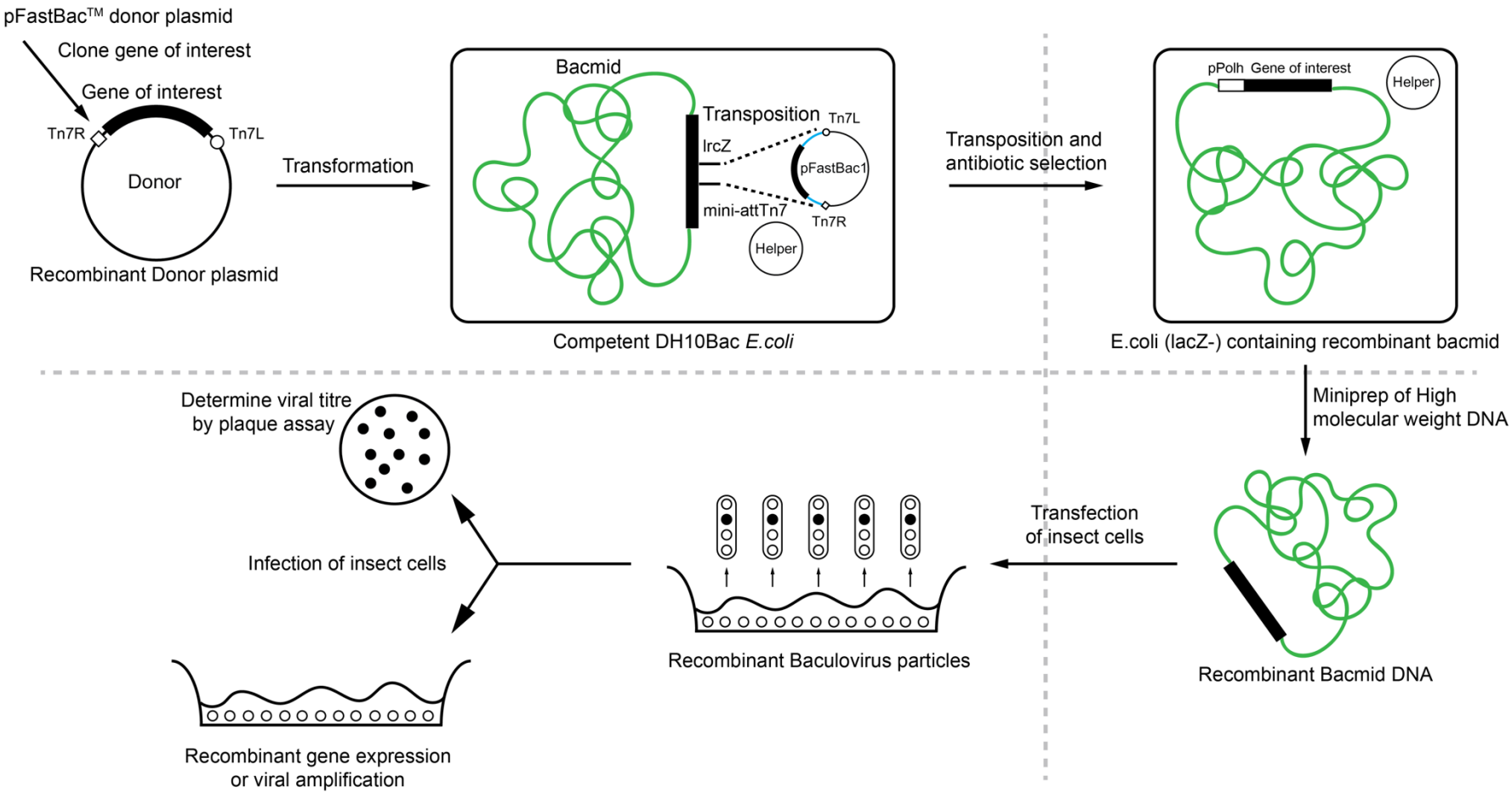


Figure 7.2: Overview of the Bac-to-Bac system from invitrogen. Schematic adapted from Bac-to-Bac® Baculovirus Expression System manual illustrating the stages in baculovirus generation.

7.2 Results

7.2.1 Preparation of FastBac1 vector

The multiple cloning site of the FastBac1 plasmid contains 16 restriction sites downstream of the polyhedrin promoter (Life Technologies, 2015). Restriction enzymes used to linearize the vector in preparation for cloning were BamHI-High Fidelity (HF) and HindIII-HF (NEB). Digestions were carried out simultaneously during this method and incubated for 4 h at 37˚C using a thermal cycler and heat inactivated at 80˚C for 20 min. Samples were isolated using the PCR and DNA clean up kit (Monarch) as per manufacturer's instructions. Aliquots of digested vector were analysed on a 1% (w/v) agarose gel and visualised using Midori Green Stain (Nippon Genetics), and a blue light box (Figure 7.3). The FastBac1 Vector is 4775 bp uncut (Life Technologies, 2015) and after the double digestion the vector loses 105 bp from the multiple cloning site resulting in a size of 4670 bp. As seen in Figure 7.3 the digested plasmid runs at approximately 4.7 kbp.

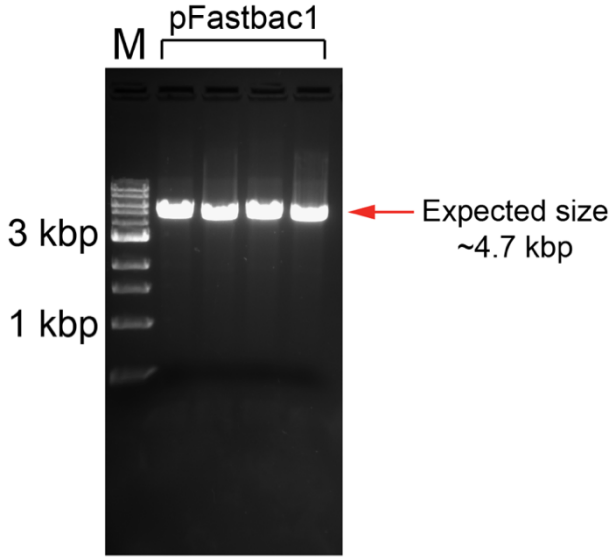


Figure 7.3: pFastBac1 digestion. 1% agarose gel of pFastBac1 digested with HindIII and BamHI. Digestion was carried out using CutSmart buffer from NEB and stained using midori green (Geneflow). M indicates 1 kb ladder (NEB).

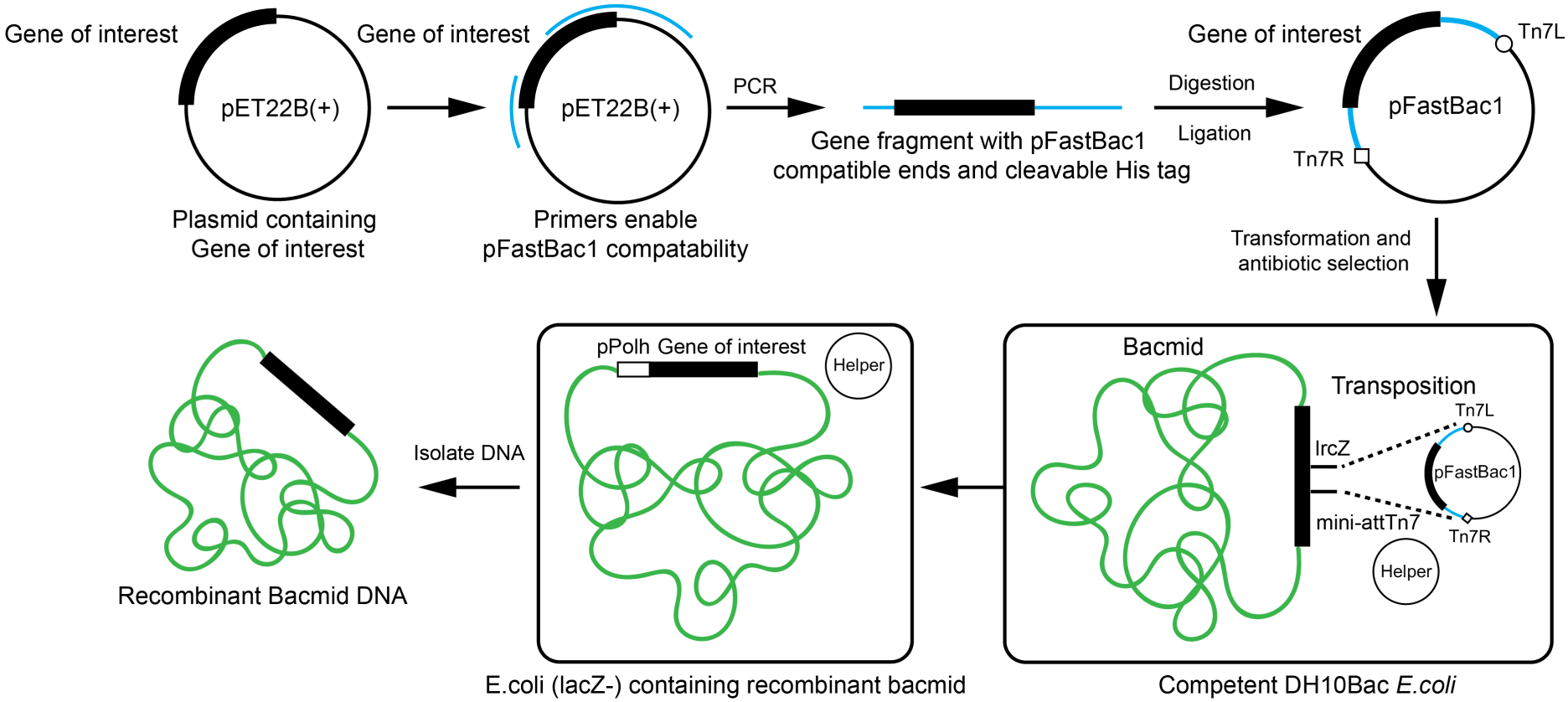


Figure 7.4: Modified bac-to-bac protocol used for generation of recombiant bacmids. Primers were designed to amplify the alphavirus core protein genes out of pET22b vectors and make them compatible with pFastBac1. These fragments were then used in a HiFi DNA assembly reaction with linearised pFastBac1 and transformed into competent cells.

7.2.2 Isolation of alphavirus core protein genes from pET22b vector.

As alphavirus core protein sequences were already present in pET22B vectors these constructs were used as templates to amplify out these fragments for use in baculoviral expression. Primers for the adaptation of the pET22b alphavirus plasmids were designed using the NEBuilder Assembly tool (v1.12.15) and ordered from IDT (Appendix B.2). These primers were used in PCR reactions using the pET22b alphavirus plasmids as templates. This method is an adaption of the Invitrogen bac-to-bac protocol and is illustrated in Figure 7.4. This method was attempted initially with the CHIKV Cp insert (Figure 7.5A) as a trial followed by the production of the remaining 4 alphavirus core protein gene fragments (Figure 7.5A). These samples were isolated using the PCR and DNA clean up kit (Monarch) as per manufacturer's instructions. Figure 7.5 shows bands of ~850 bp which is the expected size for these fragments.

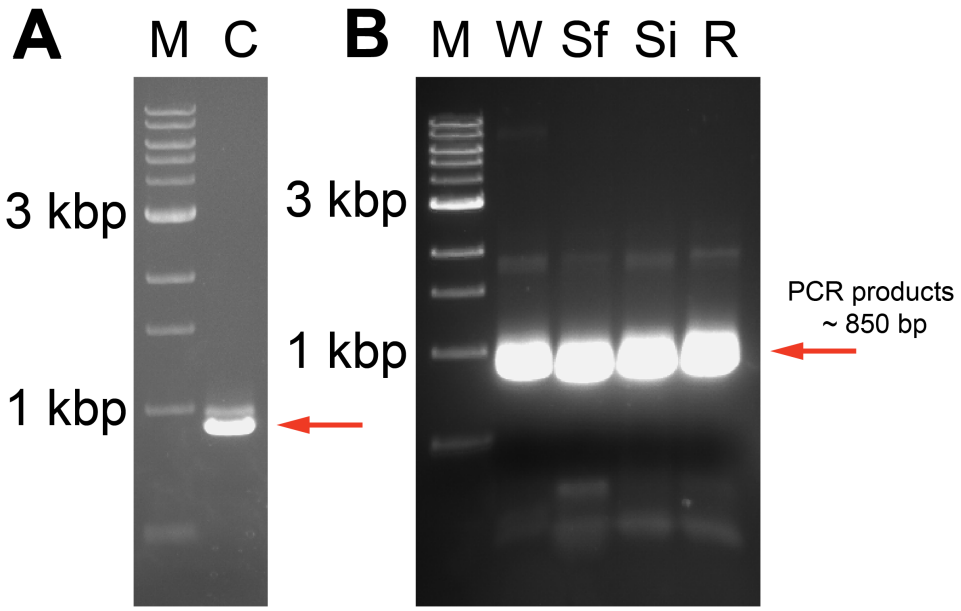


Figure 7.5: Q5 PCR of alphavirus core protein inserts. 1% agarose gel of products from Q5 PCR amplification for **(A)** CHIKV (C), **(B)** WEEV (W), SFV (Sf), SINV (Si) and RRV (R). M indicates 1 kb ladder (NEB).

7.2.3 HiFi DNA assembly with FastBac1

The fragments produced via Q5 PCR were assembled with HindIII-BamHI digested FastBac1 using the HiFi DNA assembly protocol from NEB outlined in Figure 7.6. During this procedure linearised vector from the restriction digest was incubated at 50°C with PCR fragments and the HiFi DNA Assembly Master Mix (NEB). Exonuclease within the master mix creates single stranded complementary regions within the designed overlap which anneal. DNA polymerase extends these annealed fragments and DNA ligase repairs the strands resulting in a double stranded plasmid containing the gene of interest. The plasmid was transformed into DH5α cells and selected via Carbenicillin resistance and single colonies were picked and grown overnight at 37˚C, 220 rpm. Agarose plates for all alphavirus inserts, and CHIKV Cp performed in 3 ratios (Insert:Vector) as a preliminary test, are shown in Figures 7.7 and 7.8.

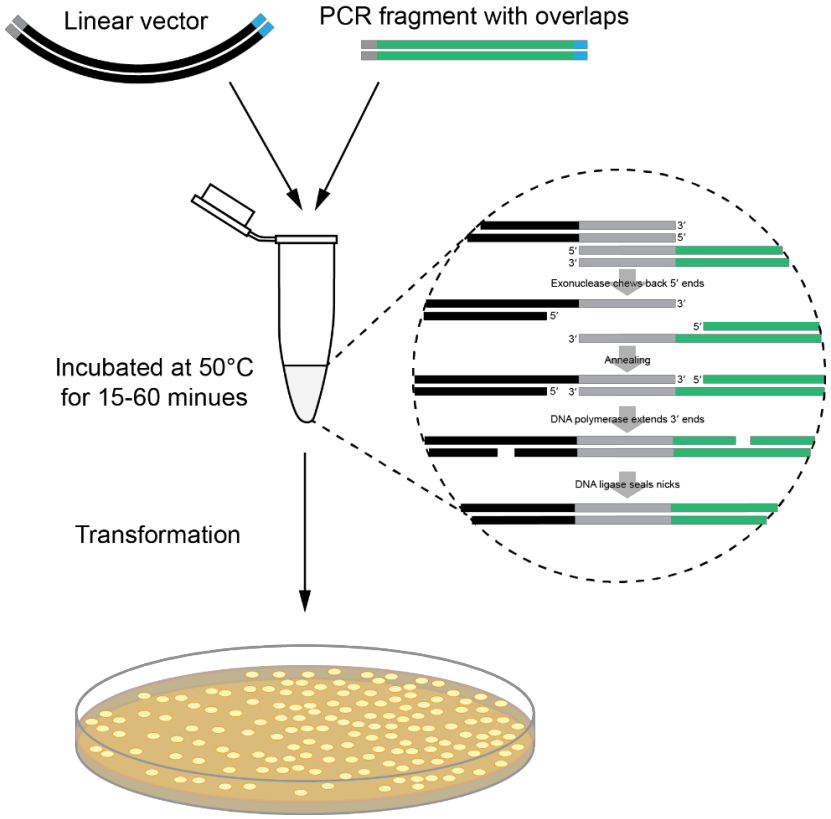


Figure 7.6: Schematic of the HiFi DNA assembly process. Adapted from the NEB website (New England Biolabs 2017)

Single colonies were picked and plasmid DNA isolated using a Qiagen miniprep kit as per manufacturer instructions, quantified using a NanoDropOne (Thermo Fisher) and submitted for sequencing with polyhedrin (Forward) and SV40polyA (Reverse) primers.

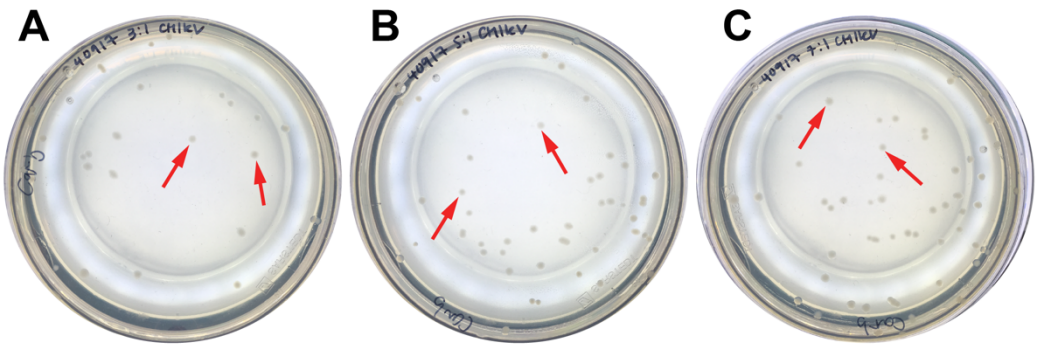


Figure 7.7: Agar plates of CHIKV HiFi DNA assembly. HiFi DNA assemblies of CHIKV Cp in 3:1 **(A)**, 5:1 **(B)** and 7:1 **(C)** ratios transformed into DH5α cells and incubated at 37˚C overnight with 100 µg/mL Carbenicillin.

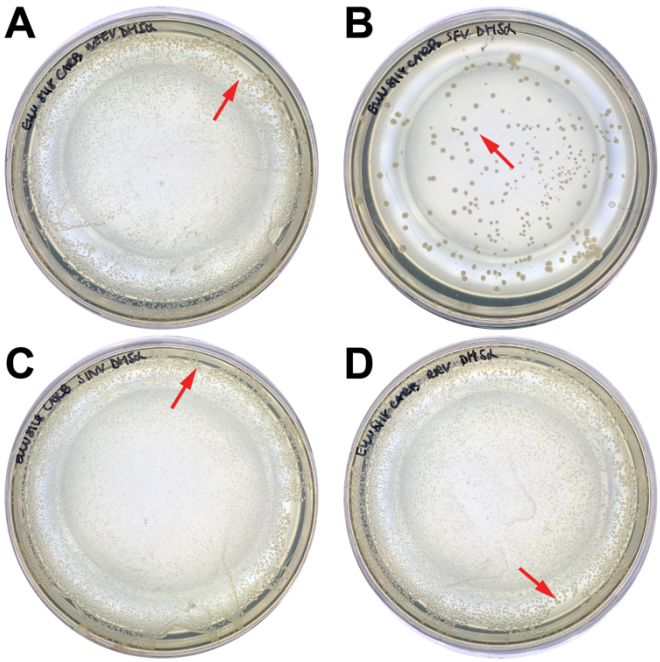


Figure 7.8: Agar plates of WEEV, SFV, SINV and RRV HiFi DNA assemblies. HiFi DNA assemblies in 3:1 ratio of WEEV Cp **(A)**, SFV Cp **(B)**, SINV Cp **(C)** and RRV Cp **(D)** transformed into DH5α cells and incubated at 37˚C overnight with 100 µg/mL Carbenicillin.

7.2.4 Transformation into DH10Bac

Sequencing showed that all the alphavirus core protein genes had been successfully inserted into FastBac vectors. These isolates were now ready to be transformed into DH10Bac cells. This cell line contains a baculovirus shuttle vector (Bacmid) hosting a target site for mini-attTn7, a kanamycin resistance marker and a mini F replicon. Tn7 transposition function is provided in *trans* by a helper plasmid encoding for transposase and tetracycline resistance. The FastBac vector contains the left and right components of the bacterial transposon Tn7, forming a mini-Tn7 element enveloping the multiple cloning site and a marker for gentamycin resistance. This allows for the transposition of the multiple cloning site, and therefore the gene of interest, into the bacmid shuttle vector. The components of this transposition are shown in Figure 7.9

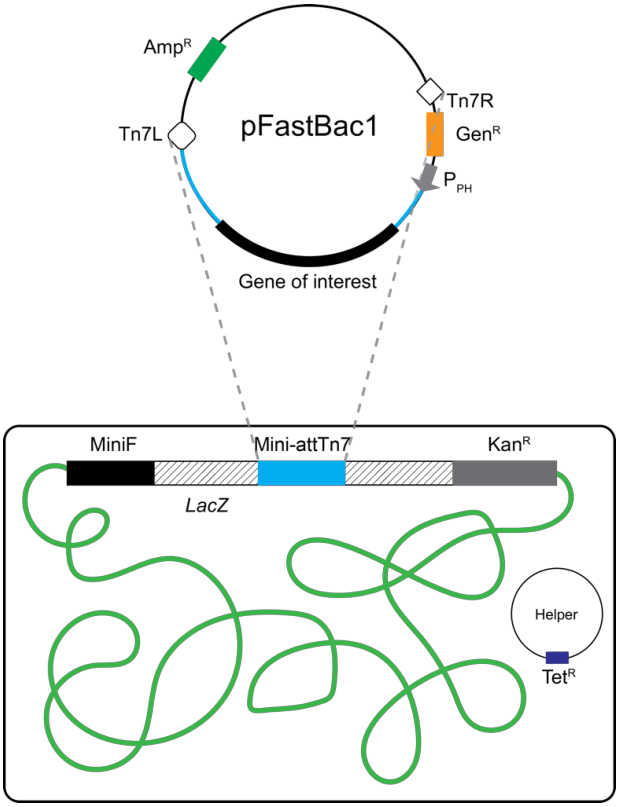


Figure 7.9: FastBac1 transposition. Diagram of transposition between the FastBac1 vector and bacmid shuttle vector in DH10Bac cells.

Transformed cells were serially diluted (10^{-1}, 10^{-2}, 10^{-3}) in SOC (Invitrogen) and 100 µL was plated out onto LB-Agar plates containing 50 µg/mL kanamycin, 100 µg/mL Bluo-gal, 7 µg/mL gentamicin, 10 µg/mL tetracycline and 40 µg/mL IPTG (referred to as KXGTI plates) and incubated at 37°C for 48 hours. The mini-attTn7 site of the bacmid lies within the LacZα gene and therefore insertion at this site disrupts expression of the peptide. This allows for selection of positive transformants via colour dominant selection when grown in the presence of Bluo-Gal (halogenated indolyl-β-galactoside). Colonies which contain the bacmid are unable to process the Bluo-Gal and will remain white, whereas colonies which contain untransposed bacmid will appear as blue colonies. Serial dilutions of all alphavirus bacmids on KXGTI plates are shown in Figures 7.10 and 7.11, CHIKV Cp (Figure 7.10) was carried out in triplicate (CHIKV-32, CHIKV-33 and CHIKV-34) as a trial run before the remaining 4 alphavirus core proteins. Transposition of the control plasmid pUC19 was also carried out and is shown alongside CHIKV Cp bacmids in Figure 7.10.

From the plates in Figures 7.10 and 7.11, 10 white colonies for each viral core protein were picked and re-streaked out onto KXGTI plates and incubated overnight at 37°C. Single colonies with a twice white phenotype were picked and used to inoculate 2 mL LB-50 µg/mL kanamycin, 7 µg/mL gentamicin, and 10 µg/mL tetracycline and incubated at 37°C, 250 rpm overnight.

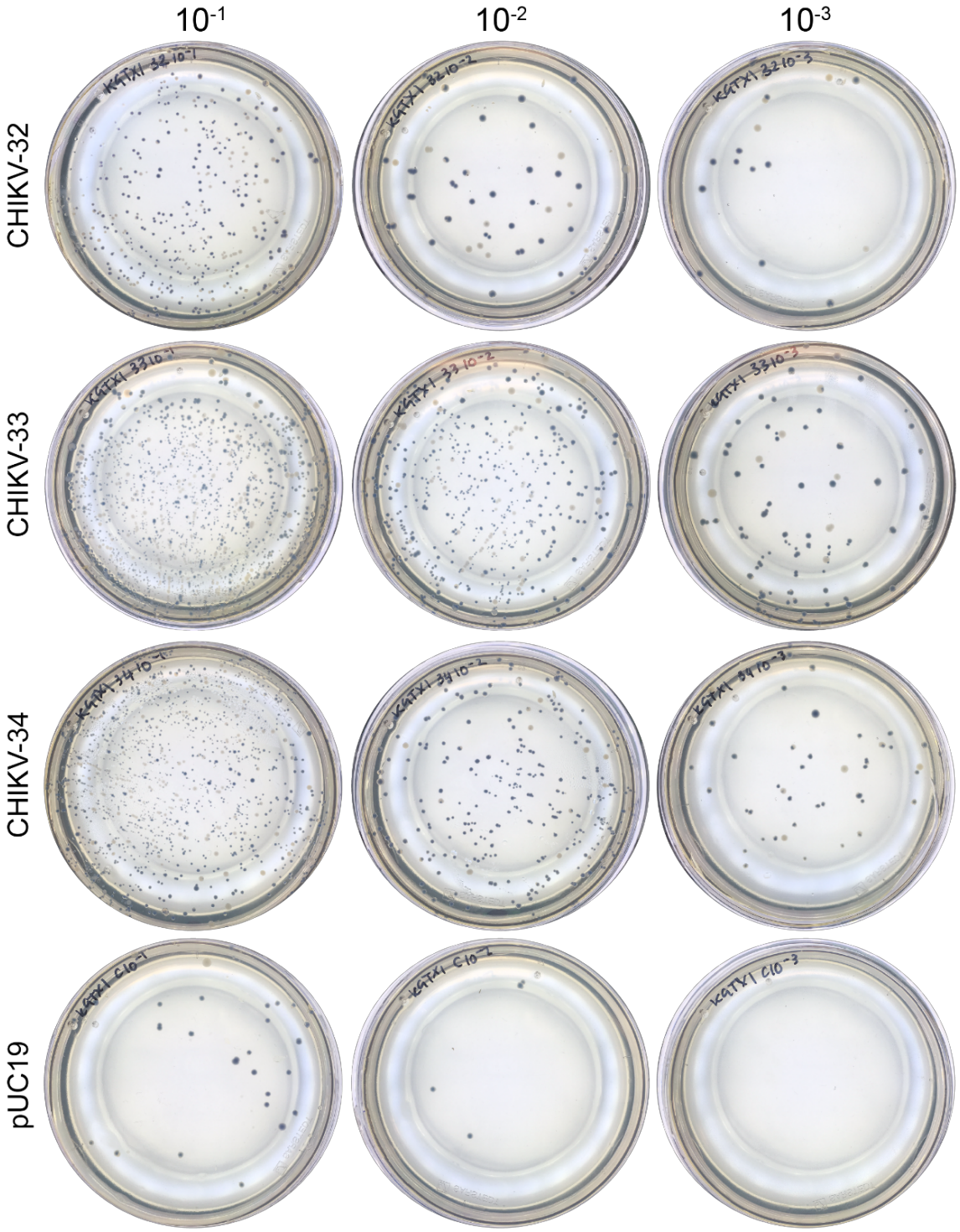


Figure 7.10: Transposition of CHIKV Cp and pUC19 control plasmid. Photographs of KXGTI plates with serial dilutions of CHIKV Cp and pUC19 control after incubation at 37°C for 48 hours. Presence of Bluo-Gal in the LB-Agar allows for the selection of white colonies which contain the recombinant bacmid.

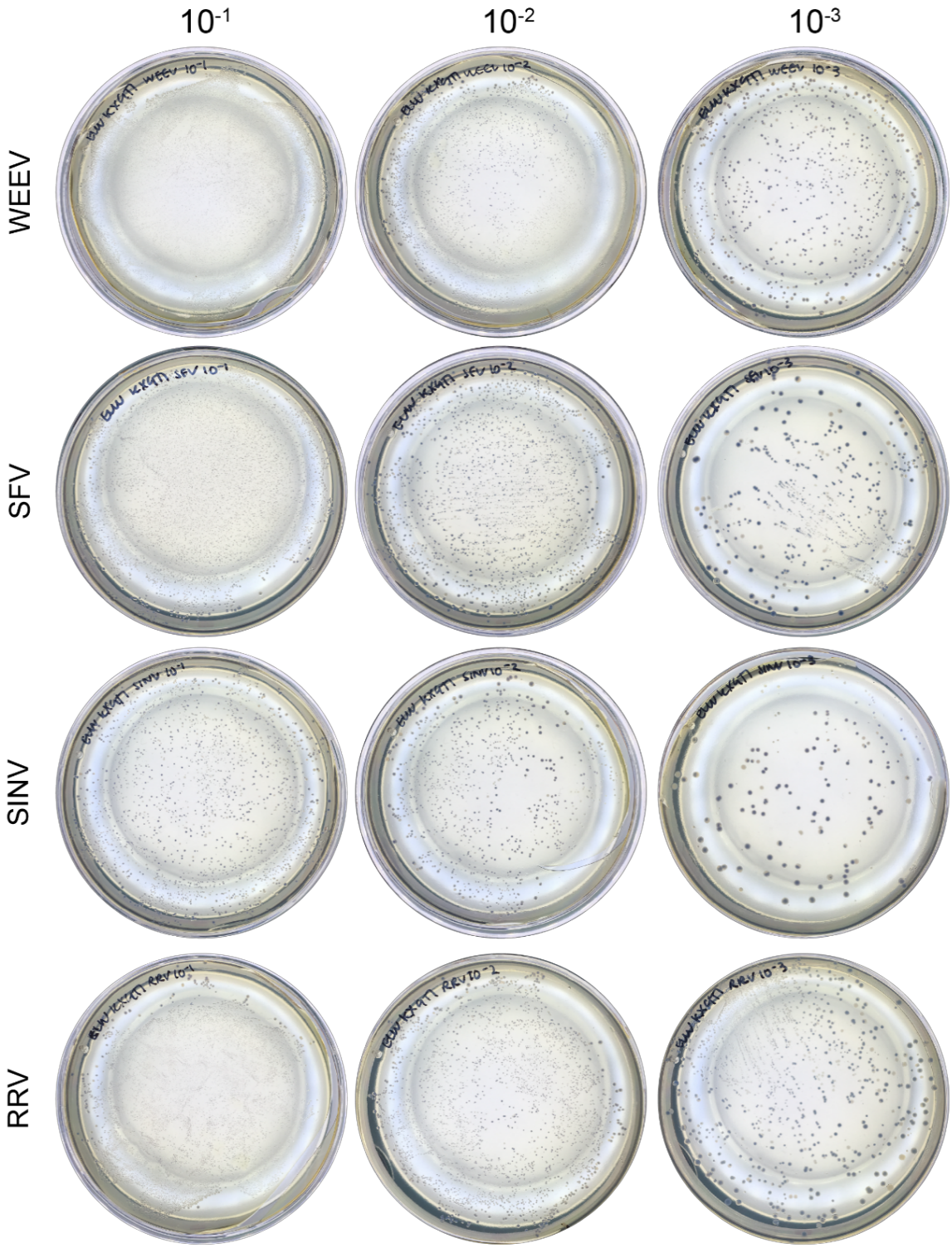


Figure 7.11: Transposition of WEEV, SFV, SINV and RRV Cp. Photographs of KXGTI plates with serial dilutions of WEEV, SFV, SINV and RRV after incubation at 37°C for 48 hours. Presence of Bluo-Gal in the LB-Agar allows for the selection of white colonies which contain the recombinant bacmid.

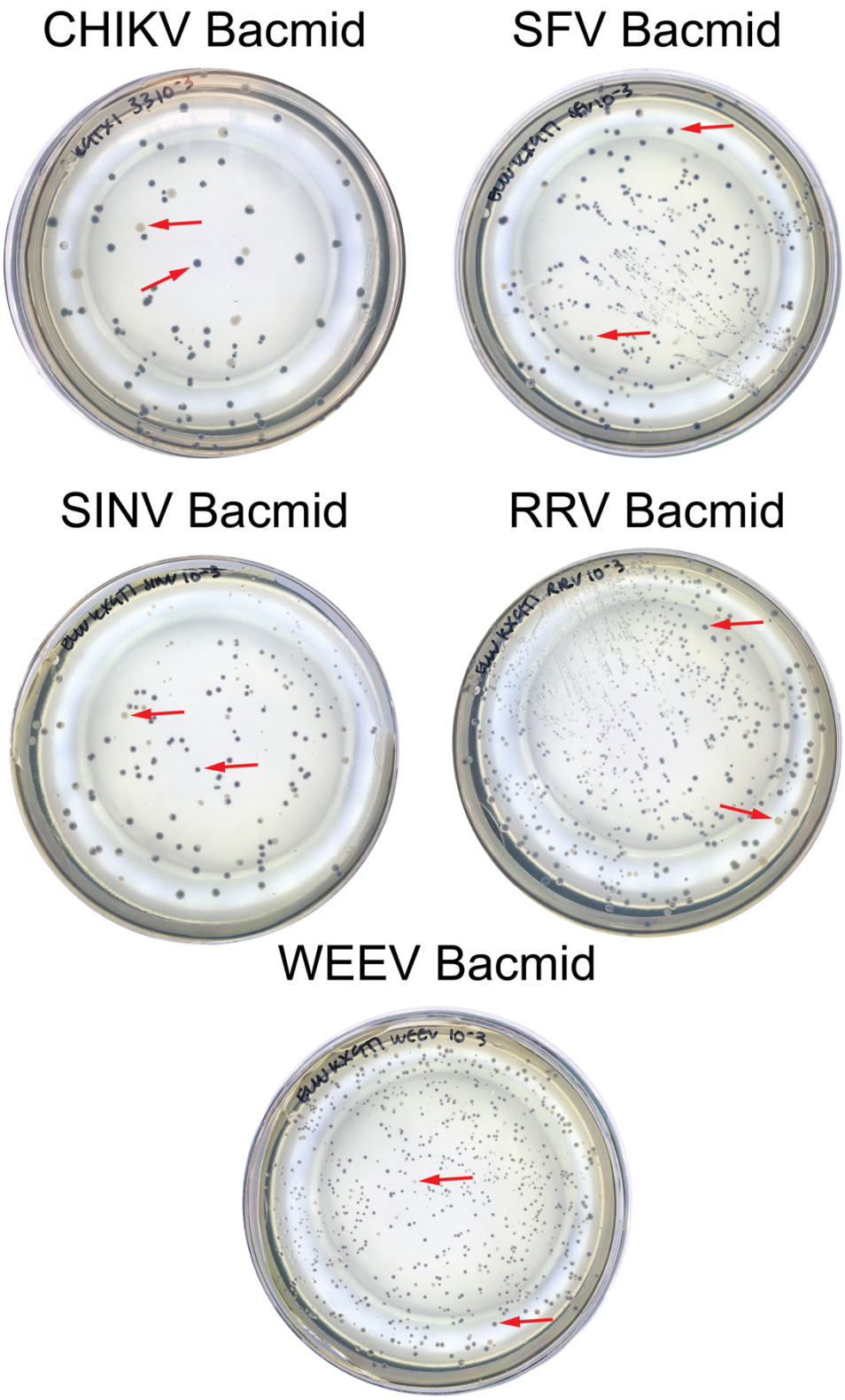


Figure 7.12: Blue-white screening of DH10Bac cells. Photographs of KXGTI plates for CHIKV, SFV, SINV, RRV and WEEV after incubation at 37°C for 48 hours. Blue and white colonies are highlighted for each plate with red arrows.

7.2.5 Purification of Alphavirus core protein Bacmids

Recombinant bacmid DNA was purified from the overnight liquid cultures of 12, twice white colonies using PureLinkTM HiPure Plasmid DNA Miniprep Kit (Invitrogen) following manufacturer's instructions. Concentrations of the resulting purified bacmids was measured using a NanoDropOne (Thermo Fisher). Before transposition, the bacmid (bMON14272) is 136 Kb (Invitrogen, 2002) so, with the addition of the genes of interest at ~ 850 bp the transposed bacmid is approximately 137 Kb. High molecular weight DNA of this size is difficult to analyse by restriction digest however, PCR analysis using pUC/M13 forward and reverse primers can be used to demonstrate the presence of the foreign gene. As shown in Figure 7.13 the distance between the pUC/M13 forward and reverse primers is approximately 300 bp, therefore a bacmid which has not undergone transposition will produce a band of ~300 bp. A bacmid transposed with pFastBac1 will produce a fragment of 2300 bp + the size of the inserted gene. In this instance the expected fragment is 2300 bp + ~850 bp = ~3150 bp. PCR analysis was performed using Q5 DNA polymerase (NEB) as outlined in 2.7.1.3.

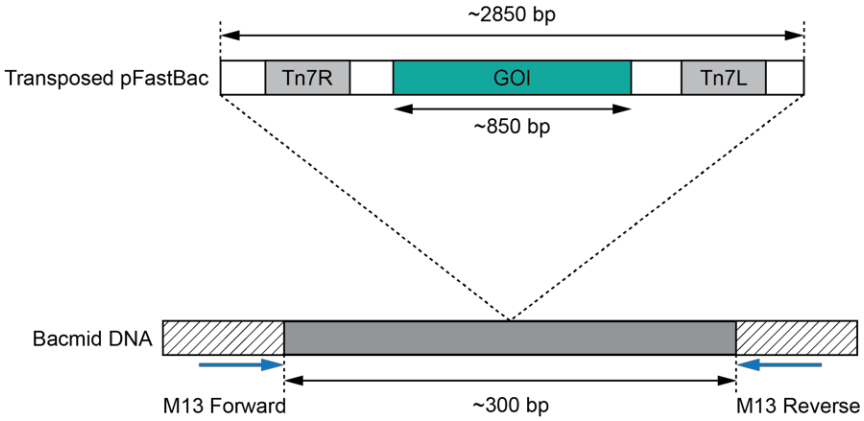


Figure 7.13: Bacmid analysis using pUC/M13 primers. Adapted from Invitrogen Bac-to-Bac handbook. Untransposed bacmid will produce a fragment of 300 bp transposed bacmid will produce a fragment of 2300 bp + insert.

After completion of the PCR program, the 12 twice white samples were analysed by 1% agarose gel as shown in Figures 7.14 and 7.15. The CHIKV isolates seen in Figure 7.14 all contain DNA fragments of the correct size after PCR amplification with pUC/M13 primers. However, there are other bands, most commonly at ~1.4 kb, seen in all lanes but 2 and 8. It is unclear as to what these bands represent as untransposed bacmid would be seen at ~300 bp and bacmid transposed with empty FastBac1 vector would produce a band at 2.3 kb. Samples with extra bands at 1.4 kb were not used for downstream investigations due to the unknown nature of this band. Isolate 2 from CHIKV was carried forward for further investigation.

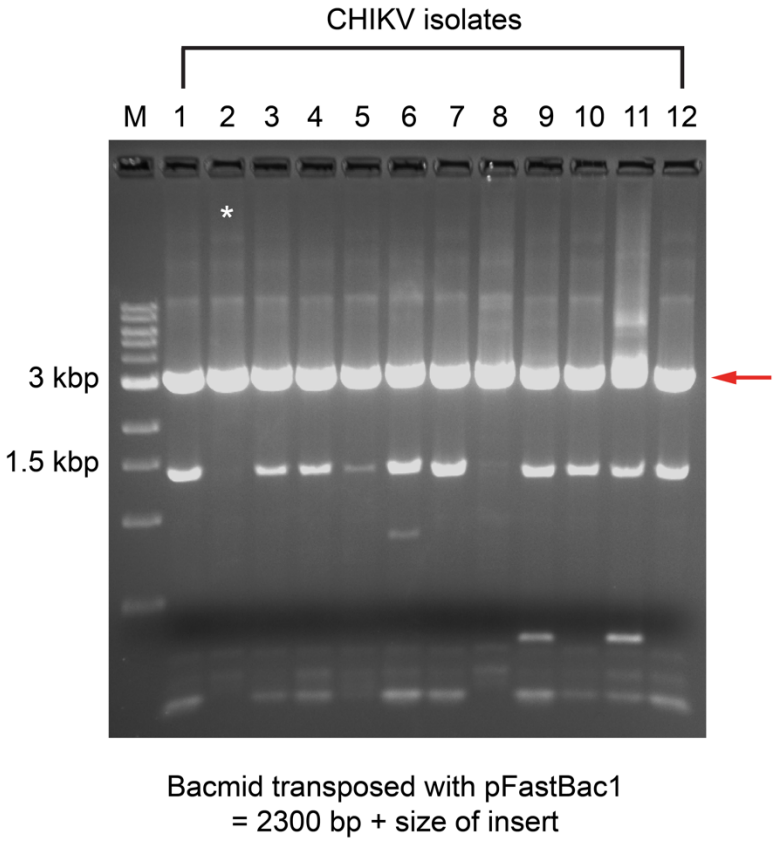


Figure 7.14: Analysis of CHIKV Bacmids. 1% Agarose gel of Bacmid samples after PCR amplification with pUC/M13 forward and reverse primers. Strong bands are seen at ~3Kb which indicates presence of gene of interest (2300 bp + ~850 bp). M = 1 kb DNA ladder from NEB. * signifies isolates carried forward for further investigation.

Of the 10 twice white samples picked for each remaining alphavirus core protein (SFV, SINV, RRV and WEEV) 5 isolates from each were taken forward for PCR analysis (Figure 7.15). As discussed previously a correctly transposed bacmid will produce a fragment of ~3 kb indicated by the red arrow. For all four remaining core proteins at least one of the five tested isolates were correctly transposed. Additionally, the band seen at ~1.4 kb in the CHIKV analysis are not seen in these samples. There are however, some samples which contain bands at ~2 kb which could indicate successful transposition occurred with an empty FastBac1 vector as this situation would also produce white colonies.

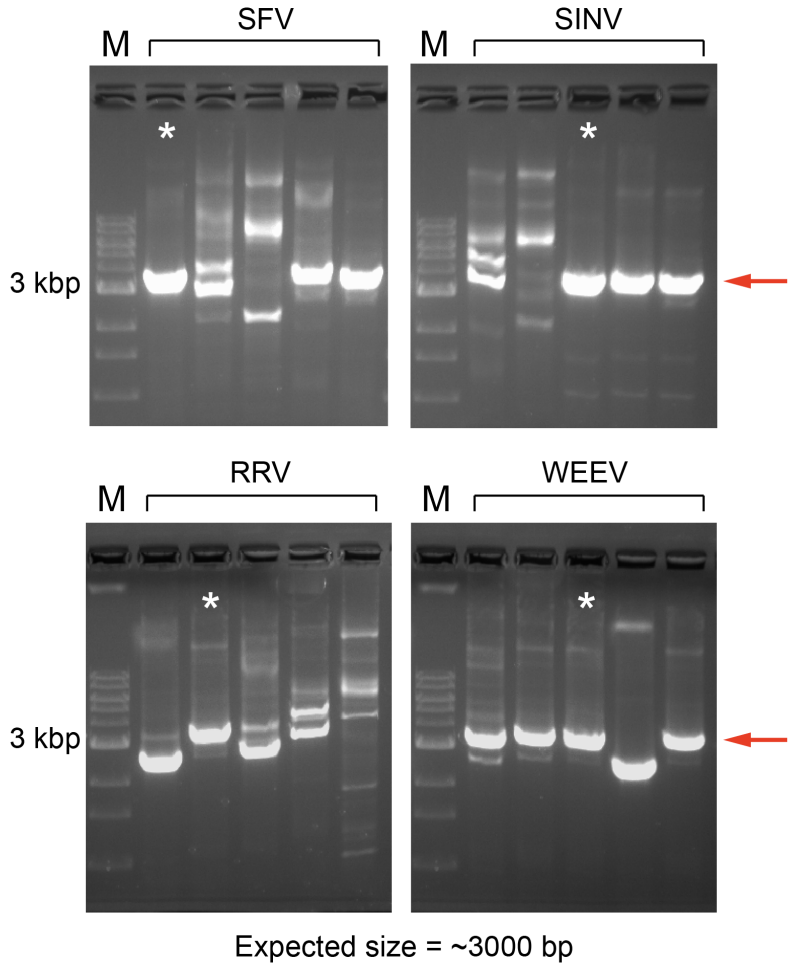


Figure 7.15: Analysis of SFV, SINV, RRV and WEEV Bacmids. 1% Agarose gel of Bacmid samples after PCR amplification with pUC/M13 forward and reverse primers. Strong bands are seen at ~3Kb which indicates presence of gene of interest (2300 bp + ~850 bp). M = 1 kb DNA ladder from NEB.* indicates isolates carried forward for further investigation.

7.2.6 Generation of recombinant baculovirus

7.2.6.1 Generation of recombinant CHIKV baculovirus

Recombinant baculoviruses for the production of protein were produced by transfection of insect cells with recombinant Bacmid DNA. For maintenance of SF9 cells and transfection conditions see materials and methods 2.7.3.1. Figure 7.16 shows non-transfected SF9 cells and cells transfected with CHIKV bacmid after 96 hours. Non-transfected cells are generally uniform in size and form a complete layer across the well of a 6 well plate. Some of the transfected cells appear larger in comparison to the control cells and there are gaps in the cell monolayer where lysis has occurred. The media from the transfected cells was harvested, centrifuged and filtered to remove cellular debris, forming a P1 viral stock.

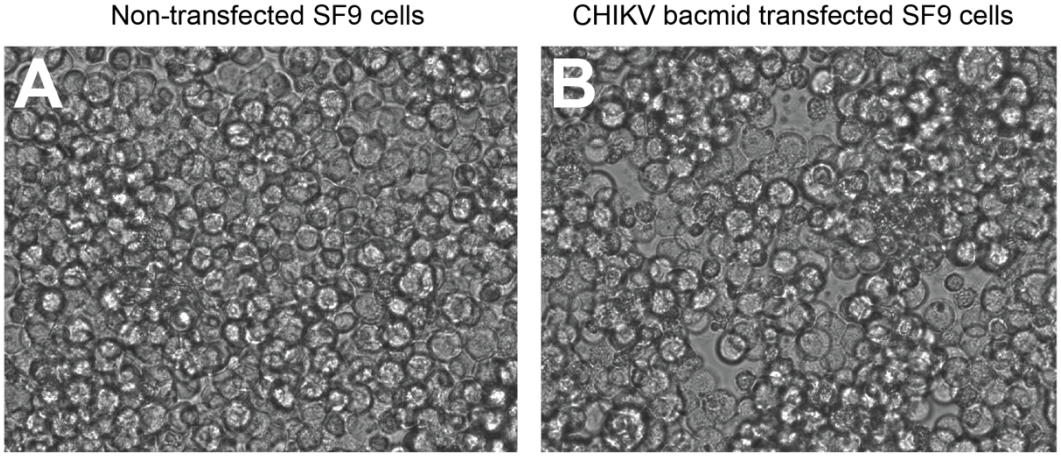


Figure 7.16: CHIKV bacmid transfection of SF9 cells. (A) Mock transfected SF9 insect cells at 40X magnification. **(B**) SF9 cells transfected with CHIKV bacmid at 40X magnification**.**

The CHIKV Cp recombinant baculovirus P1 viral stock was amplified to form a higher titre P2 viral stock which could be used for the expression of viral core proteins. This step is recommended by Invitrogen as a typical P1 stock has a titre of $1x10^6$ to $1x10^7$ plaque forming units (pfu)/mL whereas a P2 titre can range from $1x10^7$ to $1x10^8$ pfu/mL. A 30 mL suspension culture of SF9 cells ($2x10^6$ cells/mL) was infected with P1 viral stock at a presumed titre of $3x10^6$

pfu/mL (multiplicity of infection, MOI = 0.1) and the cell number and viability monitored (28°C, 120 rpm) until CHIKV Cp recombinant baculovirus infected cells diverged from the control cells. Figure 6.17 A shows the cell numbers for both the control and the infected cells, where the line indicates the average over 8 counted squares of a haemocytometer and the error bars indicate the standard deviation of those 8 counted squares. Both cultures initially increase cell number however, at 96 h post infection cell counts begin to diverge with the difference at 120 hours being almost 50%. This is mirrored in the cell viability results (Figure 7.17 B), initially both samples are ~99% viable however, as time progresses viability of the infected cells drops to 90% whereas the control cells remain 98% viable.

The P2 sample was centrifuged and filtered to remove cellular debris before being used to perform a viral plaque assay and hence determine the viral titre of the P2 viral stock. The plaque assay was performed in 2 x 6 well plates as outlined in 2.7.3.3. 8-log serial dilutions of the P2 viral stock were prepared and dilutions 10^{-4}-10^{-8} were used along with a media only negative control. The plates were incubated at 28°C and monitored every 24 hours for the appearance of plaques. After 10 days plaques were visible in all wells, photographs of which are shown in Figure 7.18. Non-infected cells appear as a monolayer, interruption of which is due to disturbance of the plaquing medium overlay. The wells containing the 10^{-4} dilution appear to have completely lysed and therefore these wells were not counted, additionally, the plaques in the 10^{-5} wells were not distinct enough to allow for accurate counting. Viral titre of P2 was calculated as 4.5×10^{8} pfu/mL.

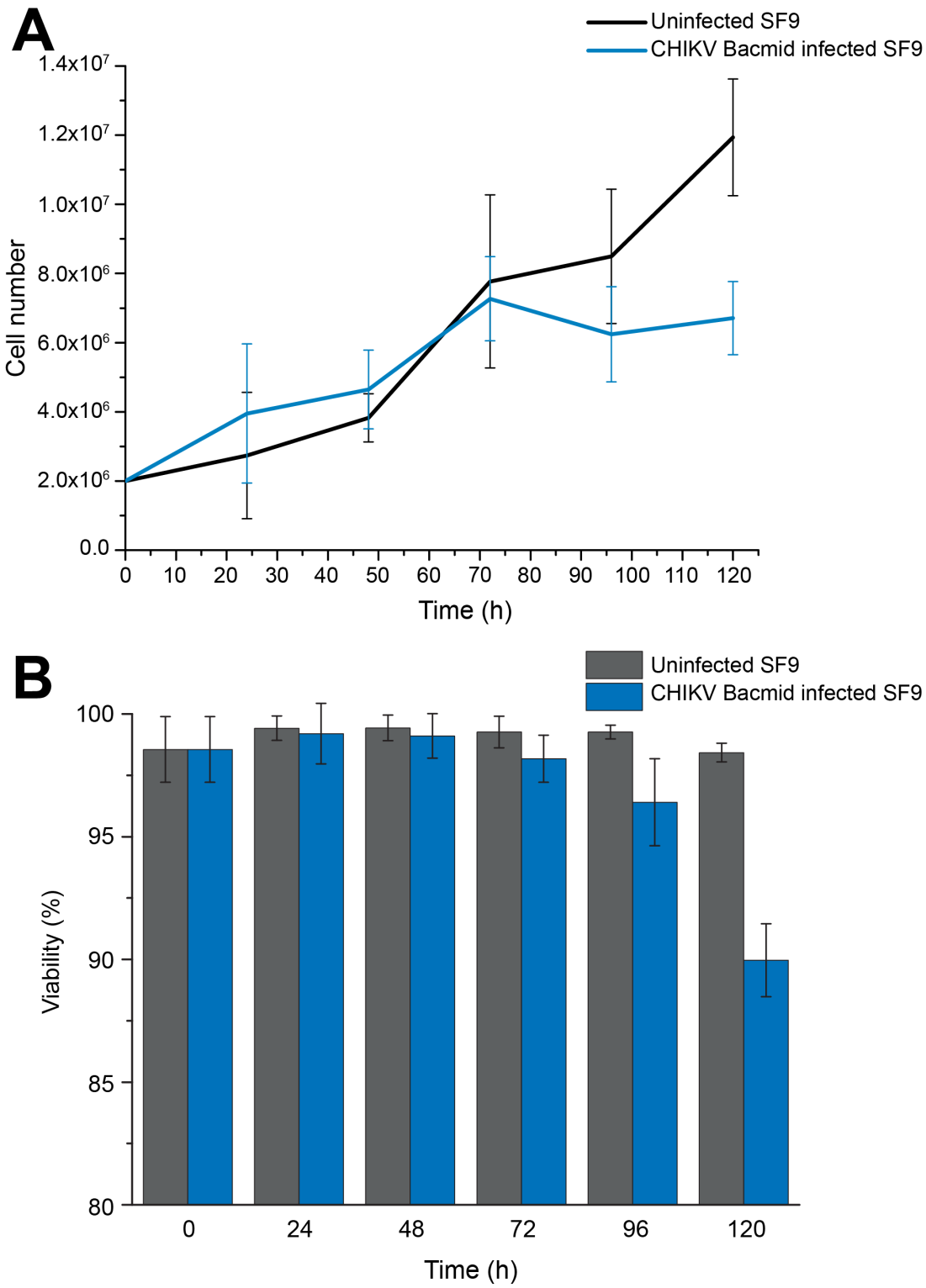


Figure 7.17: Generation of CHIKV P2 viral stock. (A) Cell counts for CHIKV Cp recombinant baculovirus infected and uninfected SF9 cells over 120 hours. Line indicates the average count over 8 squares of a haemocytometer, error bars indicated standard deviation over all counted squares. **(B)** Viability of cells over 120 hours calculated from averages of 8 counted squares.

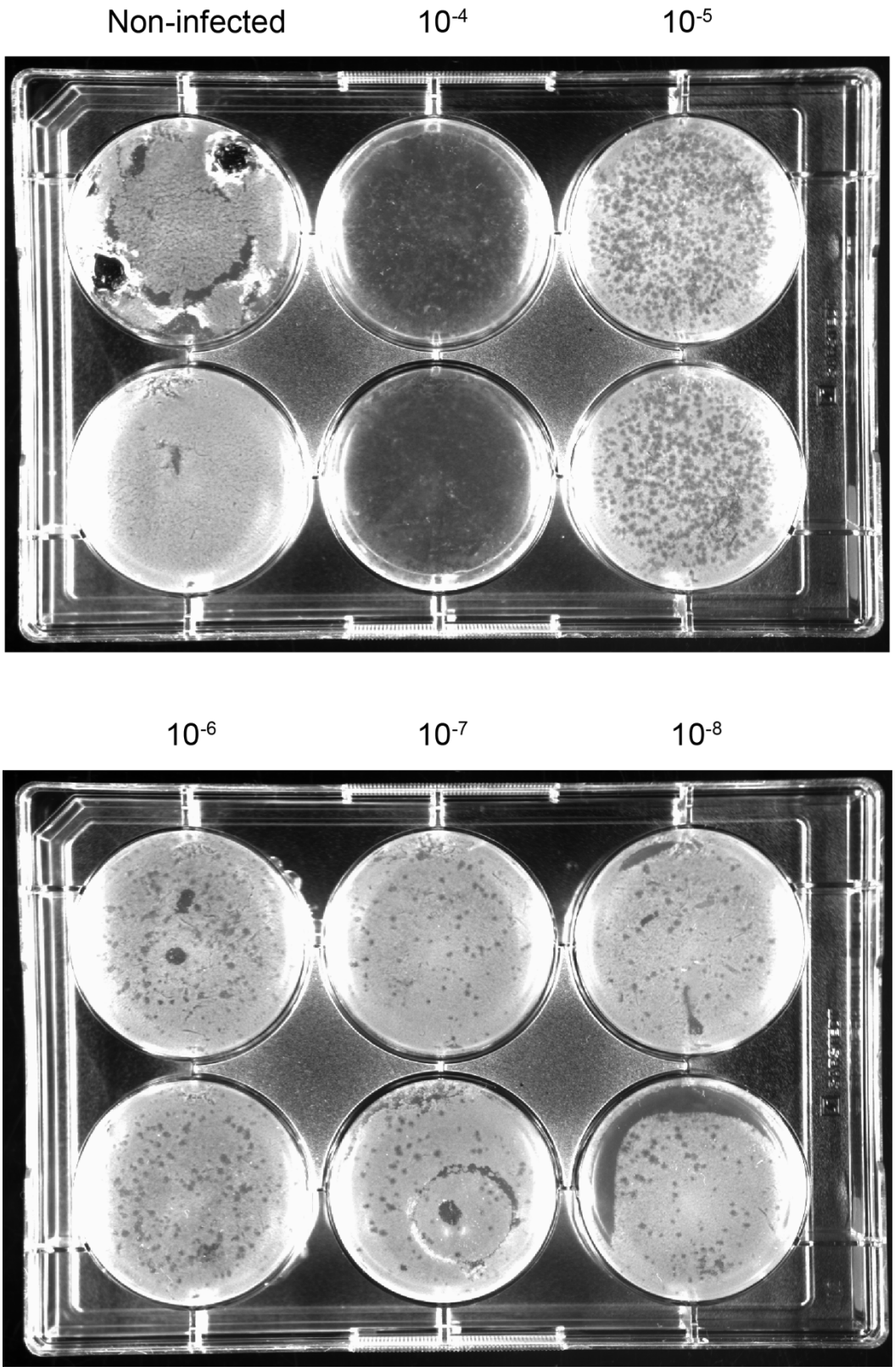


Figure 7.18: Photographs of CHIKV Cp recombinant baculovirus plaque assays. Photograph of plaque assay performed in a six well plate after incubation at 28°C for 10 days. Top: Non-infected, 10^{-4} and 10^{-5} dilutions in duplicate. Bottom: 10^{-6}, 10^{-7} and 10^{-8} dilutions in duplicate.

7.2.6.2 Generation of recombinant WEEV baculovirus

Transfection of WEEV bacmid was carried out as per 2.7.3.1. Figure 7.19 shows non-transfected SF9 cells and cells transfected with WEEV bacmid after 96 hours. As seen previously, non-transfected cells are typically uniform in size and form a monolayer, whereas a number of the WEEV bacmid transfected cells appear larger and the monolayer is incomplete. The media from the transfected cells was harvested, centrifuged and filtered to remove cellular debris, forming a P1 viral stock.

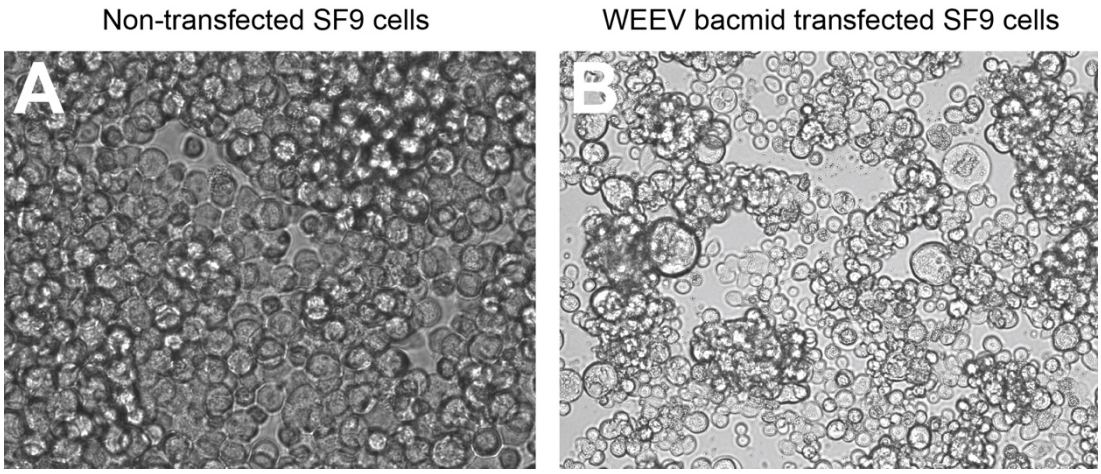


Figure 7.19: WEEV bacmid transfection of SF9 cells. (A) Mock transfected SF9 insect cells at 40X magnification. **(B**) SF9 cells transfected with WEEV bacmid at 40X magnification**.**

WEEV Cp recombinant baculovirus P1 was amplified to form a higher titre P2 viral stock to be used in the larger scale expression of viral core proteins. 30 mL of SF9 cells ($2x10^6$ cells/mL) were infected with P1 viral stock at a presumed titre of $3x10^6$ pfu/mL (MOI = 0.1) and incubated at 28°C, 120 rpm. Cell number and viability of infected and mock infected cells was monitored until infected cells diverged significantly from the control. Figure 7.20 A shows

cell numbers for both mock and infected cells, with the line indicating the average over 8 counted squares of a haemocytometer and the error bars indicate the standard deviation of those 8 counted squares. Cell numbers increase in both mock and infected cells in the initial 24 hours however, as time progresses cell numbers in the infected sample begin to decrease whereas the control cells continue to rise. This is mirrored in the cell viability results (Figure 7.20 B), initially both samples are ~96% viable however, as time progresses viability of the infected cells drops to 73% whereas the control cells remain ~96% viable.

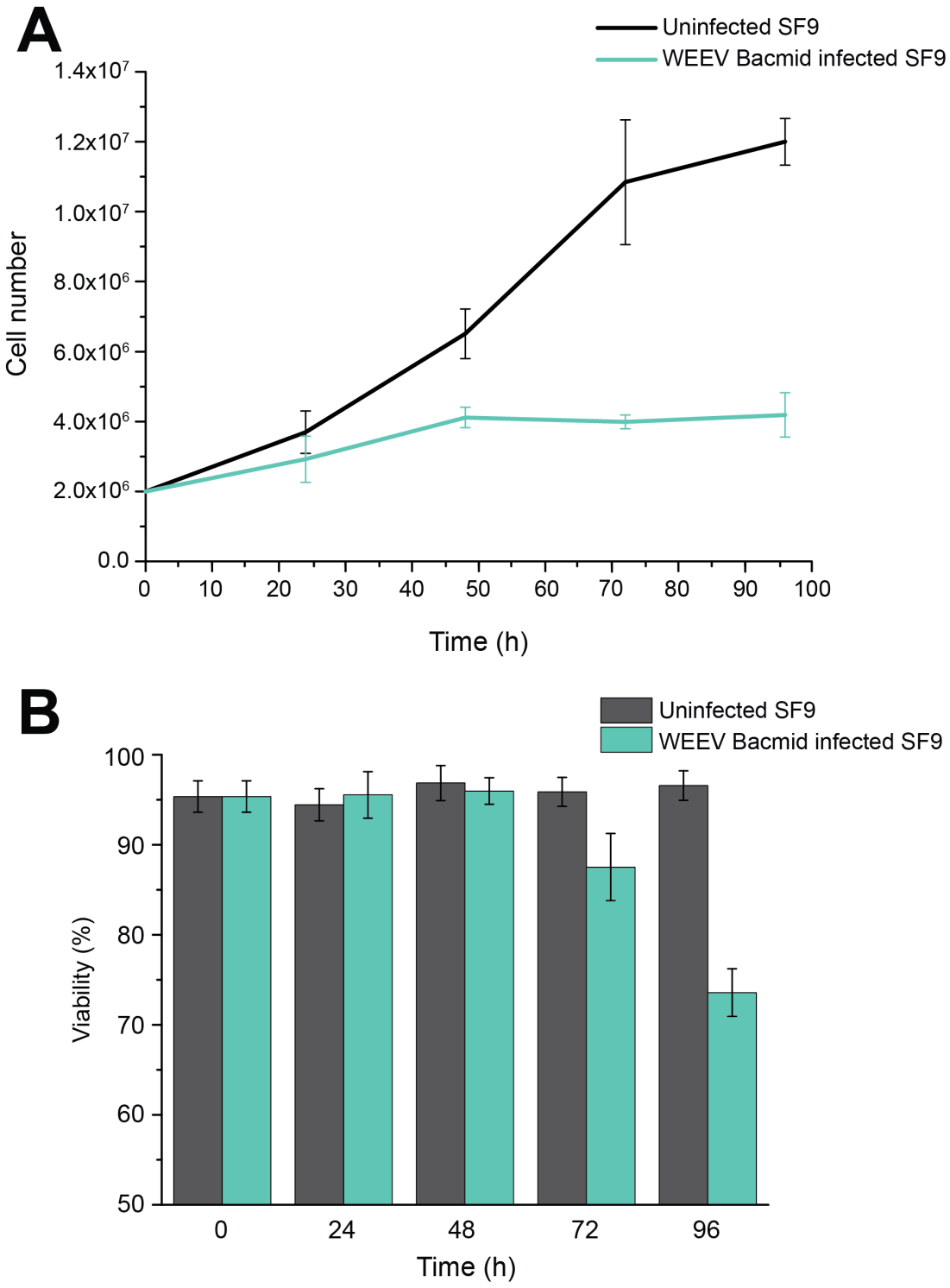


Figure 7.20: Generation of WEEV Cp recombinant baculovirus P2 viral stock. (A) Cell counts for WEEV Cp recombinant baculovirus infected and uninfected SF9 cells over 120 hours. Line indicates the average count over 8 squares of a haemocytometer, error bars indicated standard deviation over all counted squares. **(B)** Viability of cells over 120 hours calculated from averages of 8 counted squares.

7.2.6.3 Generation of recombinant SINV baculovirus

SINV bacmid transfection was carried out as in 2.7.3.1. Figure 7.21 shows non-transfected SF9 cells and cells transfected with SINV bacmid after 96 hours. Many of the cells in the transfected sample are oversized and have a granular appearance alongside evidence of cell lysis and clearing of the monolayer.

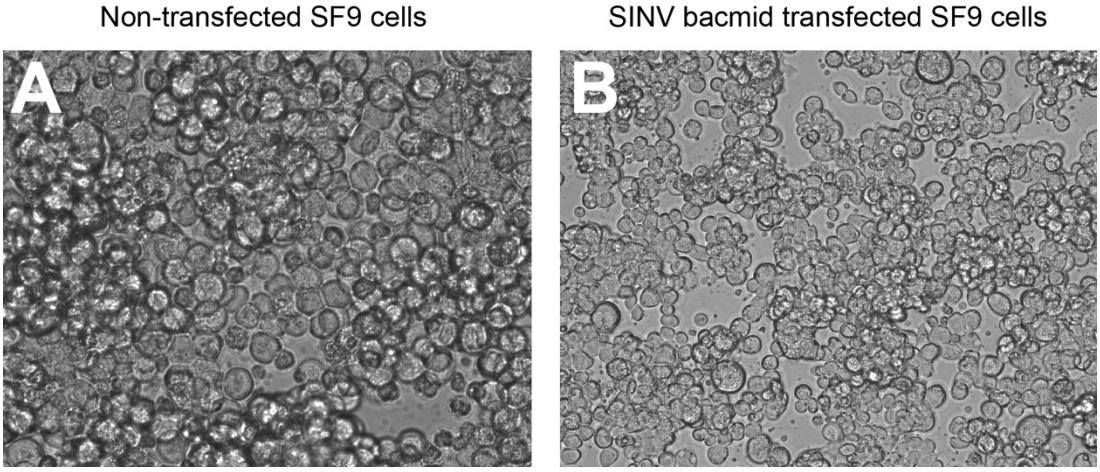


Figure 7.21: SINV bacmid transfection of SF9 cells. (A) Mock transfected SF9 insect cells at 40X magnification. **(B**) SF9 cells transfected with SINV bacmid at 40X magnification

As with previously mentioned baculoviruses, the media from the SINV bacmid transfected cells (P1 viral stock) was centrifuged and filtered before amplification to form a higher titre P2 stock for protein expression. 30 mL of SF9 cells ($2x10^6$ cells/mL) was infected with the SINV Cp recombinant baculovirus P1 viral stock at a presumed titre of $3x10^6$ pfu/mL (MOI = 0.1) and incubated at 28°C, 120 rpm. Cell number and viability of both infected and mock infected cells were monitored until infected cells deviated from the control cells as shown in Figure 7.22. Cell numbers for both the mock and the

SINV infected cells were counted, with the line indicating the average over 8 counted squares of a haemocytometer and the error bars indicate the standard deviation of those 8 counted squares. As seen in Figure 7.22 cell numbers of both mock infected and SINV Cp recombinant baculovirus infected cells increase. At approximately 72 hours cell numbers begin to differ however, viability still remains comparable to the mock infected control. At both the 96 and 120-hour time points cell numbers remain at ~$8x10^6$ cells/mL and viability had dropped to 85 and 80% respectively whereas the control cells remain >96% viable.

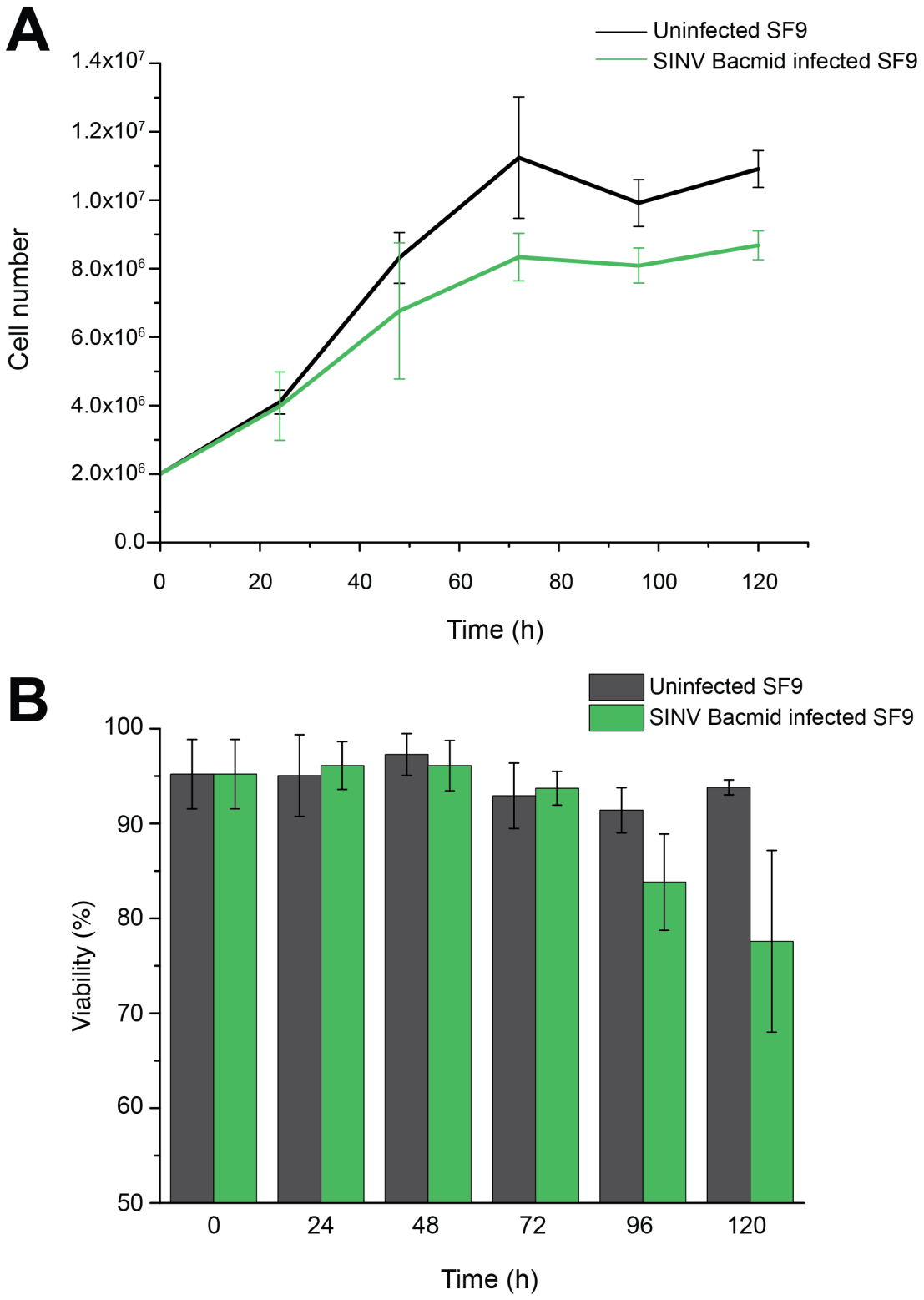


Figure 7.22: Generation of SINV Cp recombinant baculovirus P2 viral stock. (A) Cell counts for SINV Cp recombinant baculovirus infected and uninfected SF9 cells over 120 hours. Line indicates the average count over 8 squares of a haemocytometer, error bars indicated standard deviation over all counted squares. **(B)** Viability of cells over 120 hours calculated from averages of 8 counted squares.

7.2.6.4 Generation of recombinant SFV baculovirus

SFV bacmid transfection was performed as in 2.7.3.1. Figure 7.23 shows SF9 cells transfected with SFV bacmid alongside non-transfected cells after 96 hours. Cells in the transfected sample display hallmarks of infection including increased cell size, granular appearance and large nuclei.

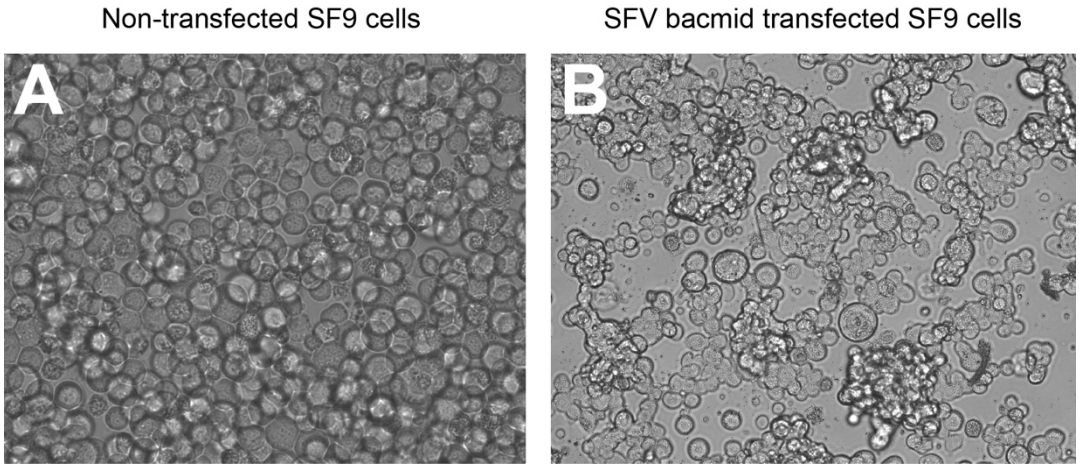


Figure 7.23: SFV bacmid transfection of SF9 cells. (A) Mock transfected SF9 insect cells. **(B**) SF9 cells transfected with SFV bacmid at 40X magnification

Media from SFV bacmid transfections (P1) was amplified to form a higher titre P2 viral stock to be used in downstream investigations. 30 mL of SF9 cells ($2x10^6$ cells/mL) were infected with SFV bacmid P1 viral stock at a presumed titre of $3x10^6$ pfu/mL (MOI = 0.1) and incubated at 28°C,120 rpm. Viability and cell numbers were monitored for both the SFV bacmid infected and mock infected control, until numbers and viability deviated significantly. Cell numbers for both the mock and the SFV infected cells were counted, with the line indicating the average over 8 counted squares of a haemocytometer and the error bars indicate the standard deviation of those 8 counted squares (Figure 7.24). Cell number of both the infected and the mock infected cells initially remains comparable until 72 hours where infected cell numbers drop significantly. Cell viability also changes with time, eventually dropping to ~70% viability at 120 hours.

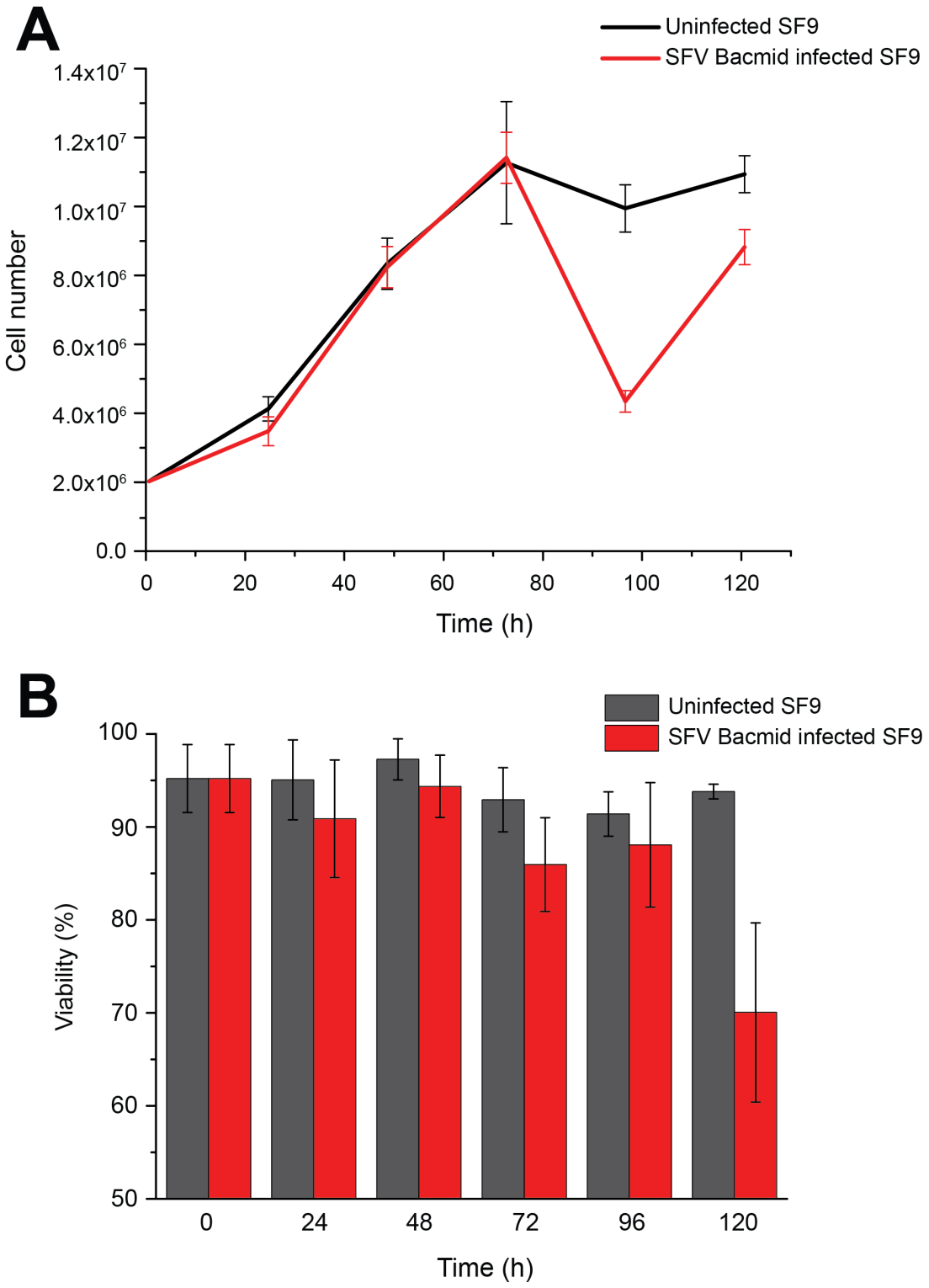


Figure 7.24: Generation of SFV Cp recombinant baculovirus P2 viral stock. (A) Cell counts for SFV baculovirus infected and uninfected SF9 cells over 120 hours. Line indicates the average count over 8 squares of a haemocytometer, error bars indicated standard deviation over all counted squares. **(B)** Viability of cells over 120 hours calculated from averages of 8 counted squares.

7.2.6.5 Generation of recombinant RRV baculovirus

Transfection of RRV bacmid was performed as described in 2.7.3.1. RRV bacmid transfected SF9 cells are shown in Figure 7.25 alongside mock transfected cells after 96 hours. As seen previously, there is clearing of the monolayer and cells appear oversized and granular.

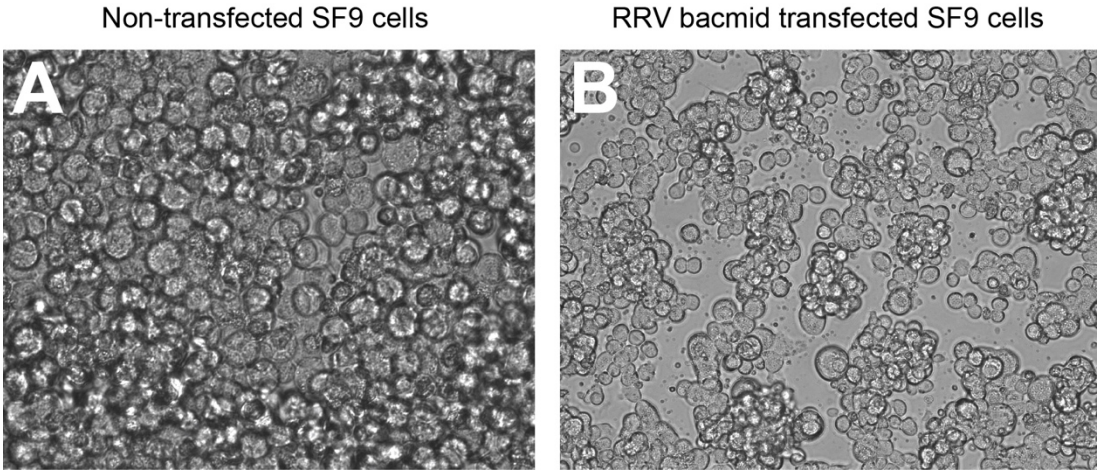


Figure 7.25: RRV bacmid transfection of SF9 cells. (A) Mock transfected SF9 insect cells at 40X magnification. **(B**) SF9 cells transfected with RRV bacmid at 40X magnification

Media from the transfected cells (P1 viral stock) was filtered and centrifuged prior to amplification to form a higher titre P2 viral stock. 30 mL of SF9 cells at a density of $2x10^6$ cells/mL were infected with the RRV bacmid P1 viral stock at an MOI of 0.1 (presumed viral titre $3x10^6$ pfu/mL) and incubated at 28°C,120 rpm. Cell number and viability were monitored every 24 hours for the infected sample and a mock infected control until numbers diverged significantly as shown in Figure 7.26, where the line indicates the average over 8 counted squares of a haemocytometer and the error bars indicate the standard deviation of those 8 counted squares. Both samples increase cell number in the initial 72 hours however, at 96 hours the cell number decreases and at 120 hours cell viability is ~75% compared to the control at ~95%.

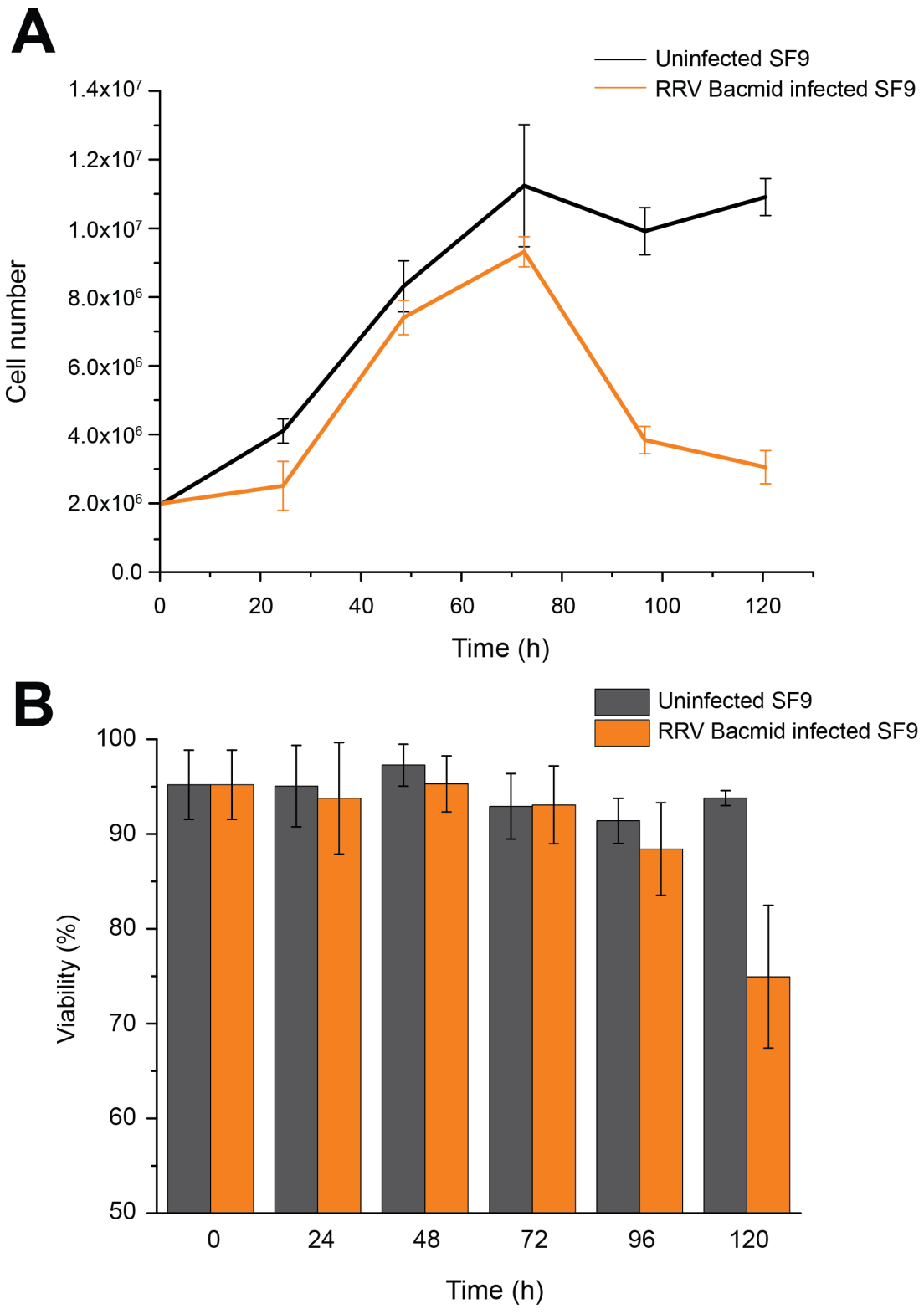


Figure 7.26: Generation of RRV Cp recombinant baculovirus P2 viral stock. (A) Cell counts for RRV bacmid infected and uninfected SF9 cells over 120 hours. Line indicates the average count over 8 squares of a haemocytometer, error bars indicated standard deviation over all counted squares. **(B)** Viability of cells over 120 hours calculated from averages of 8 counted squares.

7.2.7 Expression of recombinant CHIKV core protein

The optimal MOI varies for each individual baculovirus construct and therefore a framework of infection and optimal protein production needs to be established for each isolate.

7.2.7.1 Trial Expression of recombinant CHIKV core protein

After determining the viral titre via plaque assay, small-scale MOI investigations were performed in a 24 well plate format. $6x10^5$ SF9 cells were seeded out into each well of a 24 well plate with 300 μL of SF900II media (Gibco) and infected with P2 viral stock at varying MOIs (0, 1, 2, 5, 10 and 20). Cells and media were harvested at 144 hours and stored at -20°C before boiling with Laemmli buffer and analysis via SDS-PAGE and western blot with a CHIKV core specific antibody (Native Antigen company). SDS-PAGE and western blots of core protein expression at varying MOIs are shown in Figure 7.27. CHIKV core protein is seen at all MOI conditions excepting mock infected (MOI=0) and MOI=20 in both cells and media. Levels of protein expression in cells peaks at MOI=5 and there is no appreciable increase in expression at MOI=10 in cells.

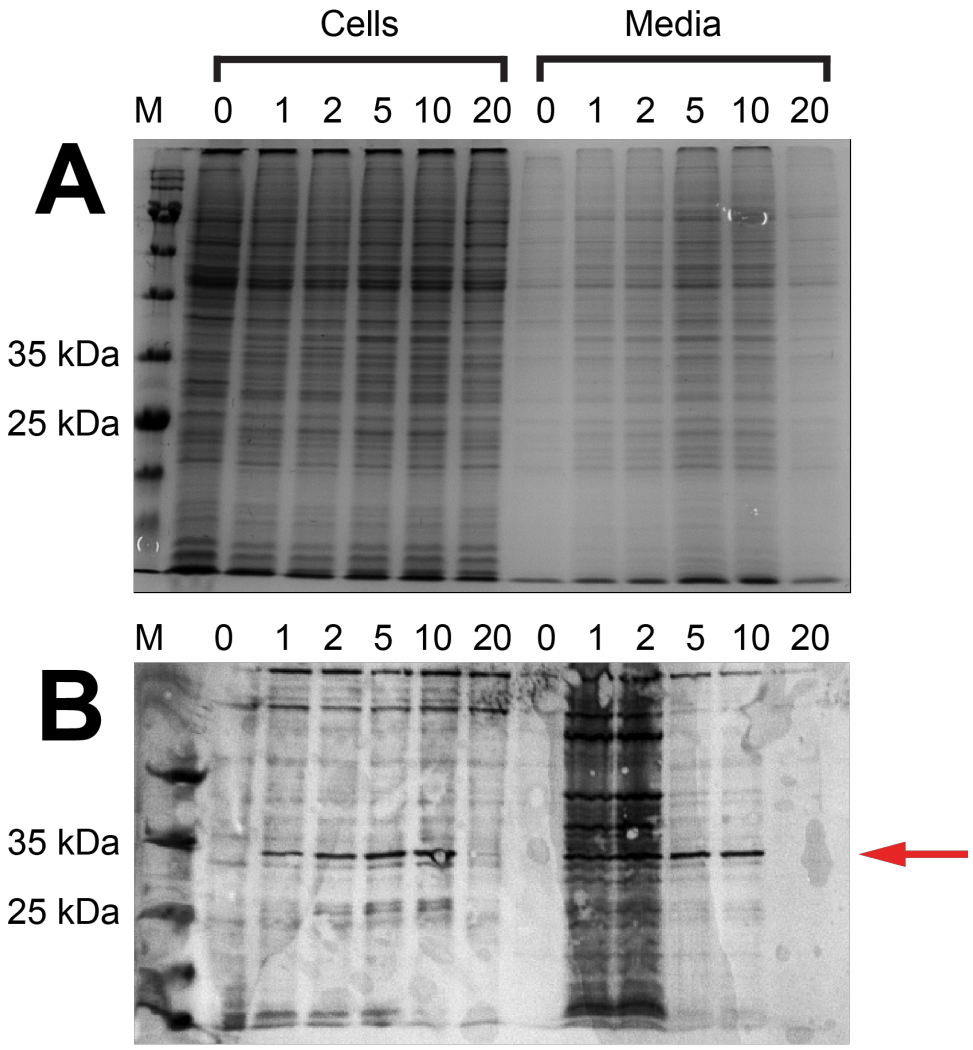


Figure 7.27: MOI investigations for optimal expression of CHIKV core protein. (A) SDS-PAGE of harvested SF9 cells and media after infection with baculovirus expressing CHIKV core protein. Cells were harvested at 144 hours post infection and separated from media by centrifugation. **(B)** Western blot of the same samples as in A stained with a CHIKV core protein antibody (native antigen company) and western blue. M = ladder, BLUeye (geneflow). Numbers indicate MOI.

An expression time-course was also performed in a 24 well plate seeded with $6x10^5$ cells/well. Cells were infected with MOI=2 and incubated at 28°C, and both cells and media were harvested at 24-hour intervals. Samples were stored at -20°C before boiling with Laemmli buffer and analysis via SDS-PAGE and western blot with a CHIKV core specific antibody. SDS-PAGE and western blots of infection timecourses are shown in Figure 7.28 and 7.29.

CHIKV core protein is visible in both harvested cells and media after approximately 96 and 120 hours respectively. Peak production in cells is seen at 168 hours post-infection. Comparatively, levels in the media peak at 144 hours after which levels appear to drop, as shown in Figure 7.29.

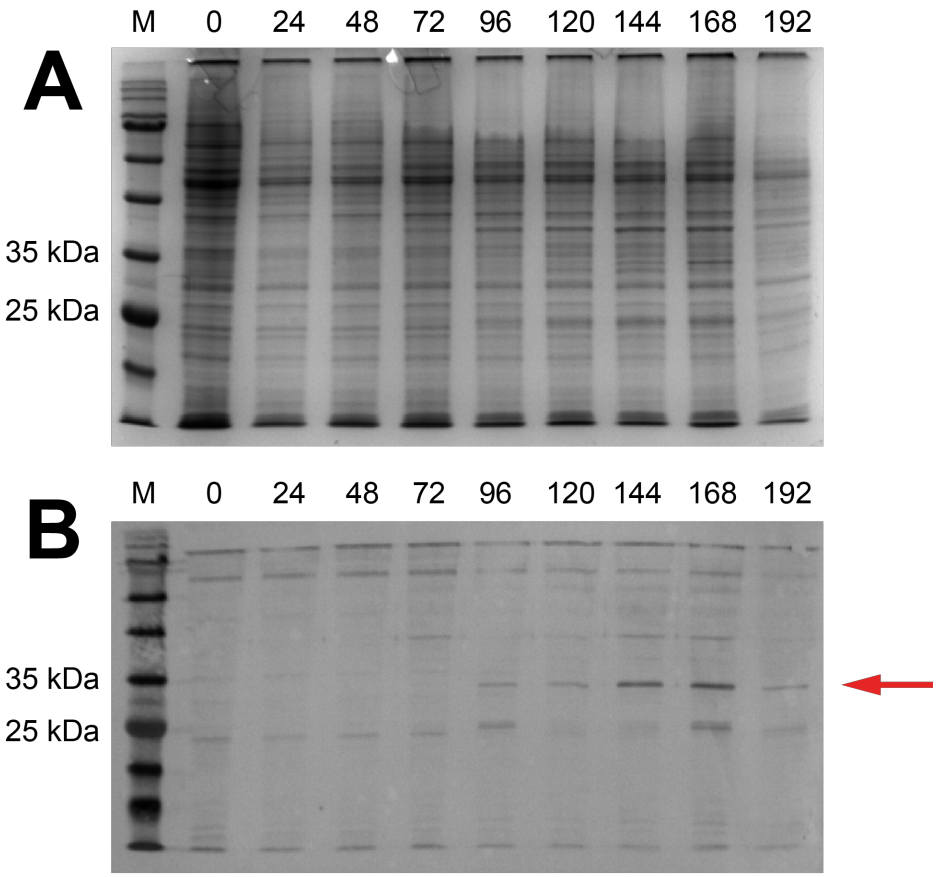


Figure 7.28: Timecourse for CHIKV core protein expression in cells. **(A)** SDS-PAGE of harvested SF9 cells after infection with baculovirus expressing CHIKV core protein. **(B)** Western blot of the same samples as in A stained with a CHIKV core protein antibody (Native antigen company) and western blue. M = ladder, BLUeye (geneflow). Numbers indicate hours post-infection.

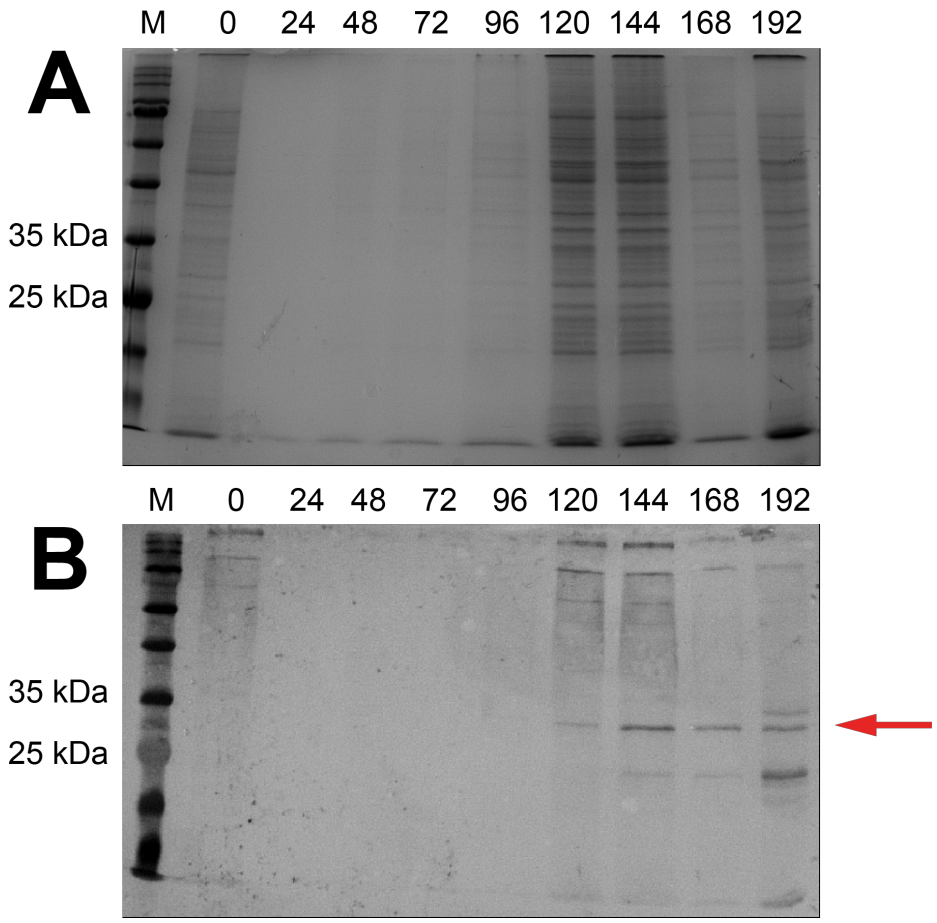


Figure 7.29: Timecourse for CHIKV core protein expression released into media. (A) SDS-PAGE of media from SF9 cells after infection with baculovirus expressing CHIKV core protein**. (B)** Western blot of the same samples as in A stained with a CHIKV core protein antibody (Native antigen company) and western blue. M = ladder, BLUeye (geneflow).

7.3 Discussion

The baculovirus expression system can be exploited to produce, by and large, any foreign gene. Since the expression of INF-β in 1983, followed quickly by *E. coli* β-galactosidase in 1984 (van Oers *et al.*, 2015), multiple proteins have been produced in this system including therapeutics and potential therapies some of which are shown in Table 7.1. Additionally, viral structural proteins expressed in this system have been observed to form VLPs which are typically immunogenic and therefore potential vaccine candidates (Contreras-Gómez *et al*., 2014). One of the most successful VLP vaccines licenced for use is the human papilloma virus (HPV) vaccine with the brand name Cervarix™ and many more are in development (Roldão *et al*., 2010).

The production of alphavirus core proteins in a baculovirus expression system began with the production of FastBac1 compatible inserts. The Cp sequences present in pET22b vectors from Chapter 6, were used as templates to amplify the Cp gene while adding regions of compatibility to the FastBac1 vector. Q5 PCR was performed with all 5 alphavirus core proteins (CHIKV, SINV, SFV, RRV and WEEV) producing products of approximately 850 bp, the expected size for correctly amplified targets with additional FastBac1 sequences. Linearised FastBac1 vector, combined with the purified inserts, were used in a HiFi DNA assembly reaction where an exonuclease creates complementary single stranded regions within the overlap which anneal. DNA polymerase extends the annealed fragments which are repaired by DNA ligase, resulting in a double stranded plasmid containing the gene of interest that is transformed into DH5α cells.

Table 7.1: Selection of VLP vaccines and therapies in preclinical development or licenced for use.

Product name	Expressed antigen	Purpose	Status
Porcilis Pesti	E2 glycoprotein	Subunit/marker vaccine against classical swine fever	Licenced
Circumvent PCV/ Porcilis PCV	Porcine circovirus ORF2	VLP vaccine against porcine circovirus type 2	Licenced
Cervarix	HPV L1 (serotypes 16 & 18)	VLP-based vaccine against cervical cancer	Licenced
CircoFLEX	Porcine circovirus ORF2	VLP vaccine against porcine circovirus type 2	Licenced
Provenge	PAP-GM-CSF	Immunotherapy against prostate cancer	Licenced
Glybera	AAV vector with lipoprotein lipase transgene	Gene therapy against familial lipoprotein lipase deficiency	Licensed
Flublok	Influenza HA	Annual trivalent flu vaccine	Licenced
rHEV	Hepatitis E ORF 2	VLP vaccine for hepatitis E capsid antigen	Phase II*
H1N1 2009 VLP	A/California/04/2009 H1N1 HA and NA A/Indonesia/05/2005 (H5N1) M1	VLP vaccine for Influenza A H1N1	Phase II
B19 VLP vaccine	VP1 and VP2 proteins	Parvovirus B19 VLP vaccine	Phase II*
EBOV VLP	Ebola VP40 and GP	VLP vaccine for Ebola virus	Preclinical
LR2006-OPY1	CHIKV structural polyprotein	VLP vaccine for CHIKV	Preclinical
HIV VLP	Pr55Gag, envelope VLP	VLP vaccine for HIV	Preclinical

Data taken from (Contreras-Gómez *et al*., 2014)

Single colonies were isolated and sequenced for the presence of insert which confirmed that the FastBac1 plasmids contained the desired gene of interest. These plasmids were transformed into DH10Bac cells, which contain a bacmid hosting a site for mini-attTn7. Multiple cloning site transposition occurs between the FastBac1 vector and the bacmid with the aid of a helper plasmid encoding for transposase. Successful transposition interrupts the LacZα gene and therefore insertion at this site disrupts expression allowing for selection of positive transformants when grown in the presence of Bluo-Gal. Additional selection was performed by re-streaking out white colonies to obtain a twice white phenotype. Confirmational PCR with M13 primers gives a product of ~3150 bp which corresponds to 2300 bp + ~850 bp, whereas an isolate that has not undergone transposition will produce a fragment of 300 bp. As seen in Figure 7.13 all sampled CHIKV Cp isolates contain a product of the correct size, however 10/12 contain a band at 1.4 Kb. Similarly, the SFV, SINV, RRV and WEEV Cp twice white isolates also contained at least one isolate of the correct size however, bands not corresponding to the gene of interest were observed in some cases. Isolates containing only one band of the correct size were carried forward.

PCR confirmed isolates were used to transfect SF9 insect cells alongside mock transfected lipofectamine only samples. In all cases, after 96 hours, SF9 cells show cytopathic effect consistent with transfection including clearance of the cell layer, enlarged cell size, granular appearance and a large nucleus (Figures 7.15, 7.18, 7.20, 7.22 and 7.24). The supernatant from transfected cells was used to infect larger flasks of SF9 cells to amplify the baculoviral

stock. This is a recommended step for the Bac-to-Bac protocol as the initial P1 viral stock is of low titre (1×10^6 to 1×10^7 pfu/mL) and on a small scale (~ 7 mL). Amplification to a P2 viral stock can achieve titres of 1×10^7 to 1×10^8 pfu/mL and a volume of ~30 mL which ensures that multiple investigations can be carried out with the same stock (Life Technologies, 2015). In all cases a mock infection was done alongside as a control, and cell counts, and viability were measured every 24 hours post infection. For all constructs after 96-120 hours cell viability diverged significantly from the control cells, dropping below 90% viability and in 4/5 instances below 80%. Additionally, cell numbers are reduced when compared to mock infected cells suggesting that cell lysis was occurring but not generation of new cells.

For CHIKV Cp, the P2 baculoviral viral stock titre was determined using a plaque assay. After 10 days of incubation at 28°C plaques were visible in the cell monolayer (Figure 7.17), and the viral titre was calculated at 4.5×10^8 pfu/mL. The P2 viral stock was then used in small scale protein expression trials to determine the optimal expression conditions for CHIKV core protein. MOI investigations show CHIKV core protein is detected in both cells and media after 144 hours at MOI = 1, 2, 5, and 10. No expression is seen in the non-induced control (Figure 7.26) or in the MOI = 20 sample.

The production of alphavirus core proteins in a baculovirus expression system in this thesis was undertaken due to the toxicity and co-elution issues discussed in Chapter 6. Out of the five alphavirus proteins chosen for expression only SFV and RRV Cp were able to be expressed and purified from

E.coli. It is hoped that the alphavirus core proteins will be better tolerated in an insect cell expression system. Due to time constraints, not all proteins were expressed in insect cells during this period of work and CHIKV Cp did not move beyond purification trials. In the future, expression of all the proteins will be investigated and scaled up before characterisation. The protein will also be used in reassembly assays and for SELEX to investigate packaging signals. The work in this chapter begins laying the groundwork for alphavirus core protein production for the investigation of packaging signal mediated assembly. Understanding mechanisms of viral assembly could lead to the generation of more stable, non-pathogenic viruses that could be used as synthetic vaccines or drug/gene delivery vehicles.

[This page has been left intentionally blank]

Chapter 8: General Discussion

[This page has been left intentionally blank]

8.1 General discussion

8.1.1 Genome packaging in ssRNA viruses

The assembly of a viral capsid is an essential step in a viral lifecycle. Regardless of morphology, the role of this capsid remains the same; to safeguard the genome and deliver it to a target cell (Prasad and Schmid, 2012). Different viral families employ different assembly and genome packaging mechanisms (Rao, 2006; Sun *et al*., 2010). The focus of this thesis has been the assembly of ss(+)RNA viruses, the largest viral group encompassing some of the most threatening human and animal pathogens (Francki *et al*., 1991). PS-mediated assembly is a recently discovered assembly mechanism that depends on multiple sequence-specific CP-RNA interactions at sites distributed throughout the genome leading to highly co-operative assembly along a defined pathway (Stockley *et al*., 2013; Stockley and Twarock., 2019). This mechanism raises two questions. What is the nature and consequence of the RNA-CP interactions, and given their importance why are structured RNAs seen so rarely in viral structures?

The interactions between RNA and CP have been thought to be dominated by electrostatic contacts, in which positively-charged areas of the CPs interact with the negatively-charged phosphodiester backbone of RNA (Bancroft *et al*., 1967; Hiebert and Bancroft, 1969; Belyi and Muthukumar, 2006; Forrey and Muthukumar, 2009; Devkota *et al*., 2009; Hagan, 2009). However, the observed selectivity of genome packaging *in vivo* suggests specificity in this process (Routh, Domitrovic and Johnson, 2012a). Packaging signals or origin

of assembly sequences are not a new concept. The first RNA packaging signal in a eukaryotic virus was characterised in turnip crinkle virus (TCV) in 1997 (Qu and Morris, 1997). Sequences necessary for assembly/genome packaging have since been identified in many other viruses (Hayashi *et al.*, 1992; Frolova *et al*., 1997; Narayanan and Makino, 2001), including pre-genomic RNA of DNA viruses (Kawamoto *et al.*, 1994). It was thought that a genome contains only one such site and the initial RNA-CP interaction facilitates further CP binding. However, recent investigations in our laboratory using a combination of SELEX, CLIP-SEQ, mass spectrometry, single molecule spectroscopy, electron microscopy and tomography, and mathematical modelling has identified many regions with CP affinity dispersed throughout viral genomes (Patel *et al.*, 2017; Shakeel *et al.*, 2017; Rolfsson *et al*., 2016; Bunka *et al.*, 2011; Patel *et al*., 2015; Stewart *et al*., 2016). Further detailed mechanistic investigation has uncovered the roles of these CP-RNA contacts during assembly.

Recent asymmetric viral reconstructions have transformed our view of multiple RNA-CP contacts within viral capsids (Koning *et al*., 2016; Dai *et al.*, 2017; Gorzelnik *et al*., 2016). Two recent structures of bacteriophage MS2 both contain significant RNA density in the capsid interior. The two models, from 2016 and 2017 assigned 95 and ~80% of the genome respectively, and lie within the capsid in a non-uniform distribution (Koning *et al.*, 2016; Dai *et al*., 2017). Unlike previous symmetry-averaged structures which do not reveal RNA densities (Valegård *et al.*, 1994), the RNA forms defined structures which

do not conform to icosahedral symmetry and extend to the interior of the capsid. The 2017 structure identified 15 SLs, all of which were previously predicted as packaging signals by the Stockley laboratory and collaborators (Dykeman, Stockley and Twarock, 2013b; Rolfsson *et al*., 2016).

8.1.2 Probing the details of a PS-mediated assembly mechanism: STNV

One of the themes explored in this thesis is the function of packaging signals and their consequences. Previous investigations by the Stockley laboratory have identified and probed the mechanism of PS-mediated assembly in STNV (Bunka *et al*., 2011; Ford *et al.*, 2013; Borodavka *et al*., 2013; Patel *et al*., 2015). The loop motif in its SL PSs, AXXA, was shown to be important for assembly as variants lacking this motif are unable to reassemble into VLPs (Ford *et al*., 2013; Patel *et al*., 2015). Multiple PSs in longer fragments act co-operatively to condense the RNA suggesting that they make assembly more efficient. These investigations also revealed that the highest affinity PS (PS3), can trigger VLP formation by itself, but was more efficient when oriented with respect to neighbouring PS sites (Patel *et al*., 2015). The PS-mediated assembly model requires a hierarchy of CP affinities among the PSs (Dykeman *et al.*, 2014). I examined the relationships between PSs given the hierarchy of affinities and their dispersal throughout the genome. Understanding the features of CP recognition in molecular detail was the reason for the work detailed in Chapter 4. Understanding the global collective action of multiple PSs was the motivation for the work detailed in Chapter 5.

Chapter 4 describes the interrogation of the CP recognition motif AXXA. SLs were produced based on the highest affinity B3, in which the 'inner' or 'outer' nucleotides of the motif were varied to probe their fitness as assembly substrates. Previously identified STNV PSs appear to have no consensus for the central nucleotides (Bunka *et al*., 2011). As expected, no base specificity was found at these positions. However, the 'outer' nucleotide variants show differing behaviour. Using svAUC to characterise VLPs reassembled around SL variants, the AUUU, UUUA, and AUUA loop variants produce a VLP peak similar to PS3, suggesting they all encompass functional recognition motifs (4.2.2.2). An adapted smFCS assay was then used to examine the relationship between loop sequence and CP affinity. In this assay the AXXA variant performs the best indicating that this motif is optimal for CP recognition and therefore VLP assembly (4.2.3). Interestingly both GUUA and UUUA variants outperform those containing a 5′ A suggesting that the 3′ A is more important for CP recognition.

One of the problems for PS-mediated assembly is the fleeting nature of the recognition elements. In single molecule assays, B3 forms a capsomer with nanomolar affinity. However, the results in Chapter 4 suggest that only one A nucleotide is critical in this recognition. How can such high affinity be extracted from just one nucleotide? One potential reason for its high affinity is that the initial CP binding event triggers formation of a trimeric capsomer i.e. the RNA-CP is stabilised by multiple CP-CP contacts.

Chapter 5 explores the collective action of multiple PSs dispersed throughout a viral genome. Genomic length variants were designed to investigate the effects on assembly in a variant construct lacking PS motifs, and probe the importance of the 5′ 127 nt sequence. In competitive reassembly assays, constructs containing AXXA SLs are packaged preferentially over their non-AXXA containing counterparts (5.2.3.2). If the assembly mechanism were solely due to electrostatics there would be no selectivity in this process and both RNAs would be equally encapsidated. These results show that the PSs are important for assembly but are they required throughout the genome?

The 5′ 127 nt of STNV contains what is thought to be the highest affinity PS (PS3). This region is also densely populated with AXXA loops when contrasted with the 3′ of the genome. One variant was designed to test whether this 5′ region of high affinity was sufficient to package a full genome. Results suggest that the 5′ 127 nt of STNV is not enough to restore packaging to WT levels under experimental conditions (5.2.3.2). This indicates that other PSs distributed throughout the genome are important in this mechanism to ensure complete capsid formation.

Multiple packaging signals within viral systems are thought to be beneficial from an evolutionary perspective. RNA viruses have a high mutation rate, estimated at 10^{-4} per base per replication cycle (Belshaw *et al.*, 2007). Therefore, the presence of only one assembly initiation site could be

vulnerable to mutation. Additionally, in three different systems studied in our laboratory, at least one of the identified PSs encodes regions of the PS binding site on the CP (Patel et al., 2015; Shakeel *et al*., 2017). Loss of the PS would also alter the sequence of the CP-PS binding site resulting in severe loss of fitness. This has a consequence of promoting assembly of RNA genomes that contain PSs. The multiple functions of RNA also protect the high affinity sites in the 5′ 127 nt. This region acts as an initiation complex for assembly (Patel *et al*., 2015), but also has other functions. Sequences within this region form a translational enhancer with sequences in the 3′ untranslated region (UTR) (Kaempfer *et al*., 1981).

The results of the titration vs one-shot assembly experiments (5.2.3.3), also highlight the roles of multiple PSs of varying affinities within a system. CP added in a series of defined steps results in a higher proportion of correctly formed species in comparison to a one-step addition. The varying affinities ensure the CP binds in a specific order leading to nucleation at specific sites. This in turn directs assembly down a small number of distinct assembly pathways and regulates the assembly process. One-step addition in contrast allows multiple assembly initiation points to occur on the RNA trapping it in unresolvable complexes.

As noted in Chapter 5, the WT reassembly assay is not 100% efficient (5.2.3.1). This suggests that the experimental conditions do not reflect those

in vivo. The natural mechanism of assembly could be part of a co-transcriptional process where PSs form on nascent chains and are therefore not available all at once preventing this problem. Such a mechanism has been suggested for other ss(+)RNA viruses such as the HPeV 'assemblysome' mechanism. In this instance the virally encoded protein 2C is thought to act as a chaperone facilitating PS folding and pentameric capsomer binding as the RNA emerges from the replicase (Shakeel *et al.*, 2017). The coupling of replication and encapsidation has also been suggested in other picornaviruses (Nugent *et al*., 1999) and plant viruses. In brome mosaic virus (BMV) it has been shown that sites of viral replication and CP synthesis overlap in large endoplasmic reticulum (ER) derived vesicles (Bamunusinghe *et al*., 2011). By coupling transcription and encapsidation, local secondary structures such as PSs on nascent chains could bind CP in a way that may not be possible on complete transcripts. The results in Chapter 5 suggest that *in vivo* STNV could assemble similarly. However, such a closed system 'assemblysome' is difficult to reproduce *in vitro*.

Overall, the work in this thesis suggests that packaging signals in STNV contribute to overall efficiency in encapsidation and highlight the importance of the AXXA motif, specifically the 3′ A. Furthermore, titration of CP results in a higher proportion of correctly formed species which is consistent with previous modelling of this mechanism (Dykeman *et al*., 2014).

8.1.3 Applications of understanding PS-mediated assembly

Understanding assembly mechanisms in detail and the roles of RNA packaging signals presents a novel antiviral target. For example, SELEX identified sites within the HBV pre-genome which in smFCS assays behave as assembly initiators, triggering the formation of *T*=4 viral cores, i.e. they are likely PSs. This investigation led to a proposed model for nucleocapsid assembly based on sequence-specific charge neutralisation of residues on the C-terminal argenine-rich domain (ARD) (Patel *et al*., 2017).

The principal HBV PS, PS1, was used as a probe against an immobilised library of over 22 thousand small molecular weight ligands designed to have RNA binding properties (Connelly *et al*., 2017). This identified 69 compounds that bound to PS1 specifically. These were then used in the Dorner laboratory at Imperial College to investigate their effects on HBV replication in cell and organoid culture. Figure 8.1 shows the results of these assays in HepG2 cells. Several of the new ligands (pink bars) are significantly more inhibitory of assembly than the current clinically used drugs targeted at the polymerase (Lamivudine and Tenofovir). These ligands are also much more potent at reducing the levels of covalently closed circular DNA (cccDNA), which is the basis of chronic infection in HBV (Peter Stockley, personal communication).

These results are proof of principle that novel antiviral drugs can emerge from these fundamental studies. Therefore, understanding a viral assembly mechanism has led to the development of novel lead compounds for a human disease.

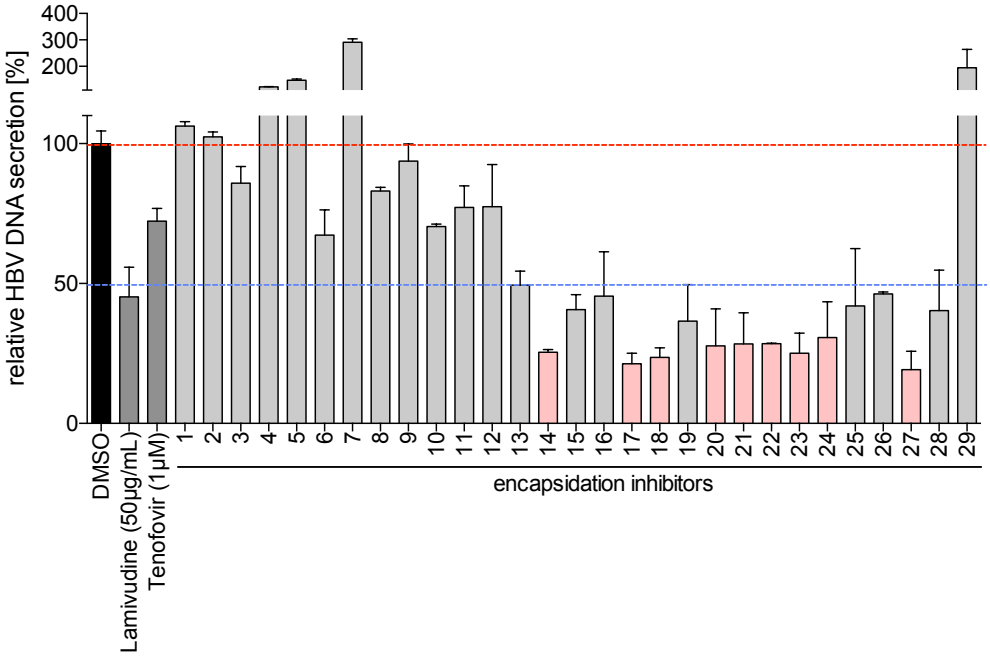


Figure 8.1: PS1 binding ligands inhibit HBV replication. Relative HBV DNA secretion of samples treated with encapsidation inhibitors 1-29 and clinically available drugs Lamivudine and Tenofovir. Several of the ligands from small molecule screening are more inhibitory than commercially available treatments.

Previous identifications of PSs have used SELEX to determine sequences (aptamers) that bind to the interior surface of CPs. These are then compared to genomic sequences to identify putative PSs (Bunka *et al*., 2011; Shakeel *et al*., 2017; Stewart *et al*., 2016; Patel *et al*., 2017). This approach requires the production and/or purification of viral capsid/core proteins. This was the aim of the work described in Chapters 6 and 7 on alphavirus Cps.

Chapter 6 describes the expression and purification of alphavirus core proteins in *E.coli*. While two of these proteins were found to be toxic (WEEV Cp and SINV Cp), three proteins were expressed and solubilised (CHIKV Cp, RRV Cp and SFV Cp). All three proteins co-eluted with a ~25 kDa protein, identified by

mass spectrometry as the 30S ribosomal subunit, S4. Both SFV and RRV core proteins were separated sufficiently from the S4 subunit to continue with further investigations (6.2.3.5). However, S4 and CHIKV Cp could not be easily or efficiently separated and purification was not attempted on a large scale from *E.coli*. SFV and RRV core proteins were used in functional assays to assess their ability to form core-like particles (6.2.4.1). As in Tellinghuisen *et al*. purified protein was reassembled with a non-specific DNA fragment forming CLPs of the expected size. As no higher order structures were observed in the Cp only control, formation of CLPs seems to be dependent on the presence of nucleic acid which is consistent with previous investigations (Tellinghuisen *et al*., 1999) and svAUC results. When the same reassembly is performed at single molecule level however, there is no change in hydrodynamic radius implying that at these concentrations nucleic acid specificity is important (6.2.4.2), this is the first single molecule assay of an alphavirus in the literature (Mendes and Kuhn, 2018). These results suggest that while high concentration can be used to induce particle formation without selectivity, at single molecule levels this process is more discriminating.

Expression of alphavirus core proteins was also achieved in insect cells, as described in Chapter 7. Due to time constraints expression conditions were only investigated for CHIKV Cp and large-scale protein production was not attempted. All five core protein genes were integrated into bacmids (7.2.3) and used to transfect insect cells (7.2.4). Preliminary western blots indicate that the CHIKV Cp is expressed (7.2.5) and work will continue on the expression of alphavirus core proteins in insect cells for use in future packaging signal

investigations. Work in both Chapters 6 and 7 paves the way for more in-depth analysis of the alphavirus assembly mechanism and identification of PSs in the alphavirus genus.

8.2 Future Work

Understanding mechanisms of viral assembly could lead to the generation of more stable, non-pathogenic organisms that can be used as synthetic vaccines or drug/gene delivery vehicles. Additionally, as shown with HBV, identification of PSs can lead to the development of novel antivirals. This is especially relevant for alphaviruses where no treatment currently exists other than management of symptoms.

Chapter 5 describes preliminary investigations into packaging of genomic length substrates. This work will be repeated, using a fluorimeter to measure fluorescence of peaks obtained from SEC and using an RI detector as the concentration source to avoid RNA/protein ambiguity in absorbance traces. This was unavailable during the work carried out in this thesis. Additionally, competitions will also be carried out with different buffer, RNA preparation conditions and new genomic length RNA variants to further probe packaging on this scale. Furthermore, during these repeated experiments further controls will be undertaken to assess the effect of post assembly interval and concentration on the samples.

In order to investigate packaging signals in alphavirus assembly, SELEX will be performed on reassembled SFV and RRV core like particles, as performed previously for HBV. This could identify areas of the genome with affinity for core protein which could subsequently be used in mutagenesis and single molecule experiments to define sequence specificity. As with HBV, potential packaging signals could be used as targets for novel antiviral drug investigations.

Chapter 7 describes the production of alphavirus core proteins in insect cells. Investigation into protein production will continue with a view of performing SELEX and downstream investigations with proteins that could not be expressed in *E.coli*.

An alternative approach for identifying RNA-CP interactions *in virio* has recently been developed by the Stockley laboratory. X-ray synchrotron footprinting is used to determine secondary structures for RNA-CP contacts within virions. It would thus be possible to use intact virions for such analysis provided that the identification of cleavage sites was carried out using next generation sequencing each site being a dip in the read count at each nucleotide.

Chapter 9: References

[This page has been left intentionally blank]

Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D. and Tsukihara, T. 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature*. **286**(5768),pp.33–39.

Adouchief, S., Smura, T., Sane, J., Vapalahti, O. and Kurkela, S. 2016. Sindbis virus as a human pathogen-epidemiology, clinical picture and pathogenesis. *Reviews in medical virology*. **26**(4),pp.221–241.

Adrian, M., Dubochet, J., Lepault, J. and McDowall, A.W. 1984. Cryo-electron microscopy of viruses. *Nature*. **308**(5954),pp.32–36.

Agilent Technologies 2015. *BL21-CodonPlus Competent Cells* [Online] 0 ed. Available from: https://www.agilent.com/cs/library/usermanuals/public/230240.pdf.

Annamalai, P., Apte, S., Wilkens, S. and Rao, A.L.N. 2005. Deletion of highly conserved arginine-rich RNA binding motif in cowpea chlorotic mottle virus capsid protein results in virion structural alterations and RNA packaging constraints. *Journal of virology*. **79**(6),pp.3277–3288.

Atkins, G.J. 2013. The Pathogenesis of Alphaviruses. *ISRN Virology*. 2013(3),pp.1–22.

Baltimore, D. 1971. Expression of Animal Virus Genomes. *Bacteriological Reviews*. **35**(3),pp.235–241.

Bamunusinghe, D., Seo, J.-K. and Rao, A.L.N. 2011. Subcellular localization and rearrangement of endoplasmic reticulum by Brome mosaic virus capsid protein. *Journal of virology*. **85**(6),pp.2953–2963.

Bancroft, J.B., Hiebert, E. and Bracker, C.E. 1969. The effects of various polyanions on shell formation of some spherical viruses. *Virology*. **39**(4),pp.924–930.

Bancroft, J.B., Hills, G.J. and Markham, R. 1967. A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits in vitro. *Virology*. **31**(2),pp.354–379.

Bandyopadhyay, A.S., Garon, J., Seib, K. and Orenstein, W.A. 2015. Polio vaccination: past, present and future. *Future microbiology*. **10**(5),pp.791–808.

Basnak, G., Morton, V.L., Rolfsson, O., Stonehouse, N.J., Ashcroft, A.E. and Stockley, P.G. 2010. Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid. *Journal of molecular biology*. **395**(5),pp.924–936.

Beijerinck, M.W. 1891. Ueber ein contagium vivum fluidum als Ursache der Flecken-krankheit der Tabaksblatter. *Verhandelingen der Koninklyke akademie van Wetten-schappen te Amsterdam*. (65),pp.3–21.

Belshaw, R., Pybus, O.G. and Rambaut, A. 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome research*. **17**(10),pp.1496–1504.

Belyi, V.A. and Muthukumar, M. 2006. Electrostatic origin of the genome packing in viruses. *Proceedings of the National Academy of Sciences of the United States of America*. **103**(46),pp.17174–17178.

Bentley, G.A., Lewit-Bentley, A., Liljas, L., Skoglund, U., Roth, M. and Unge, T. 1987. Structure of RNA in satellite tobacco necrosis virus. A low resolution neutron diffraction study using 1H2O/2H2O solvent contrast variation. *Journal of molecular biology*. **194**(1),pp.129–141.

Bergren, N.A., Auguste, A.J., Forrester, N.L., Negi, S.S., Braun, W.A. and Weaver, S.C. 2014. Western equine encephalitis virus: evolutionary analysis of a declining alphavirus based on complete genome sequences. *Journal of virology*. **88**(16),pp.9260–9267.

Bernal, J.D. and Fankuchen, I. 1941a. X-ray and crystallographic studies of plant virus preparations : I. Introduction and preparation of specimens ii. Modes of aggregation of the virus particles. *The Journal of general physiology*. **25**(1),pp.111–146.

Bernal, J.D. and Fankuchen, I. 1941b. X-ray and crystallographic studies of plant virus preparations. III. *The Journal of general physiology*. **25**(1),pp.147–165.

Blow, D.M. 2006. Noncrystallographic symmetry *In*: M. G. Rossmann and E. Arnold, eds. *International Tables for Crystallography Volume F:*. International Union of Crystallography.

Borodavka, A. 2012. Understanding the dynamics of viral RNA genomes using single-molecule fluorescence. Doctor of Philosophy. University of Leeds. Leeds, UK.

Borodavka, A., Tuma, R. and Stockley, P.G. 2013. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA biology*. **10**(4),pp.481–489.

Borodavka, A., Tuma, R. and Stockley, P.G. 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(39),pp.15769–15774.

Borth, W. 1992. Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **6**(15),pp.3345–3353.

Bredenbeek, P.J., Frolov, I., Rice, C.M. and Schlesinger, S. 1993. Sindbis virus expression vectors: packaging of RNA replicons by using defective helper RNAs. *Journal of virology*. **67**(11),pp.6439–6446.

Bringloe, D.H., Gultyaev, A.P., Pelpel, M., Pleij, C.W. and Coutts, R.H. 1998. The nucleotide sequence of satellite tobacco necrosis virus strain C and helper-assisted replication of wild type and mutant clones of the virus. *Journal of General Virology*. **79 (Pt 6)**(6),pp.1539–1546.

Briolant, S., Garin, D., Scaramozzino, N., Jouan, A. and Crance, J.M. 2004. In vitro inhibition of Chikungunya and Semliki Forest viruses replication by antiviral compounds: synergistic effect of interferon-alpha and ribavirin combination. *Antiviral research*. **61**(2),pp.111–117.

Bunka, D.H.J., Lane, S.W., Lane, C.L., Dykeman, E.C., Ford, R.J., Barker, A.M., Twarock, R., Phillips, S.E.V. and Stockley, P.G. 2011. Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. *Journal of molecular biology*. **413**(1),pp.51–65.

Cann, A.J. 2012. Genomes *In*: *Principles of Molecular Virology*. Elsevier, pp. 55–101.

Cann, A.J. 2016. *Principles of Molecular Virology*. 6th Ed. Elsevier.

Caspar, D.L. and Klug, A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology*. **27**,pp.1–24.

Caspar, D.L.D. 1956. Structure of Bushy Stunt Virus. *Nature*. **177**(4506),pp.475–476.

Chai, C., Xie, Z. and Grotewold, E. 2011. SELEX (Systematic Evolution of Ligands by EXponential Enrichment), as a powerful tool for deciphering the protein-DNA interaction space. *Methods in molecular biology (Clifton, N.J.)*. **754** (Chapter 14),pp.249–258.

Chambers, A.C., Aksular, M., Graves, L.P., Irons, S.L., Possee, R.D. and King, L.A. 2018. Overview of the Baculovirus Expression System. *Current protocols in protein science*. **91**(4),pp.5.4.1–5.4.6.

Changchien, L.M. and Craven, G.R. 1986. The use of hydroxylamine cleavage to produce a fragment of ribosomal protein S4 which retains the capacity to specifically bind 16S ribosomal RNA. *Nucleic acids research*. **14**(5),pp.1957–1966.

Chen, Chao, Daniel, M.-C., Quinkert, Z.T., De, M., Stein, B., Bowman, V.D., Chipman, P.R., Rotello, V.M., Kao, C.C. and Dragnea, B. 2006. Nanoparticle-Templated Assembly of Viral Protein Cages. *Nano Letters*. **6**(4),pp.611–615.

Chen, Chuming, Huang, H. and Wu, C.H. 2017. Protein Bioinformatics Databases and Resources *In*: *Protein Bioinformatics*. Methods in Molecular Biology. New York, NY: Humana Press, New York, NY, pp. 3–39.

Chen, Z.G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G. and Johnson, J.E. 1989. Protein-RNA interactions in an icosahedral virus at 3.0 A resolution. *Science (New York, N.Y.)*. **245**(4914),pp.154–159.

Cheng, R.H., Kuhn, R.J., Olson, N.H., Rossmann, M.G., Choi, H.K., Smith, T.J. and Baker, T.S. 1995. Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell*. **80**(4),pp.621–630.

Cheng, Y. 2015. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*. **161**(3),pp.450–457.

Choi, H.K., Lu, G., Lee, S., Wengler, G. and Rossmann, M.G. 1997. Structure of Semliki Forest virus core protein. *Proteins*. **27**(3),pp.345–359.

Clokie, M.R.J., Millard, A.D., Letarov, A.V. and Heaphy, S. 2014. Phages in nature. *Bacteriophage*. **1**(1),pp.31–45.

Communicable Diseases Branch, Department of Health, Government of the Commonwealth of Australia 2018. National Notifiable Diseases Surveillance System. Available from: http://www9.health.gov.au/cda/source/cda-index.cfm.

Connelly, C.M., Abulwerdi, F.A. and Schneekloth, J.S. 2017. Discovery of RNA Binding Small Molecules Using Small Molecule Microarrays. *Methods in molecular biology (Clifton, N.J.)*. **1518**(4),pp.157–175.

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E. and Chisti, Y. 2014. Protein production using the baculovirus-insect cell expression system. *Biotechnology progress*. **30**(1),pp.1–18.

Crick, F.H. and Watson, J.D. 1956. Structure of small viruses. *Nature*. **177**(4506),pp.473–475.

Crick, F.H.C. and Watson, J.D. 1956. Structure of Small Viruses. *Nature*. **177**(4506),pp.473–475.

Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J. and Klug, A. 1970. Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature*. **226**(5244),pp.421–425.

Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z.H. and Sun, R. 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature*. **541**(7635),pp.112–116.

Danthinne, X., Seurinck, J., Van Montagu, M., Pleij, C.W. and van Emmelo, J. 1991. Structural similarities between the RNAs of two satellites of tobacco necrosis virus. *Virology*. **185**(2),pp.605–614.

Dent, K.C., Thompson, R., Barker, A.M., Hiscox, J.A., Barr, J.N., Stockley, P.G. and Ranson, N.A. 2013. The asymmetric structure of an icosahedral

virus bound to its receptor suggests a mechanism for genome release. *Structure (London, England : 1993)*. **21**(7),pp.1225–1234.

Department for Environment, F. 2002. Defra, UK - Science Search. Available from: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=9781.

Devkota, B., Petrov, A.S., Lemieux, S., Boz, M.B., Tang, L., Schneemann, A., Johnson, J.E. and Harvey, S.C. 2009. Structural and electrostatic characterization of Pariacoto virus: Implications for viral assembly. *Biopolymers*. **91**(7),pp.530–538.

Doeven, M.K., Folgering, J.H.A., Krasnikov, V., Geertsma, E.R., van den Bogaart, G. and Poolman, B. 2005. Distribution, lateral mobility and function of membrane proteins incorporated into giant unilamellar vesicles. *Biophysical journal*. **88**(2),pp.1134–1142.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W. and Schultz, P. 1988. Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics*. **21**(2),pp.129–228.

Duong-Ly, K.C. and Gabelli, S.B. 2014. Salting out of proteins using ammonium sulfate precipitation. *Methods in enzymology*. **541**,pp.85–94.

Dykeman, E.C., Stockley, P.G. and Twarock, R. 2013a. Building a viral capsid in the presence of genomic RNA. *Physical review. E, Statistical, nonlinear, and soft matter physics*. **87**(2),p.022717.

Dykeman, E.C., Stockley, P.G. and Twarock, R. 2013b. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of molecular biology*. **425**(17),pp.3235–3249.

Dykeman, E.C., Stockley, P.G. and Twarock, R. 2014. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(14),pp.5361–5366.

Ellis, R. 1991. Molecular Chaperones. *Annual Review of Biochemistry*. **60**(1),pp.321–347.

Ellis, R.J. 2006. Molecular chaperones: assisting assembly in addition to folding. *Trends in Biochemical Sciences*. **31**(7),pp.395–401.

European Centre for Disease Prevention and Control 2017. *Clusters of autochthonous chikungunya cases in Italy* [Online]. Stockholm: ECDC. Available from: https://ecdc.europa.eu/sites/portal/files/documents/RRA-chikungunya-Italy-update-9-Oct-2017.pdf.

European Centre for Disease Prevention and Control 2018. Factsheet about chikungunya. Available from: https://ecdc.europa.eu/en/chikungunya/facts/factsheet.

Famulare, M., Chang, S., Iber, J., Zhao, K., Adeniji, J.A., Bukbuk, D., Baba, M., Behrend, M., Burns, C.C. and Oberste, M.S. 2016. Sabin Vaccine Reversion in the Field: a Comprehensive Analysis of Sabin-Like Poliovirus Isolates in Nigeria. R. M. Sandri-Goldin, ed. *Journal of virology*. **90**(1),pp.317–331.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. and Ysebaert, M. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*. **260**(5551),pp.500–507.

Fitzpatrick, J.A.J. and Lillemeier, B.F. 2011. Fluorescence correlation spectroscopy: linking molecular dynamics to biological function in vitro and in situ. *Current Opinion in Structural Biology*. **21**(5),pp.650–660.

Ford, R.J. 2012. The roles of RNA in the assembly and diassembly of single stranded RNA icosahedral viruses. Doctor of Philosophy. University of Leeds. Leeds, UK.

Ford, R.J., Barker, A.M., Bakker, S.E., Coutts, R.H., Ranson, N.A., Phillips, S.E.V., Pearson, A.R. and Stockley, P.G. 2013. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *Journal of molecular biology*. **425**(6),pp.1050–1064.

Forrey, C. and Muthukumar, M. 2009. Electrostatics of capsid-induced viral RNA organization. *The Journal of Chemical Physics*. **131**(10),p.105101.

Fouts, D.E., True, H.L. and Celander, D.W. 1997. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. *Nucleic acids research*. **25**(22),pp.4464–4473.

Fraenkel-Conrat, H and Williams, R.C. 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. Proc Natl Acad Sci U S A. **41**(10), pp.690–698.

Francki, R.I.B., Fauquet, C.M., Knudson, D.L. and Brown, F. (eds.). 1991. *Classification and Nomenclature of Viruses*. Vienna: Springer Vienna.

Freundlich, H. 1940. The Ultracentrifuge. By The Svedberg and Kai O. Pedersen. *The Journal of Physical Chemistry*. **44**(7),pp.952–953.

Frolova, E., Frolov, I. and Schlesinger, S. 1997. Packaging signals in alphaviruses. *Journal of virology*. **71**(1),pp.248–258.

GE Healthcare 2014 UNICORN 7 - GE Healthcare Life Sciences. Available from:

https://www.gelifesciences.com/en/gb/shop/chromatography/software/unicorn-7-p-05649.

Garmann, R.F., Comas-Garcia, M., Koay, M.S.T., Cornelissen, J.J.L.M., Knobler, C.M. and Gelbart, W.M. 2014. Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *Journal of virology*. **88**(18),pp.10472–10479.

Garmann, R.F., Sportsman, R., Beren, C., Manoharan, V.N., Knobler, C.M. and Gelbart, W.M. 2015. A Simple RNA-DNA Scaffold Templates the Assembly of Monofunctional Virus-Like Particles. *Journal of the American Chemical Society*. **137**(24),pp.7584–7587.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server *In*: *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H. and Ferrin, T.E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein science : a publication of the Protein Society*. **27**(1),pp.14–25.

GoldBiotechnology 2015. *Aprotinin Stock Solution* [Online]. Available from: https://www.goldbio.com/documents/3595/Aprotinin+Stock+Solution.pdf.

Goldsmith, C.S. and Miller, S.E. 2009. Modern uses of electron microscopy for detection of viruses. *Clinical microbiology reviews*. **22**(4),pp.552–563.

Gorzelnik, K.V., Cui, Z., Reed, C.A., Jakana, J., Young, R. and Zhang, J. 2016. Asymmetric cryo-EM structure of the canonical Allolevivirus Qβ reveals a single maturation protein and the genomic ssRNA in situ. *Proceedings of the National Academy of Sciences of the United States of America*. **113**(41),pp.11519–11524.

Gosch, M. and Rigler, R. 2005. Fluorescence correlation spectroscopy of molecular motions and kinetics. *Advanced Drug Delivery Reviews*. **57**(1),pp.169–190.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic acids research*. **38**(Web Server issue),pp.W695–9.

Grassly, N.C. 2018. Eradicating polio with a vaccine we must stop using. *The Lancet. Infectious diseases*. **18**(6),pp.590–591.

Haaheim, L.R., Pattison, J.R. and Whitley, R.J. (eds.). 2002. *A Practical Guide to Clinical Virology* 2nd ed. Chichester, UK: John Wiley & Sons, Ltd.

Hafenstein, S., Palermo, L.M., Kostyuchenko, V.A., Xiao, C., Morais, M.C., Nelson, C.D.S., Bowman, V.D., Battisti, A.J., Chipman, P.R., Parrish,

C.R. and Rossmann, M.G. 2007. Asymmetric binding of transferrin receptor to parvovirus capsids. *Proceedings of the National Academy of Sciences*. **104**(16),pp.6585–6589.

Hagan, M.F. 2009. A theory for viral capsid assembly around electrostatic cores. *The Journal of Chemical Physics*. **130**(11),p.114902.

Haneef, I., Talbot, S.J. and Stockley, P.G. 1989. Modeling loop structures in proteins and nucleic acids: an RNA stem-loop. *Journal of molecular graphics*. **7**(4),pp.186–195.

Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K. and Bricogne, G. 1978. Tomato bushy stunt virus at 2.9 A resolution. *Nature*. **276**(5686),pp.368–373.

Hayashi, T., Shioda, T., Iwakura, Y. and Shibuta, H. 1992. RNA packaging signal of human immunodeficiency virus type 1. *Virology*. **188**(2),pp.590–599.

Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., Saxena, P., Thompson, R.F., Cockburn, J.J., Lomonossoff, G.P. and Ranson, N.A. 2015. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nature communications*. **6**(1),p.10113.

Hiebert, E. and Bancroft, J.B. 1969. Factors affecting the assembly of some spherical viruses. *Virology*. **39**(2),pp.296–311.

Hirao, I., Spingola, M., Peabody, D. and Ellington, A.D. 1998. The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Molecular diversity*. **4**(2),pp.75–89.

Hodgkin, D.C. 1950. X-RAY ANALYSIS AND PROTEIN STRUCTURE. *Cold Spring Harbor symposia on quantitative biology*. **14**(0),pp.65–78.

Hollings, M. and Stone, O.M. 1971. Viruses that Infect Fungi. *Annual Review of Phytopathology*. **9**(1),pp.93–118.

Holmes, K.C., Stubbs, G.J., Mandelkow, E. and Gallwitz, U. 1975. Structure of tobacco mosaic virus at 6.7 å resolution. *Nature*. **254**(5497),pp.192–196.

Hong, P., Koza, S. and Bouvier, E.S.P. 2012. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of liquid chromatography & related technologies*. **35**(20),pp.2923–2950.

Hu, Y., Zandi, R., Anavitarte, A., Knobler, C.M. and Gelbart, W.M. 2008. Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. *Biophysical journal*. **94**(4),pp.1428–1436.

Integrated DNA Technologies 2018. IDT Codon Optimization Tool. Available from: https://eu.idtdna.com/CodonOpt#. [Accessed September 2018]

Invitrogen 2002. *Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture Techniques* [Online]. Available from: https://tools.thermofisher.com/content/sfs/manuals/bevtest.pdf.

Ivanowski, D. 1892. Ueber die Mosaikkrankheit der tabakspflanze. *St Petersburg Academy Imperial Science Bulletin*. (35),pp.67–70.

James Martin Center for Nonproliferation Studies 2018. Chemical and Biological Weapons. Available from: https://www.nonproliferation.org/chemical-and-biological-weapons-possession-and-programs-past-and-present/.

Johansson, H.E., Liljas, L. and Uhlenbeck, O.C. 1997. RNA Recognition by the MS2 Phage Coat Protein. *Seminars in Virology*. **8**(3),pp.176–185.

Jones, T.A. and Liljas, L. 1984. Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 A resolution. *Journal of molecular biology*. **177**(4),pp.735–767.

Jose, J., Snyder, J.E. and Kuhn, R.J. 2009. A structural and functional perspective of alphavirus replication and assembly. *Future microbiology*. **4**(7),pp.837–856.

Kaempfer, R., van Emmelo, J. and Fiers, W. 1981. Specific binding of eukaryotic initiation factor 2 to satellite tobacco necrosis virus RNA at a 5'-terminal sequence comprising the ribosome binding site. *Proceedings of the National Academy of Sciences*. **78**(3),pp.1542–1546.

Kausche, G.A., Pfankuch, E. and Ruska, H. 1939. Die Sichtbarmachung von pflanzlichem Virus im □bermikroskop. *Die Naturwissenschaften*. **27**(18),pp.292–299.

Kawamoto, S., Ueda, K., Mita, E. and Matsubara, K. 1994. The packaging signal in hepatitis B virus pregenome functions only at the 5' end. *Journal of virological methods*. **49**(2),pp.113–127.

Keifer, D.Z., Pierson, E.E. and Jarrold, M.F. 2017. Charge detection mass spectrometry: weighing heavier things. *The Analyst*. **142**(10),pp.1654–1671.

Kelly, S.M., Jess, T.J. and Price, N.C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. **1751**(2),pp.119–139.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. **181**(4610),pp.662–666.

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**(16),pp.8022–8036.

King, A.M.Q., Adams, M.J., Carstens, E.B. and Lefkowitz, E. (eds.). 2012. *Virus Taxonomy: Classification and Nomenclature of Viruses*. San Diego, CA: Elsevier Academic Press.

King, L.A. and Possee, R.D. 1992. *The Baculovirus Expression System*. Dordrecht: Springer Netherlands.

Knipe, D.M. and Howley, P.M. (eds.). 2013. *Fields's Virology* 6 ed. Lippincott Williams & Wilkins, 2013.

Koning, R.I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., Carazo, J.M. and Koster, A.J. 2016. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nature communications*. **7**,p.12524.

Korsman, S.N.J., van Zyl, G.U., Nutt, L., Andersson, M.I. and Preiser, W. 2012. Togaviruses *In*: *Virology*. Elsevier, pp. 98–99.

Kostyuchenko, V.A., Jakana, J., Liu, X., Haddow, A.D., Aung, M., Weaver, S.C., Chiu, W. and Lok, S.-M. 2011. The structure of barmah forest virus as revealed by cryo-electron microscopy at a 6-angstrom resolution has detailed transmembrane protein architecture and interactions. *Journal of virology*. **85**(18),pp.9327–9333.

Krichevsky, O. and Bonnet, G. 2002. Fluorescence correlation spectroscopy: the technique and its applications. *Reports on Progress in Physics*. **65**(2),pp.251–297.

Krupovic, M., Kuhn, J.H. and Fischer, M.G. 2016. A classification system for virophages and satellite viruses. *Archives of Virology*. **161**(1),pp.233–247.

Laine, M., Luukkainen, R. and Toivanen, A. 2004. Sindbis viruses and other alphaviruses as cause of human arthritic disease. *Journal of Internal Medicine*. **256**(6),pp.457–471.

Lakowicz, J.R. (ed.). 2006. *Principles of Fluorescence Spectroscopy* [Online]. Boston, MA: Springer US. Available from: http://link.springer.com/10.1007/978-0-387-46312-4.

Lane, S.W., Dennis, C.A., Lane, C.L., Trinh, C.H., Rizkallah, P.J., Stockley, P.G. and Phillips, S.E.V. 2011. Construction and crystal structure of recombinant STNV capsids. *Journal of molecular biology*. **413**(1),pp.41–50.

Larson, S.B., Day, J., Greenwood, A. and McPherson, A. 1998. Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *Journal of molecular biology*. **277**(1),pp.37–59.

Larson, S.B., Koszelak, S., Day, J., Greenwood, A., Dodds, J.A. and McPherson, A. 1993. Double-helical RNA in satellite tobacco mosaic virus. *Nature*. **361**(6408),pp.179–182.

Laue, T.M., Shah, B.D., Ridgeway, T.M. and Pelletier, S.L. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. *In*: S. E. Harding, A. J. Rowe and J. C. Horton, eds. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Cambridge, Royal Society of Chemistry, pp. 90–125.

Leica Microsystems 2010. *Fluorescence Correlation Spectroscopy : The Femtoliter Test Tube – System Calibration and In Vitro Applications*

[Online]. Available from: https://www.leica-microsystems.com/fileadmin/downloads/Leica%20TCS%20SP8%20SMD/Application%20Notes/SMD_FCS_Basics-AppLetter.EN.pdf.

Levis, R., Weiss, B.G., Tsiang, M., Huang, H. and Schlesinger, S. 1986. Deletion mapping of Sindbis virus DI RNAs derived from cDNAs defines the sequences essential for replication and packaging. *Cell*. **44**(1),pp.137–145.

Life Technologies 2015. *Bac-to-Bac® Baculovirus Expression System* [Online] 0 ed. Available from: https://tools.thermofisher.com/content/sfs/manuals/bactobac_man.pdf.

Liljas, L., Unge, T., Jones, T.A., Fridborg, K., Lövgren, S., Skoglund, U. and Strandberg, B. 1982. Structure of satellite tobacco necrosis virus at 3.0 A resolution. *Journal of molecular biology*. **159**(1),pp.93–108.

Lin, T., Cavarelli, J. and Johnson, J.E. 2003. Evidence for assembly-dependent folding of protein and RNA in an icosahedral virus. *Virology*. **314**(1),pp.26–33.

Littlefield, J.W., Keller, E.B., Gross, J. and Zamecknik, P. 1955. Studies on Cytoplasmic Ribonucleoprotein particles from the liver of the rat. *Journal of Biological Chemistry*. (217),pp.111–124.

Lundstrom, K. 2017. Oncolytic Alphaviruses in Cancer Immunotherapy. *Vaccines*. **5**(2),p.9.

Lustig, A. and Levine, A.J. 1992. *One hundred years of virology.* American Society for Microbiology (ASM).

Lührmann, R., Kastner, B. and Bach, M. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochimica et biophysica acta*. **1087**(3),pp.265–292.

Mancini, E.J., Clarke, M., Gowen, B.E., Rutten, T. and Fuller, S.D. 2000. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Molecular cell*. **5**(2),pp.255–266.

Marbach, A. and Bettenbrock, K. 2012. lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of biotechnology*. **157**(1),pp.82–88.

Mathiot, C.C., Grimaud, G., Garry, P., Bouquety, J.C., Mada, A., Daguisy, A.M. and Georges, A.J. 1990. An outbreak of human Semliki Forest virus infections in Central African Republic. *The American journal of tropical medicine and hygiene*. **42**(4),pp.386–393.

McElwee, M., Vijayakrishnan, S., Rixon, F. and Bhella, D. 2018. Structure of the herpes simplex virus portal-vertex. B. Sugden, ed. *PLoS biology*. **16**(6),p.e2006191.

Mendes, A. and Kuhn, R.J. 2018. Alphavirus Nucleocapsid Packaging and Assembly. *Viruses*. **10**(3),p.138.

Minor, P.D. 2015. Live attenuated vaccines: Historical successes and current challenges. *Virology*. **479-480**,pp.379–392.

Mukhopadhyay, S., Chipman, P.R., Hong, E.M., Kuhn, R.J. and Rossmann, M.G. 2002. In vitro-assembled alphavirus core-like particles maintain a structure similar to that of nucleocapsid cores in mature virus. *Journal of virology*. **76**(21),pp.11128–11132.

Mukhopadhyay, S., Zhang, W., Gabler, S., Chipman, P.R., Strauss, E.G., Strauss, J.H., Baker, T.S., Kuhn, R.J. and Rossmann, M.G. 2006. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure (London, England : 1993)*. **14**(1),pp.63–73.

Mulvey, T. 2002. Origins and historical development of the electron microscope. *British Journal of Applied Physics*. **13**(5),pp.197–207.

Nabel, G.J. 2013. Designing tomorrow's vaccines. *The New England journal of medicine*. **368**(6),pp.551–560.

Nagai, K. 1996. RNA—protein complexes. *Current Opinion in Structural Biology*. **6**(1),pp.53–61.

Narayanan, K. and Makino, S. 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *Journal of virology*. **75**(19),pp.9059–9067.

National Audit Office 2002. *The 2001 Outbreak of Foot and Mouth Disease* [Online]. The House of Commons. Available from: https://www.nao.org.uk/wp-content/uploads/2002/06/0102939.pdf.

Nicaise, V. 2014. Crop immunity against viruses: outcomes and future challenges. *Frontiers in plant science*. **5**(227),p.660.

Novagen 2018. *Competent Cells* [Online] 220006 ed. Available from: http://www.merckmillipore.com/GB/en/product/RosettaDE3-Competent-Cells-Novagen,EMD_BIO-70954#anchor_BRO.

Novagen 2003. pET systems manual. Available from: http://lifeserv.bgu.ac.il/wp/zarivach/wp-content/uploads/2017/11/Novagen-pET-system-manual-1.pdf.

Nugent, C.I., Johnson, K.L., Sarnow, P. and Kirkegaard, K. 1999. Functional coupling between replication and packaging of poliovirus replicon RNA. *Journal of virology*. **73**(1),pp.427–435.

Pan American Health Organization 2017. Chikungunya Data, Maps and Statistics 2014-2017. Available from: https://www.paho.org/hq/index.php?option=com_topics&view=article&id=343&Itemid=40931&lang=en.

Parent, K.N., Doyle, S.M., Anderson, E. and Teschke, C.M. 2005. Electrostatic interactions govern both nucleation and elongation during phage P22 procapsid assembly. *Virology*. **340**(1),pp.33–45.

Pasetti, M.F., Simon, J.K., Sztein, M.B. and Levine, M.M. 2011. Immunology of Gut Mucosal Vaccines. *Immunol Rev*. 2011 Jan; **239**(1): 125–148.

Patel, N., Dykeman, E.C., Coutts, R.H.A., Lomonossoff, G.P., Rowlands, D.J., Phillips, S.E.V., Ranson, N., Twarock, R., Tuma, R. and Stockley, P.G. 2015. Revealing the density of encoded functions in a viral RNA. *Proceedings of the National Academy of Sciences of the United States of America*. **112**(7),pp.2227–2232.

Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiß, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E., Tuma, R., Twarock, R., Ranson, N.A. and Stockley, P.G. 2017. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nature microbiology*. **2**,p.17098.

Patel, N., Wroblewski, E., Leonov, G., Phillips, S.E.V., Tuma, R., Twarock, R. and Stockley, P.G. 2017. Rewriting nature's assembly manual for a ssRNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. **114**(46),pp.12255–12260.

Pennock, G.D.,Shoemaker,C. and Miller, L K. 1984. Strong and Regulated Expression of Escherichia coli β-Galactosidase in Insect Cells with a Baculovirus Vector. *Molecular and Cellular Biology*. **4**(3) p.399-406.

Perlmutter, J.D. and Hagan, M.F. 2015. Mechanisms of virus assembly. *Annual review of physical chemistry*. **66**(1),pp.217–239.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T. 1960. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*. **185**(4711),pp.416–422.

Pierce 2014. Instructions Pierce protease inhibitor tablets. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0011744_Pierce_Protease_Inhibit_MiniTab_UG.pdf.

Pierson, E. E., Keifer, D. Z., Kukreja, A. A., Wang, J. C. Y., Zlotnick, A. Jarrold, M. F. 2016. Charge Detection Mass Spectrometry Identifies Preferred Non-Icosahedral Polymorphs in the Self-Assembly of Woodchuck Hepatitis Virus Capsids. *Journal of Molecular Biology*. **428(**2), pp 292-300.

Pineda, C., Muñoz-Louis, R., Caballero-Uribe, C.V. and Viasus, D. 2016. Chikungunya in the region of the Americas. A challenge for rheumatologists and health care systems. *Clinical rheumatology*. **35**(10),pp.2381–2385.

Plotkin, S. 2014. History of vaccination. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(34),pp.12283–12287.

Plotkin, S.A. and Plotkin, S.L. 2011. The development of vaccines: how the past led to the future. *Nature reviews. Microbiology*. **9**(12),pp.889–893.

Powers, A.M. and Roehrig, J.T. 2010. Alphaviruses *In*: *Diagnostic Virology Protocols*. Methods in Molecular Biology. Totowa, NJ: Humana Press, pp. 17–38.

Prangishvili, D., Forterre, P. and Garrett, R.A. 2006. Viruses of the Archaea: a unifying view. *Nature reviews. Microbiology*. **4**(11),pp.837–848.

Prasad, B.V.V. and Schmid, M.F. 2012. Principles of virus structural organization. *Advances in experimental medicine and biology*. **726**(Chapter 3),pp.17–47.

Qu, F. and Morris, T.J. 1997. Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *Journal of virology*. **71**(2),pp.1428–1435.

Ralston, G. 1993. *Introduction to Analytical Ultracentrifugation*. Fullerton, CA.

Rao, A.L.N. 2006. Genome packaging by spherical plant RNA viruses. *Annual Review of Phytopathology*. **44**(1),pp.61–87.

Reed, D.S., Larsen, T., Sullivan, L.J., Lind, C.M., Lackemeyer, M.G., Pratt, W.D. and Parker, M.D. 2005. Aerosol exposure to western equine encephalitis virus causes fever and encephalitis in cynomolgus macaques. *The Journal of infectious diseases*. **192**(7),pp.1173–1182.

Rees, M.W., Short, M.N. and Kassanis, B. 1970. The amino acid composition, antigenicity, and other characteristics of the satellite viruses of tobacco necrosis virus. *Virology*. **40**(3),pp.448–461.

Roche 2004. Roche applied science: the complete guide for protease inhibition. Available from: https://iti.stanford.edu/content/dam/sm/iti/documents/himc/immunoassays/ProteaseInhibitionGuide.pdf.

Roldão, A., Mellado, M.C.M., Castilho, L.R., Carrondo, M.J.T. and Alves, P.M. 2010. Virus-like particles in vaccine development. *Expert review of vaccines*. **9**(10),pp.1149–1176.

Rolfsson, O., Middleton, S., Manfield, I.W., White, S.J., Fan, B., Vaughan, R., Ranson, N.A., Dykeman, E., Twarock, R., Ford, J., Kao, C.C. and Stockley, P.G. 2016. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *Journal of molecular biology*. **428**(2 Pt B),pp.431–448.

Rolfsson, O., Toropova, K., Morton, V., Francese, S., Basnak, G., Thompson, G.S., Homans, S.W., Ashcroft, A.E., Stonehouse, N.J., Ranson, N.A. and Stockley, P.G. 2008. RNA Packing Specificity and Folding during Assembly of the Bacteriophage MS2. *Computational and Mathematical Methods in Medicine*. **9**(3-4),pp.339–349.

Ross, W., Aiyar, S.E., Salomon, J. and Gourse, R.L. 1998. Escherichia coli promoters with UP elements of different strengths: modular structure of bacterial promoters. *Journal of bacteriology*. **180**(20),pp.5375–5383.

Rossmann, M.G. and Blow, D.M. 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*. **15**(1),pp.24–31.

Routh, A., Domitrovic, T. and Johnson, J.E. 2012a. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(6),pp.1907–1912.

Routh, A., Domitrovic, T. and Johnson, J.E. 2012b. Packaging host RNAs in small RNA viruses: an inevitable consequence of an error-prone polymerase? *Cell cycle (Georgetown, Tex.)*. **11**(20),pp.3713–3714.

Rowsell, S., Stonehouse, N.J., Convery, M.A., Adams, C.J., Ellington, A.D., Hirao, I., Peabody, D.S., Stockley, P.G. and Phillips, S.E. 1998. Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nature structural biology*. **5**(11),pp.970–975.

Sahin, E. and Roberts, C.J. 2012. Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. *Methods in molecular biology (Clifton, N.J.)*. **899**(Chapter 25),pp.403–423.

Schneemann, A. 2006. The structural and functional role of RNA in icosahedral virus assembly. *Annual review of microbiology*. **60**(1),pp.51–67.

Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal*. **78**(3),pp.1606–1619.

Schuck, P., Zhao, H., Brautigam, C. and Ghirlando, R. 2015. Analytical Ultracentrifugation Basics *In*: *Basic Principles of Analytical Ultracentrifugation*. CRC Press, pp. 1–22.

Schwille, P. and Haustein, E. 2001. Fluorescence Correlation Spectroscopy - An Introduction to its Concepts and Applications. **94(22)**(3). Available from: https://www.researchgate.net/publication/235410358_Fluorescence_Correlation_Spectroscopy_-_An_Introduction_to_its_Concepts_and_Applications.

Schwille, P., Oehlenschläger, F. and Walter, N.G. 1996. Quantitative hybridization kinetics of DNA probes to RNA in solution followed by diffusional fluorescence correlation analysis. *Biochemistry*. **35**(31),pp.10182–10193.

Scott, W.G. 1998. RNA catalysis. *Current Opinion in Structural Biology*. **8**(6),pp.720–726.

Scott, W.G. and Klug, A. 1996. Ribozymes: structure and mechanism in RNA catalysis. *Trends in Biochemical Sciences*. **21**(6),pp.220–224.

Serdyuk, I.N., Zaccai, N.R. and Zaccai, J. 2007. *Methods in Molecular Biophysics by Igor N. Serdyuk* [Online]. Cambridge: Cambridge University Press. Available from: https://www.cambridge.org/core/books/methods-in-molecular-biophysics/26C040D1C6ED5B6B9937470F83A890CE?pageNum=2.

Shakeel, S., Dykeman, E.C., White, S.J., Ora, A., Cockburn, J.J.B., Butcher, S.J., Stockley, P.G. and Twarock, R. 2017. Genomic RNA folding mediates assembly of human parechovirus. *Nature communications*. **8**(1),p.5.

Sharma, R., Fatma, B., Saha, A., Bajpai, S., Sistla, S., Dash, P.K., Parida, M., Kumar, P. and Tomar, S. 2016. Inhibition of chikungunya virus by picolinate that targets viral capsid protein. *Virology*. **498**,pp.265–276.

Shirako, Y., Niklasson, B., Dalrymple, J.M., Strauss, E.G. and Strauss, J.H. 1991. Structure of the Ockelbo virus genome and its relationship to other Sindbis viruses. *Virology*. **182**(2),pp.753–764.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. **7**(1),pp.539–539.

Sigma 2018. *Aprotinin product information* [Online]. Available from: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/6/a6103dat.pdf.

Sigma 2001. *Pepstatin A product information* [Online]. Available from: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/1/p4265pis.pdf.

Simon, A.E. and Gehrke, L. 2009. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochimica et biophysica acta*. **1789**(9-10),pp.571–583.

Smith, G E., Summers, M D. and Fraser, M J. 1983. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular and Cellular Biology*. **3**(12) pp. 2156-2165.

Snijder, J., Rose, R.J., Veesler, D., Johnson, J.E. and Heck, A.J.R. 2013. Studying 18 MDa Virus Assemblies with Native Mass Spectrometry. *Angewandte Chemie International Edition*. **52**(14),pp.4020–4023.

Sorger, P.K., Stockley, P.G. and Harrison, S.C. 1986. Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *Journal of molecular biology*. **191**(4),pp.639–658.

Sottrup-Jensen, L. 1989. Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *The Journal of biological chemistry*. **264**(20),pp.11539–11542.

Spurgers, K.B. and Glass, P.J. 2011. Vaccine Development for Biothreat Alpha viruses. *Journal of Bioterrorism & Biodefense*. **01**(01),pp.1–9.

Sreerama, N. and Woody, R.W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*. **287**(2),pp.252–260.

Stanley, W.M. 1935. Isolation of a crystalline protein posessing the properties of tobacco-mosaic virus. *Science (New York, N.Y.)*. **81**(2113),pp.644–645.

Stewart, H., Bingham, R.J., White, S.J., Dykeman, E.C., Zothner, C., Tuplin, A.K., Stockley, P.G., Twarock, R. and Harris, M. 2016. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Scientific reports*. **6**,p.22952.

Stockley, P.G., Rolfsson, O., Thompson, G.S., Basnak, G., Francese, S., Stonehouse, N.J., Homans, S.W. and Ashcroft, A.E. 2007. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *Journal of molecular biology*. **369**(2),pp.541–552.

Stockley, P.G., Stonehouse, N.J., Murray, J.B., Goodman, S.T., Talbot, S.J., Adams, C.J., Liljas, L. and Valegård, K. 1995. Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic acids research*. **23**(13),pp.2512–2518.

Stockley, P.G., Twarock, R., Bakker, S.E., Barker, A.M., Borodavka, A., Dykeman, E., Ford, R.J., Pearson, A.R., Phillips, S.E.V., Ranson, N.A. and Tuma, R. 2013. Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *Journal of biological physics*. **39**(2),pp.277–287.

Strauss, J.H. and Strauss, E.G. 2008. Plus-Strand RNA Viruses *In*: *Viruses and Human Disease*. Elsevier, pp. 63–136.

Strauss, J.H. and Strauss, E.G. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiological reviews*. **58**(3),pp.491–562.

Sun, S., Rao, V.B. and Rossmann, M.G. 2010. Genome packaging in viruses. *Current Opinion in Structural Biology*. **20**(1),pp.114–120.

Talbot, S.J., Goodman, S., Bates, S.R.E., Fishwick, C.W.G. and Stockley, P.G. 1990. Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex. *Nucleic acids research*. **18**(12),pp.3521–3528.

Tang, L., Johnson, K.N., Ball, L.A., Lin, T., Yeager, M. and Johnson, J.E. 2001. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nature structural biology*. **8**(1),pp.77–83.

Tarasow, T.M. and Eaton, B.E. 1998. Dressed for success: Realizing the catalytic potential of RNA. *Biopolymers*. **48**(1),pp.29–37.

Tellinghuisen, T.L., Hamburger, A.E., Fisher, B.R., Ostendorp, R. and Kuhn, R.J. 1999. In vitro assembly of alphavirus cores by using nucleocapsid protein expressed in Escherichia coli. *Journal of virology*. **73**(7),pp.5309–5319.

Thermo Fisher Scientific 2018. DNA and RNA Molecular Weights and Conversions. Available from: https://www.thermofisher.com/uk/en/home/references/ambion-tech-support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html.

Thermo Fisher Scientific 2015. *PMSF Protease Inhibitor - Thermo Fisher Scientific* [Online]. Available from: https://www.thermofisher.com/order/catalog/product/36978.

Tian, Y., Martinez, M.M. and Pappas, D. 2011. Fluorescence correlation spectroscopy: a review of biochemical and microfluidic applications. *Applied spectroscopy*. **65**(4),pp.115A–124A.

Tito, M.A., Tars, K., Valegard, K., Hajdu, J. and Robinson, C.V. 2000. Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid. *Journal of the American Chemical Society*. **122**(14),pp.3550–3551.

Twarock, R. and Stockley, P.G. 2019. RNA-mediated virus assembly mechanisms and consequences for viral evolution and therapy. *Annual Reviews in Biophysics. In Press.*

Uetrecht, C., Versluis, C., Watts, N.R., Roos, W.H., Wuite, G.J.L., Wingfield, P.T., Steven, A.C. and Heck, A.J.R. 2008. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(27),pp.9216–9220.

Unge, T., Liljas, L., Strandberg, B., Vaara, I., Kannan, K.K., Fridborg, K., Nordman, C.E. and Lentz, P.J. 1980. Satellite tobacco necrosis virus structure at 4.0 Å resolution. *Nature*. **285**(5764),pp.373–377.

UniProt Consortium, T. 2018. UniProt: the universal protein knowledgebase. *Nucleic acids research*. **46**(5),pp.2699–2699.

Valegård, K., Liljas, L., Fridborg, K. and Unge, T. 1990. The three-dimensional structure of the bacterial virus MS2. *Nature*. **345**(6270),pp.36–41.

Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*. **371**(6498),pp.623–626.

van der Schoot, P. and Bruinsma, R. 2005. Electrostatics and the assembly of an RNA virus. *Physical review. E, Statistical, nonlinear, and soft matter physics*. **71**(6 Pt 1),p.061928.

van Emmelo, J., Ameloot, P., Plaetinck, G. and Fiers, W. 1984. Controlled synthesis of the coat protein of satellite tobacco necrosis virus in Escherichia coli. *Virology*. **136**(1),pp.32–40.

Van Etten, J.L. 2011. Another Really, Really Big Virus. *Viruses*. **3**(1),pp.32–46.

van Oers, M.M., Pijlman, G.P. and Vlak, J.M. 2015. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *The Journal of general virology*. **96**(Pt 1),pp.6–23.

Voet, D. and Voet, J.G. 2010. *Biochemistry* 4 ed. New York: John Wiley & Sons, Inc.

Wang, A. 1991. Viruses Of The Protozoa. *Annual review of microbiology*. **45**(1),pp.251–263.

Weaver, S.C. and Lecuit, M. 2015. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease E. W. Campion, ed. *New England Journal of Medicine*. **372**(13),pp.1231–1239.

Weiss, B., Geigenmüller-Gnirke, U. and Schlesinger, S. 1994a. Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. *Nucleic acids research*. **22**(5),pp.780–786.

Weiss, B., Geigenmüller-Gnirke, U. and Schlesinger, S. 1994b. Interactions between Sindbis virus RNAs and a 68 amino acid derivative of the viral capsid protein further defines the capsid binding site. *Nucleic acids research*. **22**(5),pp.780–786.

Weiss, B., Nitschko, H., Ghattas, I., Wright, R. and Schlesinger, S. 1989. Evidence for specificity in the encapsidation of Sindbis virus RNAs. *Journal of virology*. **63**(12),pp.5310–5318.

Weiss, B., Rosenthal, R. and Schlesinger, S. 1980. Establishment and maintenance of persistent infection by Sindbis virus in BHK cells. *Journal of virology*. **33**(1),pp.463–474.

Wengler, G., Boege, U., Bischoff, H. and Wahn, K. 1982. The core protein of the alphavirus Sindbis virus assembles into core-like nucleoproteins with the viral genome RNA and with other single-stranded nucleic acids in vitro. *Virology*. **118**(2),pp.401–410.

Wheaton, R.M. and Bauman, W.C. 1953. Non-ionic separations with ion exchange resins. *Annals of the New York Academy of Sciences*. **57**(3),pp.159–176.

White, C.L., Thompson, M. and Dimmock, N.J. 1998. Deletion Analysis of a Defective Interfering Semliki Forest Virus RNA Genome Defines a Region in the nsP2 Sequence That Is Required for Efficient Packaging of the Genome into Virus Particles. *Journal of virology*. **72**(5),pp.4320–4326.

Whitmore, L. and Wallace, B.A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research*. **32**(Web Server),pp.W668–W673.

Wiley, D.C. and Bonsdorff, Von, C.H. 1978. Three-dimensional crystals of the lipid-enveloped Semliki Forest virus. *Journal of molecular biology*. **120**(3),pp.375–379.

Wilks, C.R. 2016. Alphaviruses: current biology S. Malingham, L. J. Herrero, & B. L. Herring, eds. *Australian Veterinary Journal*. **94**(3),pp.59–59.

Willems, W.R., Kaluza, G., Boschek, C.B., Bauer, H., Hager, H., Schütz, H.J. and Feistner, H. 1979. Semliki forest virus: cause of a fatal case of human encephalitis. *Science (New York, N.Y.)*. **203**(4385),pp.1127–1129.

Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C. 1991. Specific Interaction between RNA Phage Coat Proteins and RNA *In*: Progress in Nucleic Acid Research and Molecular Biology. Elsevier, pp. 185–220.

Woolhouse, M., Scott, F., Hudson, Z., Howey, R. and Chase-Topping, M. 2012. Human viruses: discovery and emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **367**(1604),pp.2864–2871.

Woolhouse, M.E.J. and Gowtage-Sequeria, S. 2005. Host range and emerging and reemerging pathogens. *Emerging infectious diseases*. **11**(12),pp.1842–1847.

World Health Organization 2016. Global health estimates 2016 summary tables:. Available from: http://www.who.int/healthinfo/global_burden_disease/en/.

World Health Organization 2018. HIV/AIDS. Available from: http://www.who.int/news-room/fact-sheets/detail/hiv-aids.

Wressnigg, N., van der Velden, M.V.W., Portsmouth, D., Draxler, W., O'Rourke, M., Richmond, P., Hall, S., McBride, W.J.H., Redfern, A., Aaskov, J., Barrett, P.N. and Aichinger, G. 2015. An inactivated Ross River virus vaccine is well tolerated and immunogenic in an adult population in a randomized phase 3 trial. S. A. Plotkin, ed. *Clinical and vaccine immunology : CVI*. **22**(3),pp.267–273.

Würkner, D. and Wengler, G. 1992. Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. *Virology*. **191**(2),pp.880–888.

Wyatt Technology, 2016. Astra 6 software. Wyatt Technology.

Wyatt Technology 2018a. Understanding Light Scattering and Chromatography Mode. *Wyatt Technology*. [Online]. [Accessed 24 March 2018a]. Available from: https://www.wyatt.com/library/theory/light-scattering-chromatography-mode.html.

Wyatt Technology 2018b. Understanding Multi-Angle Static Light Scattering. [Accessed 9 August 2018b]. Available from: https://www.wyatt.com/library/theory/understanding-multi-angle-static-light-scattering.html.

Ylösmäki, E., Martikainen, M., Hinkkanen, A. and Saksela, K. 2013. Attenuation of Semliki Forest virus neurovirulence by microRNA-mediated detargeting. *Journal of virology*. **87**(1),pp.335–344.

Ysebaert, M., van Emmelo, J. and Fiers, W. 1980. Total nucleotide sequence of a nearly full-size DNA copy of satellite tobacco necrosis virus RNA. *Journal of molecular biology*. **143**(3),pp.273–287.

Yu, W., Mengersen, K., Dale, P., Mackenzie, J.S., Toloo, G.S., Wang, X. and Tong, S. 2014. Epidemiologic patterns of Ross River virus disease in Queensland, Australia, 2001-2011. *The American journal of tropical medicine and hygiene*. **91**(1),pp.109–118.

Zandi, R., van der Schoot, P., Reguera, D., Kegel, W. and Reiss, H. 2006. Classical nucleation theory of virus capsids. *Biophysical journal*. **90**(6),pp.1939–1948.

Zhang, R., Hryc, C.F., Cong, Y., Liu, X., Jakana, J., Gorchakov, R., Baker, M.L., Weaver, S.C. and Chiu, W. 2011. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *The EMBO journal*. **30**(18),pp.3854–3863.

Zhang, W., Fisher, B.R., Olson, N.H., Strauss, J.H., Kuhn, R.J. and Baker, T.S. 2002. Aura virus structure suggests that the T=4 organization is a fundamental property of viral structural proteins. *Journal of virology*. **76**(14),pp.7239–7246.

Zhang, W., Heil, M., Kuhn, R.J. and Baker, T.S. 2005. Heparin binding sites on Ross River virus revealed by electron cryo-microscopy. *Virology*. **332**(2),pp.511–518.

Zlotnick, A. and Mukhopadhyay, S. 2011. Virus assembly, allostery and antivirals. *Trends in microbiology*. **19**(1),pp.14–23.

Appendix A: Sequences for Protein Expression

[This page has been left intentionally blank]

A.1 Synthetic STNV gene

ATGGCGAAACAGCAGAACAACCGCCGCAAATCTGCGACCATGCGCGCG
GTGAAACGCATGATCAACACCCACCTGGAACACAAACGCTTCGCGCTG
ATCAACTCTGGTAACACCAACGCGACCGCGGGCACCGTGCAGAACCTG
TCTAACGGTATCATCCAGGGTGACAGCATCAACCAGCGCTCTGGTGAC
CAGGTGCGCATCGTGTCTCACAAACTGCACGTGCGCGGCACCGCGATC
ACCGTGTCTCAGACCTTCCGCTTCATCTGGTTCCGCGACAACATGAACC
GCGGTAACCACCCCGACCGTGCTGGAAGTGCTGAACACCGCGAACTTC
ATGTCTCAGTACAACCCGATCACCCTGCAGCAGAAACGCTTCACCATCC
TGAAAGACGTGACCCTGAACTGCTCTCTGACCGGTGAATCTATCAAAGA
CCGCATCATCAACCTGCCGGGTCAGCTGGTGAACTACAACGGTGCGAC
CGCGGTGGCGGCGTCTAACGGTCCGGGTGCGATCTTCATGCTGCAGAT
CGGTGACTCTCTGGTGGGTCTGTGGGACTCCTCTTACGAAGCGGTGTA
CACTGACGCGTA

Sequence from (Lane *et al* 2011)

A.2 Wild type STNV gene

ATGGCAAAACAACAGAACAACAGGCGAAAATCCGCAACAATGCGTGCA
GTGAAGCGCATGATAAATACACACTTGGAGCATAAAAGGTTTGCACTGA
TCAACTCAGGGAACACCAATGCAACTGCTGGTACAGTACAAAATCTGTC
CAACGGTATAATCCAAGGAGATGATATCAACCAGAGAAGTGGTGATCAA
GTGCGTATAGTTTCACATAAACTTCACGTACGAGGCACTGCCATCACCG
TCAGCCAGACCTTTAGATTTATCTGGTTTCGTGATAACATGAACCGTGG
GAACCACTCCCACAGTTCTTGAGGTGTTGAACACTGCGAATTTCATGTC
GCAGTATAACCCAATCACGTTGCAGCAAAAGAGATTTACAATACTCAAG
GATGTAACTCTCAATTGTTCGCTGACAGGGGAGAGCATTAAAGATCGGA
TAATTAACCTTCCAGGACAACTGGTGAACTATAATGGAGCGACGGCTGT
AGCAGCCTCCAATGGTCCCGGCGCAATATTTATGTTGCAGATTGGCGA
CTCCTTGGTTGGTCTGTGGGACTCCTCTTATGAGGCTGTGTACACAGAT
GCATAA

Genbank accession V01468.1

A.3 STNV CP amino acid sequence

MAKQQNNRRKSATMRAVKRMINTHLEHKRFALINSGNTNATAGTVQNLSN
GIIQGDDINQRSGDQVRIVSHKLVRGTAITVSQTFRFIWFRDNMNRGTTPTV
LEVLNTANFMSQYNPITLQQKRFTILKDVTLNCSLTGESIKDRIINLPGQLVN
YNGATAVAASNGPGAIFMLQIGDSLVGLWDSSYEAVYTDA

Extinction coefficient = 16960 M^{-1} cm^{-1}

A.4 Chikungunya virus (CHIKV) Cp sequence

ATGGAGTTTATCCCGACCCAGACCTTCTATAACCGCCGTTACCAACCGC
GTCCTTGGACGCCGCGCCCTACTATCCAGGTGATTGGTCCCCGCCCGC
GCCCGCAACGTCAGGCTGGCCAACTGGCCCAGCTGATCTCTGCCGTGA
ATAAACTGACTATGCGTGCCGTTCCGCAACAGAAGCCGCGCCGCAACC
GCAAGAACAAGAAACAGAAACAAAAGCAGCAGGCACCTCAAAATAATAC
GAATCAGAAGAAACAGCCTCCTAAGAAGAAGCCCGCGCAAAAGAAGAA
GAAACCCGGCCGTCGTGAACGCATGTGCATGAAAATTGAAAATGATTGT
ATCTTCGGTGTCAAACATGAAGGAAAAGTTACCGGTTATGCGTGTCTGG
TGGGCGATAAGGTGATGAAACCAGCTCATGTGAAAGGCACGATCGATA
ACGCTGACCTGGCCAAACTGGCGTTCAAACGCAGTTCAAAGTACGATCT
TGAATGCGCCCAAATCCCCGTCCACATGAAATCGGACGCTTCTAAATTC
ACGCACGAAAAGCCGGAAGGCTACTACAATTGGCATCACGGTGCGGTT
CAGTACTCAGGCGGTCGTTTCACCATCCCAACCGGGGCCGGTAAACCG
GGCGACTCCGGTCGCCCTATTTTCGATAACAAAGGTCGTGTAGCAGCA
ATTGTCTTGGGCGGGGCTAACGAAGGTGCTCGTACCGCTTTATCAGTG
GTGACCTGGAACAAAGATATTGTAACGAAAATTACACCGGAAGGTGCCG
AAGAATGGTAA

A.5 Chikungunya virus (CHIKV) Cp amino acid sequence

MEFIPTQTFYNRRYQPRPWTPRPTIQVIGPRPRPQRQAGQLAQLISAVNKL
TMRAVPQQKPRRNRKNKKQKQKQQAPQNNTNQKKQPPKKKPAQKKKKP
GRRERMCMKIENDCIFGVKHEGKVTGYACLVGDKVMKPAHVKGTIDNADL
AKLAFKRSSKYDLECAQIPVHMKSDASKFTHEKPEGYYNWHHGAVQYSG
GRFTIPTGAGKPGDSGRPIFDNKGRVAAIVLGGANEGARTALSVVTWNKDI
VTKITPEGAEEW
Genbank accession ADJ93861.1

Extinction coefficient = 32,680 M^{-1} cm^{-1}

A.6 Semliki Forest Virus (SFV) Cp sequence

ATGAATTACATCCCTACGCAAACGTTTTACGGTCGTCGTTGGCGTCCGC
GTCCGGCGGCACGTCCGTGGCCGTTGCAGGCCACTCCGGTGGCTCCG
GTCGTCCCGGACTTCCAGGCCCAGCAGATGCAGCAACTCATCAGCGCC
GTAAATGCGCTGACCATGCGCCAGAACGCAATTGCTCCTGCTGGTCCT
CCGAAACCAAAGAAGAAGAAGACGACCAAACCAAAGCCGAAAACGCAG
CCGAAGAAGATCAAAGGCAAAACGCAGCAGCAAAAGAAGAAAGACAAG
CAAGCCGACAAGAAGAAGAAGAAACCGGGTAAACGTGAACGTATGTGC
ATGAAGATTGAAAATGACTGTATCTTCGAAGTCAAACACGAAGGTAAGG
TCACTGGCTACGCCTGCCTGGTGGGCGACAAAGTCATGAAACCTGCCC
ACGTGAAAGGTGTCATCGACAACGCGGACCTGGCAAAGCTGGCTTTCA
AGAAATCGAGCAAGTATGACCTTGAGTGTGCCCAGATTCCAGTTCACAT
GCGTTCGGATGCCTCAAAGTACACGCATGAGAAGCCGGAGGGTCACTA
CAACTGGCACCACGGTGCTGTTCAGTACAGCGGCGGTCGTTTCACTAT
TCCGACCGGTGCGGGCAAACCGGGTGACAGTGGCCGTCCGATCTTTG
ACAACAAGGGCCGTGTAGTCGCTATCGTCCTGGGCGGTGCCAACGAG
GGCTCACGCACCGCACTGTCGGTGGTCACCTGGAACAAAGATATGGTG
ACTCGTGTGACCCCGGAAGGTTCCGAAGAGTGGTAA

A.7 Semliki Forest Virus (SFV) Cp amino acid sequence

MNYIPTQTFYGRRWRPRPAARPWPLQATPVAPVVPDFQAQQMQQLISAV
NALTMRQNAIAPAGPPKPKKKKTTKPKPKTQPKKIKGKTQQQKKKDKQAD
KKKKKPGKRERMCMKIENDCIFEVKHEGKVTGYACLVGDKVMKPAHVKGV
IDNADLAKLAFKKSSKYDLECAQIPVHMRSDASKYTHEKPEGHYNWHHGA
VQYSGGRFTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGSRTALSVVT
WNKDMVTRVTPEGSEEW

Genbank accession: AAM64227.1

Extinction coefficient = 37,930 M^{-1} cm^{-1}

A.8 Ross River Virus (RRV) Cp sequence

ATGAACTACATTCCGACCCAAACCTTTTATGGCCGCCGTTGGCGCCCG
CGCCCGGCATTCCGTCCATGGCAGGTCCCCATGCAGCCAACGCCGAC
GATGGTTACGCCAATGCTGCAAGCACCCGACTTGCAAGCGCAGCAGAT
GCAGCAGCTGATTTCTGCAGTGTCTGCACTGACGACAAAACAGAATGTG
AAAGCCCCCAAGGGTCAGCGGAAAAAAAAGCAGCAGAAGCCGAAGGA
GAAAAAAGAGAATCAGAAAAAGAAACCGACCCAAAAAAAAAAACAACAA
CAGAAGCCGAAGCCTCAGGCCAAAAAGAAAAAACCAGGTCGGCGTGAA
CGTATGTGCATGAAAATTGAAAACGACTGCATTTTCGAGGTAAAACTGG
ATGGCAAAGTGACGGGCTATGCGTGCCTGGTGGGAGACAAAGTCATGA
AACCGGCCCATGTTAAAGGAACCATCGATAATCCGGACCTGGCCAAGT
TGACTTACAAAAAAAGTTCTAAGTACGATTTAGAATGTGCTCAGATTCCG

GTGCACATGAAATCGGATGCATCTAAGTATACCCACGAGAAGCCGGAG
GGTCATTACAACTGGCATCATGGCGCGGTCCAGTATTCAGGGGGACGC
TTTACCATTCCGACGGGTGCCGGTAAACCTGGCGATAGCGGTCGCCCA
ATCTTTGACAATAAAGGTCGTGTCGTGGCAATTGTTTTGGGAGGCGCCA
ATGAGGGAGCTCGGACCGCTTTATCTGTCGTAACTTGGACAAAAGATAT
GGTTACGCGCGTCACCCCGGAGGGCACCGAAGAATGGTAA

A.9 Ross River Virus (RRV) Cp amino acid sequence

MNYIPTQTFYGRRWRPRPAFRPWQVPMQPTPTMVTPMLQAPDLQAQQM
QQLISAVSALTTKQNVKAPKGQRKKKQQKPKEKKENQKKKPTQKKKQQQK
PKPQAKKKKPGRRERMCMKIENDCIFEVKLDGKVTGYACLVGDKVMKPAH
VKGTIDNPDLAKLTYKKSSKYDLECAQIPVHMKSDASKYTHEKPEGHYNWH
HGAVQYSGGRFTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTAL
SVVTWTKDMVTRVTPEGTEEW
Genbank accession ABB53381.1

Extinction coefficient = 39,420 M^{-1} cm^{-1}

A.10 Sindbis Virus (SINV) Cp sequence

ATGAACCGCGGCTTCTTTAACATGCTTGGACGCCGGCCGTTCCCCGCC
CCAACGGCGATGTGGCGCCCTCGTCGCCGGCGTCAGGCAGCGCCTAT
GCCGGCCCGCAACGGTCTGGCCAGCCAGATTCAACAACTGACGACGG
CAGTGAGCGCCCTGGTGATTGGTCAAGCGACACGCCCTCAGCCGCCTC
GTCCACGGCCTCCGCCTCGTCAGAAGAAACAGGCTCCGAAACAGCCGC
CGAAACCTAAGAAACCGAAAACACAAGAAAAGAAGAAGAAACAGCCTG
CGAAACCGAAGCCAGGCAAACGTCAGCGGATGGCCCTGAAACTTGAAG
CCGATCGCCTGTTCGATGTCAAGAATGAGGATGGCGACGTCATCGGTC
ATGCGTTGGCGATGGAGGGCAAAGTGATGAAACCGCTTCACGTTAAAG
GGACTATTGATCACCCCGTATTAAGTAAGCTGAAATTCACCAAAAGCAG
CGCCTACGACATGGAATTCGCGCAGCTTCCGGTGAATATGCGCAGTGA
AGCCTTCACATATACCTCGGAGCACCCGGAGGGCTTTTACAACTGGCAT
CACGGGGCAGTTCAGTATTCCGGCGGCCGCTTCACCATCCCGCGTGGT
GTTGGTGGACGCGGCGACTCCGGTCGCCCGATTATGGACAACTCCGG
CCGGGTTGTAGCCATTGTATTAGGTGGCGCAGATGAAGGTACCCGGAC
GGCGCTGAGCGTAGTGACTTGGAATTCCAAGGGCAAAACCATCAAAAC
AACCCCAGAAGGTACCGAGGAATGGTAA

A.11 Sindbis Virus (SINV) Cp amino acid sequence

MNRGFFNMLGRRPFPAPTAMWRPRRRRQAAPMPARNGLASQIQQLTTAV
SALVIGQATRPQPPRPRPPPRQKKQAPKQPPKPKKPKTQEKKKKQPAKPK
PGKRQRMALKLEADRLFDVKNEDGDVIGHALAMEGKVMKPLHVKGTIDHP

VLSKLKFTKSSAYDMEFAQLPVNMRSEAFTYTSEHPEGFYNWHHGAVQYS
GGRFTIPRGVGGRGDSGRPIMDNSGRVVAIVLGGADEGTRTALSVVTWNS
KGKTIKTTPEGTEEW
Genbank accession NP_740673
Extinction coefficient = 27,960 M^{-1} cm^{-1}

A.12 Western Equine Encephalitis Virus (WEEV) Cp sequence

ATGTTTCCGTATCCGCAATTGAACTTTCCGCCAGTGTACCCTACCAACC
CAATGGCATATCGCGATCCGAACCCACCCCGCTGCCGTTGGCGTCCCT
TTCGCCCCCCTCTGGCAGCACAGATTGAAGATTTGCGCCGTTCCATTGC
CAATCTTACCTTCAAACAGCGCAGTCCGAACCCACCGCCGGGCCCCCC
ACCTAAAAAAAAAAAAATCTGCGCCCAAACCGAAACCTACCCAGCCGAAA
AAGAAAAAACAACAGGCTAAGAAAACCAAACGCAAGCCGAAACCAGGC
AAGCGTCAACGTATGTGCATGAAACTTGAAAGTGACAAAACCTTTCCAA
TTATGTTGAACGGACAGGTTAATGGGTACGCGTGCGTCGTGGGGGGTC
GCCTGATGAAACCGCTGCACGTGGAAGGCAAGATCGATAACGAACAAC
TTGCGGCCGTGAAGTTGAAAAAAGCATCCATGTATGATCTCGAGTACGG
TGATGTGCCGCAAAATATGAAATCGGACACATTACAATACACCAGCGAC
AAACCTCCAGGTTTTTATAACTGGCATCATGGCGCCGTGCAGTACGAAA
ACGGCCGCTTTACGGTTCCGCGGGGGGTAGGTGGCAAAGGGGACTCT
GGTCGTCCGATTCTGGATAACCGCGGCCGTGTTGTAGCTATTGTGCTG
GGCGGCGCAAACGAAGGCACGCGTACCGCCCTGTCCGTGGTTACCTG
GAACCAGAAAGGCGTGACGATCAAAGATACCCCCGAAGGTTCGGAACC
ATGGTAA

A.13 Western Equine Encephalitis Virus (WEEV) Cp amino acid sequence

MFPYPQLNFPPVYPTNPMAYRDPNPPRCRWRPFRPPLAAQIEDLRRSIAN
LTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQAKKTKRKPKPGKRQ
RMCMKLESDKTFPIMLNGQVNGYACVVGGRLMKPLHVEGKIDNEQLAAVK
LKKASMYDLEYGDVPQNMKSDTLQYTSDKPPGFYNWHHGAVQYENGRFT
VPRGVGGKGDSGRPILDNRGRVVAIVLGGANEGTRTALSVVTWNQKGVTI
KDTPEGSEPW
Genbank accession P13897.1

Extinction coefficient = 35,410 M^{-1} cm^{-1}

[This page has been left intentionally blank]

Appendix B: Primers

[This page has been left intentionally blank]

B.1 B3 short variant primers

Primer Name	Sequence 5′→3′	Tm °C
Forward	GACATTAATACGACTCACTATAGGGACATGCA	65.5
AUUArev	GTGTGCATAATTGCATGTCCCTATAGTGAGTCG	68.2
GUUGrev	GTGTGCACAACTGCATGTCCCTATAGTGAGTCG	70.0
AUUGrev	GTGTGCACAATTGCATGTCCCTATAGTGAGTCG	69.5
GUUArev	GTGTGCATAACTGCATGTCCCTATAGTGAGTCG	69.5
UUUArev	GTGTGCATAAATGCATGTCCCTATAGTGAGTCG	68.2
AUUUrev	GTGTGCAAAATTGCATGTCCCTATAGTGAGTCG	68.2
GUUUrev	GTGTGCAAAACTGCATGTCCCTATAGTGAGTCG	69.5
UUUGrev	GTGTGCACAAATGCATGTCCCTATAGTGAGTCG	69.5

B3 sequence variants. Loop motif (left), full sequence (middle) and the Mfold folding free energy values (right) of the B3 sequence variants.

Loop motif	Sequence	∆G of folding (kcal mol^{-1})
ACAA	ACAUGCAACAAUGCACAC	-2.6
AUUU	ACAUGCAAUUUUGCACAC	-2.6
UUUA	ACAUGCAUUUAUGCACAC	-2.6
GUUU	ACAUGCAGUUUUGCACAC	-2.2
UUUG	ACAUGCAUUUGUGCACAC	-2.9
AUUG	ACAUGCAAUUGUGCACAC	-2.6
GUUA	ACAUGCAGUUAUGCACAC	-3.4
GUUG	ACAUGCAGUUGUGCACAC	-2.5
AUUA	ACAUGCAAUUAUGCACAC	-2.6

B.2 Alphavirus FastBac1 primers

Primer name	Sequence 5′→3′	Annealing temperature ˚C
CHIKV_5'his_tev FWD	CCGTCCCACCATCGGGCGCGATGGAGTTTATCCCGACC	61.8
CHIKV_5'his_tev REV	TCCTCTAGTACTTCTCGACATTATTATTAGTGGTGGTGATGATGATGTTCGTTCAGATAAAACTGGCCCCATTCTTCGGCACCTTC	
WEEV_5'his_tev FWD	CCGTCCCACCATCGGGCGCGATGTTTCCGTATCCGCAATTG	64.1
WEEV_5'his_tev REV	TCCTCTAGTACTTCTCGACATTATTATTAGTGGTGGTGATGATGATGTTCGTTCAGATAAAACTGGCCCCATGGTTCCGAACCTTC	
SINV_5'his_tev FWD	CCGTCCCACCATCGGGCGCGATGAACCGCGGCTTCTTTAAC	62.9
SINV_5'his_tev REV	TCCTCTAGTACTTCTCGACATTATTATTAGTGGTGGTGATGATGATGTTCGTTCAGATAAAACTGGCCCCATTCCTCGGTACCTTC	
SFV_5'his_tev FWD	CCGTCCCACCATCGGGCGCGATGAATTACATCCCTACGC	59.6
SFV_5'his_tev REV	TCCTCTAGTACTTCTCGACATTATTATTAGTGGTGGTGATGATGATGTTCGTTCAGATAAAACTGGCCCCACTCTTCGGAACCTTC	
RRV_5'his_tev FWD	CCGTCCCACCATCGGGCGCGATGAACTACATTCCGACCC	62.2
RRV_5'his_tev REV	TCCTCTAGTACTTCTCGACATTATTATTAGTGGTGGTGATGATGATGTTCGTTCAGATAAAACTGGCCCCATTCTTCGGTGCCCTC	

B.3 Sequencing primers

Primer name	Sequence 5′→3′
Polyhedrin Forward	GAAATTTGTGATGCTATTGC
SV40 reverse	AAATGATAACCATCTCGC
pBamRev	GGTGATGTCGGCGATATAGG

B.4 Bacmid PCR primers

Primer name	Sequence 5′→3′
pUC/M13 Forward	CCCAGTCACGACGTTGTAAAACG
pUC/M13 Reverse	AGCGGATAACAATTTCACACAGG

Appendix C: Supplementary Information for Chapter 3

[This page has been left intentionally blank]

Name	Molecular weight (Da)
WT-STNV CP	21715
Recombinant VLP CP	21577.52
1239 nt genome	1239 x 320.5 + 159 =397258.5

M.W. of ssRNA = (# nucleotides x 320.5) + 159.0 (Thermo Fisher Scientific, 2018). STNV genome length = 1239 nt. VLP packaged fragment = ~500 bp (Lane et al., 2011)

WT STNV

Component	Molecular weight (Da)	Quantity	Total (Da)
WT-STNV CP	21715	60	1302900
Ca^{2+} ions	40.078	92	3687.176
Genome	397258.5	1	397258.5
			1703845.676

Empty VLP

Component	Molecular weight (Da)	Quantity	Total (Da)
Recombinant VLP CP	21577.52	60	1294651.2
Ca^{2+} ions	40.078	92	3687.176
			1298338.376

Packaged 1239 nt fragment

Component	Molecular weight (Da)	Quantity	Total (Da)
Recombinant VLP CP	21577.52	60	1294651.2
Ca^{2+} ions	40.078	92	3687.176
1239 nt RNA	397258.5	1	397258.5
			1695596.88

Packaged 500 nt fragment

Component	Molecular weight (Da)	Quantity	Total (Da)
Recombinant VLP CP	21577.52	60	1294651.2
Ca^{2+} ions	40.078	92	3687.176
500 nt fragment	160409	1	160409
			1458747.376

500 nt fragment as 20% of total packaged RNA

Component	Molecular weight (Da)	Quantity	Total (Da)
Recombinant VLP CP	21577.52	60	1294651.2
Ca^{2+} ions	40.078	92	3687.176
Packaged*	802045	1	802045
			2100383.376

*(1239 x 320.5) + 159 = 397258.5, (397258.5/20) x 100 = 1986292.5
M.W. of ssRNA = (# nucleotides x 320.5) + 159.0 (Thermo Fisher Scientific, 2018)

O

[This page has been left intentionally blank]

Appendix D: gBlock Sequences

[This page has been left intentionally blank]

gBlock WT-ΔAXXA

ATCGGGATCCTAATACGACTCACTATAGGGAGTAAAGACAGGAAACTTT
ACTGACTAACATGGCAAAACAACAGAACAACAGGCGAAAATCCGCAACA
ATGCGTGCAGTGAAGCGCATGATAAATACACACTTGGAGCATAAAAGGT
TTGCACTGATTCTTGGAGCATTTTTGGTTTGCACTGTTCTTCTCAGGGTT
CTCCAATGCAACTGCTGGTACAGTTCTTTTTCTGTCCAACGGTTTTTTCC
TTGGTGATGATATCAACCTGTGTTGTGGTGATCAAGTGCGTATAGTTTCT
CTTTAACTTCACGTACGAGGCACTGCCTTCTCCGTCTGCCTGACCTTTA
GTTTTTTCTGGTTTCGTGTTTTCATGAACCGTGGGTCCTCTCCCACAGTT
CTTGAGGTGTTGAACACTGCGAATTTCATGTCGCAGTATAACCCAATCA
CGTTGCAGCAAAAGAGATTTACAATACTCAAGGATGTAACTCTCAATTGT
TCGCTGACAGGGGAGAGCATTAAAGATCGGTTTTTTTTCCTTCCAGGTC
TTCTGGTGAACTATAATGGAGCGACGGCTGTAGCAGCCTCCAATGGTC
CCGGCGCAATTTTTTTGTTGCAGATTGGCGACTCCTTGGTTGGTCTGTG
GGACTCCTCTTATGAGGCTGTGTACTCTGTTGCATAATCCCAGAGGTTC
ACAATGTTAGTGATGGGGCGCTGTTTGTTGCGTAGCTACCCTTCTGGAG
CCACTTCCTGGTGGTAAGCAGAAATCCAAGGGTACGGTGGTACGGTGG
AAAGCAGTCCCAGCTCTGCATTGGGAACCGGCTTACACCCAGCTTAGG
GCTAAAGTGTACTACTTGCTCATTTGTAGTCTAAATGAGACGTTGGCCT
CGACGTGTCGAGGTGGCCTAAGGGATTGGAACCCCTGATGGTCGTAGT
CGAATTTCCCGTGTTTCATTCCGAGTCTCTTGGTCTTTTTGCCATTAGTA
GGTCTAGCACTCAACGTAACTTCAAAGATATCCTCCTTGCAACATGTTTT
TGTGCGCCGTCTGTGTTTAAAGCGGTATATTAAGTGCGCCGGCATATCG
TTGTTTGGACCAGGGCCCCACGCCGGTTGGTACCCGGGTGGCTTCCCC
TCGTTCACAGGGCTTTAGGTGTTGTTAAGGTATAGTTTTTTGACAAATGC
GGACAAACCTGAAAAGCTCGCTAGTGGTGGGCTGGCCAAGCGATGTTC
CTCATCCAGGTATAGTTCTACATGGGAAATTTGGTTCCTTCCAAACTTCT
ATGAAGTCCTCGACTACCCCAAGCTTATCG

Sequence WT-ΔAXXA

GAGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAAC
AACAGGCGAAAAUCCGCAACAAUGCGUGCAGUGAAGCGCAUGAUAAA
UACACACUUGGAGCAUAAAAGGUUUGCACUGAUUCUUGGAGCAUUUU
UGGUUUGCACUGUUCUUCUCAGGGUUCUCCAAUGCAACUGCUGGUA
CAGUUCUUUUUCUGUCCAACGGUUUUUUCCUUGGUGAUGAUAUCAAC
CUGUGUUGUGGUGAUCAAGUGCGUAUAGUUUCUCUUUAACUUCACGU
ACGAGGCACUGCCUUCUCCGUCUGCCUGACCUUUAGUUUUUUCUGG
UUUCGUGUUUUCAUGAACCGUGGGUCCUCUCCCACAGUUCUUGAGG
UGUUGAACACUGCGAAUUUCAUGUCGCAGUAUAACCCAAUCACGUUG
CAGCAAAAGAGAUUUACAAUACUCAAGGAUGUAACUCUCAAUUGUUCG
CUGACAGGGGAGAGCAUUAAAGAUCGGUUUUUUUUCCUUCCAGGUCU
UCUGGUGAACUAUAAUGGAGCGACGGCUGUAGCAGCCUCCAAUGGUC
CCGGCGCAAUUUUUUUGUUGCAGAUUGGCGACUCCUUGGUUGGUCU
GUGGGACUCCUCUUAUGAGGCUGUGUACUCUGUUGCAUAAUCCCAGA
GGUUCACAAUGUUAGUGAUGGGGCGCUGUUUGUUGCGUAGCUACCC
UUCUGGAGCCACUUCCUGGUGGUAAGCAGAAAUCCAAGGGUACGGU
GGUACGGUGGAAAGCAGUCCCAGCUCUGCAUUGGGAACCGGCUUAC
ACCCAGCUUAGGGCUAAAGUGUACUACUUGCUCAUUUGUAGUCUAAA
UGAGACGUUGGCCUCGACGUGUCGAGGUGGCCUAAGGGAUUGGAAC

CCCUGAUGGUCGUAGUCGAAUUUCCCGUGUUUCAUUCCGAGUCUCU
UGGUCUUUUUGCCAUUAGUAGGUCUAGCACUCAACGUAACUUCAAAG
AUAUCCUCCUUGCAACAUGUUUUUGUGCGCCGUCUGUGUUUAAAGCG
GUAUAUUAAGUGCGCCGGCAUAUCGUUGUUUGGACCAGGGCCCCAC
GCCGGUUGGUACCCGGGUGGCUUCCCCUCGUUCACAGGGCUUUAGG
UGUUGUUAAGGUAUAGUUUUUUGACAAAUGCGGACAAACCUGAAAAG
CUCGCUAGUGGUGGGCUGGCCAAGCGAUGUUCCUCAUCCAGGUAUA
GUUCUACAUGGGAAAUUUGGUUCCUUCCAAACUUCUAUGAAGUCCUC
GACUACCCC

Extinction coefficient = 11,477,000 M^{-1} cm^{-1}

gBlock ΔAXXA

ATCGGGATCCTAATACGACTCACTATAGGGAGTAAAGACTTTTAACTTTA
CTGACTAACATGGCAAAACAACTTTTCAACAGGCGAAAATCCGCATTTTT
GCGTGCAGTGAAGCGCATGATATTTTCACACTTGGAGCATTTTTGGTTT
GCACTGATCAACTCAGGGAACACCAATGCAACTGCTGGTACAGTACAAA
ATCTGTCCAACGGTATAATCCAAGGAGATGATATCAACCAGAGAAGTGG
TGATCAAGTGCGTATAGTTTCACATAAACTTCACGTACGAGGCACTGCC
ATCACCGTCAGCCAGACCTTTAGATTTATCTGGTTTCGTGATAACATGAA
CCGTGGGACCACTCCCACAGTTCTTGAGGTGTTGAACACTGCGAATTTC
ATGTCGCAGTATAACCCAATCACGTTGCAGCAAAAGAGATTTACAATAC
TCAAGGATGTAACTCTCAATTGTTCGCTGACAGGGGAGAGCATTAAAGA
TCGGATAATTAACCTTCCAGGACAACTGGTGAACTATAATGGAGCGACG
GCTGTAGCAGCCTCCAATGGTCCCGGCGCAATATTTATGTTGCAGATTG
GCGACTCCTTGGTTGGTCTGTGGGACTCCTCTTATGAGGCTGTGTACAC
AGATGCATAATCCCAGAGGTTCACAATGTTAGTGATGGGGCGCTGAAA
GATGCGTAGCTACCCTTCTGGAGCCACTTCCTGGTGGTAAGCAGAAAT
CCAAGGGTACGGTGGTACGGTGGAAAGCAGTCCCAGCTCTGCATTGGG
AACCGGCTTACACCCAGCTTAGGGCTAAAGTGTACTACTTGCTCATTTG
TAGTCTAAATGAGACGTTGGCCTCGACGTGTCGAGGTGGCCTAAGGGA
TTGGAACCCCTGATGGTCGTAGTCGAATTTCCCGTGTTTCATTCCGAGT
CTCTTGGTCATAATGCCATTAGTAGGTCTAGCACTCAACGTAACTTCAAA
GATATCCTCCTTGCAACAAGAATATGTGCGCCGTCTGTGTTTAAAGCGG
TATATTAAGTGCGCCGGCATATCGTTGTTTGGACCAGGGCCCCACGCC
GGTTGGTACCCGGGTGGCTTCCCCTCGTTCACAGGGCTTTAGGAGATG
ATAAGGTATAGTTATTAGACAAATGCGGACAAACCTGAAAAGCTCGCTA
GTGGTGGGCTGGCCAAGCGAAGAACCTCATCCAGGTATAGTTCTACAT
GGGAAATTTGGTACCATCCAAACTTCTATGAAGTCCTCGACTACCCCAA
GCTTATCG

Sequence ΔAXXA

GAGUAAAGACUUUUAACUUUACUGACUAACAUGGCAAAACAACUUUUC
AACAGGCGAAAAUCCGCAUUUUUGCGUGCAGUGAAGCGCAUGAUAUU
UUCACACUUGGAGCAUUUUUGGUUUGCACUGAUCAACUCAGGGAACA
CCAAUGCAACUGCUGGUACAGUACAAAAUCUGUCCAACGGUAUAAUC
CAAGGAGAUGAUAUCAACCAGAGAAGUGGUGAUCAAGUGCGUAUAGU

UUCACAUAAACUUCACGUACGAGGCACUGCCAUCACCGUCAGCCAGA
CCUUUAGAUUUAUCUGGUUUCGUGAUAACAUGAACCGUGGGACCACU
CCCACAGUUCUUGAGGUGUUGAACACUGCGAAUUUCAUGUCGCAGUA
UAACCCAAUCACGUUGCAGCAAAAGAGAUUUACAAUACUCAAGGAUGU
AACUCUCAAUUGUUCGCUGACAGGGGAGAGCAUUAAAGAUCGGAUAA
UUAACCUUCCAGGACAACUGGUGAACUAUAAUGGAGCGACGGCUGUA
GCAGCCUCCAAUGGUCCCGGCGCAAUAUUUAUGUUGCAGAUUGGCGA
CUCCUUGGUUGGUCUGUGGGACUCCUCUUAUGAGGCUGUGUACACA
GAUGCAUAAUCCCAGAGGUUCACAAUGUUAGUGAUGGGGCGCUGAAA
GAUGCGUAGCUACCCUUCUGGAGCCACUUCCUGGUGGUAAGCAGAAA
UCCAAGGGUACGGUGGUACGGUGGAAAGCAGUCCCAGCUCUGCAUU
GGGAACCGGCUUACACCCAGCUUAGGGCUAAAGUGUACUACUUGCUC
AUUUGUAGUCUAAAUGAGACGUUGGCCUCGACGUGUCGAGGUGGCC
UAAGGGAUUGGAACCCCUGAUGGUCGUAGUCGAAUUUCCCGUGUUU
CAUUCCGAGUCUCUUGGUCAUAAUGCCAUUAGUAGGUCUAGCACUCA
ACGUAACUUCAAAGAUAUCCUCCUUGCAACAAGAAUAUGUGCGCCGU
CUGUGUUUAAAGCGGUAUAUUAAGUGCGCCGGCAUAUCGUUGUUUG
GACCAGGGCCCCACGCCGGUUGGUACCCGGGUGGCUUCCCCUCGUU
CACAGGGCUUUAGGAGAUGAUAAGGUAUAGUUAUUAGACAAAUGCGG
ACAAACCUGAAAAGCUCGCUAGUGGUGGGCUGGCCAAGCGAAGAACC
UCAUCCAGGUAUAGUUCUACAUGGGAAAUUUGGUACCAUCCAAACUU
CUAUGAAGUCCUCGACUACCCC

Extinction coefficient = 11,551,200 M^{-1} cm^{-1}

gBlock C4-WT

ATCGGGATCCTAATACGACTCACTATAGGGCTGCCCTCAAGGACCAGG
GCAGAAAAGAGGAAAAGAAAAGTGACAGAACACTTATAAGGAAAAAACG
TACAAACGTTTTAAGGAAAAAAGGAAGCTGCAATAGCGCAAGGAATCCG
AAAATTCGGAAAGGAATTCTTCTCAGGGTTCTCCAATGCAACTGCTGGT
ACAGTTCTTTTTCTGTCCAACGGTTTTTTCCTTGGTGATGATATCAACCT
GTGTTGTGGTGATCAAGTGCGTATAGTTTCTCTTTAACTTCACGTACGA
GGCACTGCCTTCTCCGTCTGCCTGACCTTTAGTTTTTTCTGGTTTCGTGT
TTTCATGAACCGTGGGTCCTCTCCCACAGTTCTTGAGGTGTTGAACACT
GCGAATTTCATGTCGCAGTATAACCCAATCACGTTGCAGCAAAAGAGAT
TTACAATACTCAAGGATGTAACTCTCAATTGTTCGCTGACAGGGGAGAG
CATTAAAGATCGGTTTTTTTTCCTTCCAGGTCTTCTGGTGAACTATAATG
GAGCGACGGCTGTAGCAGCCTCCAATGGTCCCGGCGCAATTTTTTTGT
TGCAGATTGGCGACTCCTTGGTTGGTCTGTGGGACTCCTCTTATGAGG
CTGTGTACTCTGTTGCATAATCCCAGAGGTTCACAATGTTAGTGATGGG
GCGCTGTTTGTTGCGTAGCTACCCTTCTGGAGCCACTTCCTGGTGGTAA
GCAGAAATCCAAGGGTACGGTGGTACGGTGGAAAGCAGTCCCAGCTCT
GCATTGGGAACCGGCTTACACCCAGCTTAGGGCTAAAGTGTACTACTTG
CTCATTTGTAGTCTAAATGAGACGTTGGCCTCGACGTGTCGAGGTGGC
CTAAGGGATTGGAACCCCTGATGGTCGTAGTCGAATTTCCCGTGTTTCA
TTCCGAGTCTCTTGGTCTTTTTGCCATTAGTAGGTCTAGCACTCAACGTA
ACTTCAAAGATATCCTCCTTGCAACATGTTTTTGTGCGCCGTCTGTGTTT
AAAGCGGTATATTAAGTGCGCCGGCATATCGTTGTTTGGACCAGGGCC

CCACGCCGGTTGGTACCCGGGTGGCTTCCCCTCGTTCACAGGGCTTTA
GGTGTTGTTAAGGTATAGTTTTTTGACAAATGCGGACAAACCTGAAAAG
CTCGCTAGTGGTGGGCTGGCCAAGCGATGTTCCTCATCCAGGTATAGT
TCTACATGGGAAATTTGGTTCCTTCCAAACTTCTATGAAGTCCTCGACTA
CCCCAAGCTTATCG

Sequence C4-WT
GCUGCCCUCAAGGACCAGGGCAGAAAAGAGGAAAAGAAAAGUGACAG
AACACUUAUAAGGAAAAAACGUACAAACGUUUUAAGGAAAAAAGGAAG
CUGCAAUAGCGCAAGGAAUCCGAAAAUUCGGAAAGGAAUUCUUCUCA
GGGUUCUCCAAUGCAACUGCUGGUACAGUUCUUUUUCUGUCCAACGG
UUUUUUCCUUGGUGAUGAUAUCAACCUGUGUUGUGGUGAUCAAGUG
CGUAUAGUUUCUCUUUAACUUCACGUACGAGGCACUGCCUUCUCCGU
CUGCCUGACCUUUAGUUUUUUCUGGUUUCGUGUUUUCAUGAACCGU
GGGUCCUCUCCCACAGUUCUUGAGGUGUUGAACACUGCGAAUUUCAU
GUCGCAGUAUAACCCAAUCACGUUGCAGCAAAAGAGAUUUACAAUACU
CAAGGAUGUAACUCUCAAUUGUUCGCUGACAGGGGAGAGCAUUAAAG
AUCGGUUUUUUUUCCUUCCAGGUCUUCUGGUGAACUAUAAUGGAGCG
ACGGCUGUAGCAGCCUCCAAUGGUCCCGGCGCAAUUUUUUUGUUGC
AGAUUGGCGACUCCUUGGUUGGUCUGUGGGACUCCUCUUAUGAGGC
UGUGUACUCUGUUGCAUAAUCCCAGAGGUUCACAAUGUUAGUGAUGG
GGCGCUGUUUGUUGCGUAGCUACCCUUCUGGAGCCACUUCCUGGUG
GUAAGCAGAAAUCCAAGGGUACGGUGGUACGGUGGAAAGCAGUCCCA
GCUCUGCAUUGGGAACCGGCUUACACCCAGCUUAGGGCUAAAGUGUA
CUACUUGCUCAUUUGUAGUCUAAAUGAGACGUUGGCCUCGACGUGUC
GAGGUGGCCUAAGGGAUUGGAACCCCUGAUGGUCGUAGUCGAAUUU
CCCGUGUUUCAUUCCGAGUCUCUUGGUCUUUUUGCCAUUAGUAGGU
CUAGCACUCAACGUAACUUCAAAGAUAUCCUCCUUGCAACAUGUUUU
UGUGCGCCGUCUGUGUUUAAAGCGGUAUAUUAAGUGCGCCGGCAUA
UCGUUGUUUGGACCAGGGCCCCACGCCGGUUGGUACCCGGGUGGCU
UCCCCUCGUUCACAGGGCUUUAGGUGUUGUUAAGGUAUAGUUUUUU
GACAAAUGCGGACAAACCUGAAAAGCUCGCUAGUGGUGGGCUGGCCA
AGCGAUGUUCCUCAUCCAGGUAUAGUUCUACAUGGGAAAUUUGGUUC
CUUCCAAACUUCUAUGAAGUCCUCGACUACCCC

Extinction coefficient = 11,359,500 M^{-1} cm^{-1}

Sequence WT STNV
GAGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAAC
AACAGGCGAAAAUCCGCAACAAUGCGUGCAGUGAAGCGCAUGAUAAA
UACACACUUGGAGCAUAAAAGGUUUGCACUGAUCAACUCAGGGAACA
CCAAUGCAACUGCUGGUACAGUACAAAAUCUGUCCAACGGUAUAAUC
CAAGGAGAUGAUAUCAACCAGAGAAGUGGUGAUCAAGUGCGUAUAGU
UUCACAUAAACUUCACGUACGAGGCACUGCCAUCACCGUCAGCCAGA
CCUUUAGAUUUAUCUGGUUUCGUGAUAACAUGAACCGUGGGACCACU
CCCACAGUUCUUGAGGUGUUGAACACUGCGAAUUUCAUGUCGCAGUA
UAACCCAAUCACGUUGCAGCAAAAGAGAUUUACAAUACUCAAGGAUGU
AACUCUCAAUUGUUCGCUGACAGGGGAGAGCAUUAAAGAUCGGAUAA

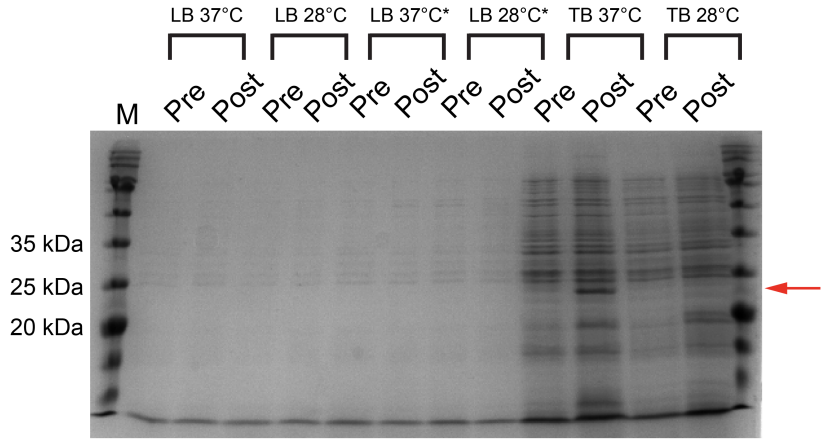
UUAACCUUCCAGGACAACUGGUGAACUAUAAUGGAGCGACGGCUGUA
GCAGCCUCCAAUGGUCCCGGCGCAAUAUUUAUGUUGCAGAUUGGCGA
CUCCUUGGUUGGUCUGUGGGACUCCUCUUAUGAGGCUGUGUACACA
GAUGCAUAAUCCCAGAGGUUCACAAUGUUAGUGAUGGGGCGCUGAAA
GAUGCGUAGCUACCCUUCUGGAGCCACUUCCUGGUGGUAAGCAGAAA
UCCAAGGGUACGGUGGUACGGUGGAAAGCAGUCCCAGCUCUGCAUU
GGGAACCGGCUUACACCCAGCUUAGGGCUAAAGUGUACUACUUGCUC
AUUUGUAGUCUAAAUGAGACGUUGGCCUCGACGUGUCGAGGUGGCC
UAAGGGAUUGGAACCCCUGAUGGUCGUAGUCGAAUUUCCCGUGUUU
CAUUCCGAGUCUCUUGGUCAUAAUGCCAUUAGUAGGUCUAGCACUCA
ACGUAACUUCAAAGAUAUCCUCCUUGCAACAAGAAUAUGUGCGCCGU
CUGUGUUUAAAGCGGUAUAUUAAGUGCGCCGGCAUAUCGUUGUUUG
GACCAGGGCCCCACGCCGGUUGGUACCCGGGUGGCUUCCCCUCGUU
CACAGGGCUUUAGGAGAUGAUAAGGUAUAGUUAUUAGACAAAUGCGG
ACAAACCUGAAAAGCUCGCUAGUGGUGGGCUGGCCAAGCGAAGAACC
UCAUCCAGGUAUAGUUCUACAUGGGAAAUUUGGUACCAUCCAAACUU
CUAUGAAGUCCUCGACUACCCC

Extinction coefficient = 11,645,100 M^{-1} cm^{-1}

[This page has been left intentionally blank]

Appendix E: Supplementary Information for Chapter 6

[This page has been left intentionally blank]



* Flasks were also heat shocked to 42°C for 30 min and cooled to expresison temp for 20 min.

Figure E.1: CHIKV core protein expression. SDS-PAGE of CHIKV core protein expression under differing conditions. Pre = pre-induction, Post = post-induction. M = BLUeye protein ladder (Geneflow).

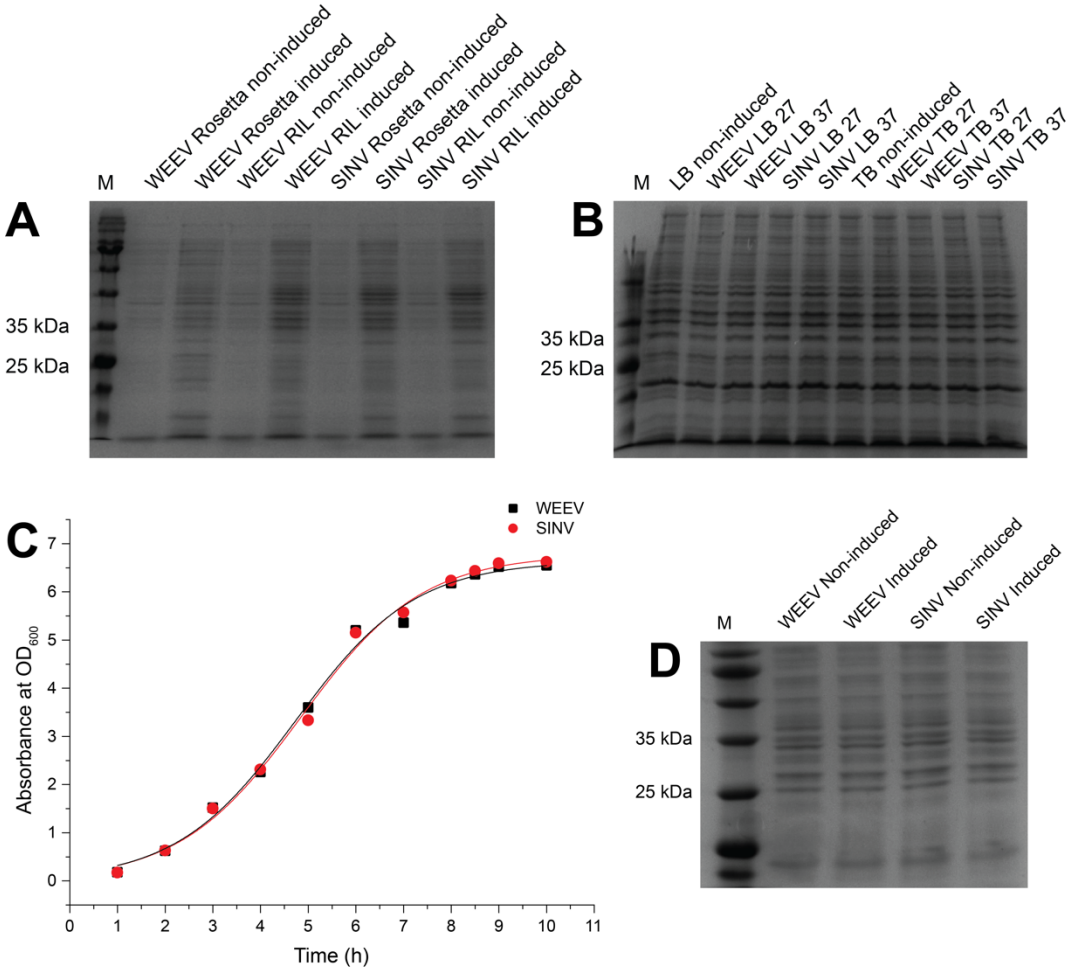


Figure E.2: Expression of SINV and WEEV core proteins. (A) SDS-PAGE of WEEV and SINV Cp expression in Rosetta and RIL cells at 37°C and 27°C. **(B)** SDS-PAGE of WEEV and SINV Cp expression (Rosetta cells) in LB and TB at 37°C and 27°C. **(C)** OD_{600} of cell growth of Rosetta-WEEV and Rosetta-SINV. **(D)** SDS-PAGE of non-induced cells and cells induced at stationary phase for Rosetta-WEEV Cp and Rosetta-SINV Cp. M = BLUeye protein ladder (Geneflow).

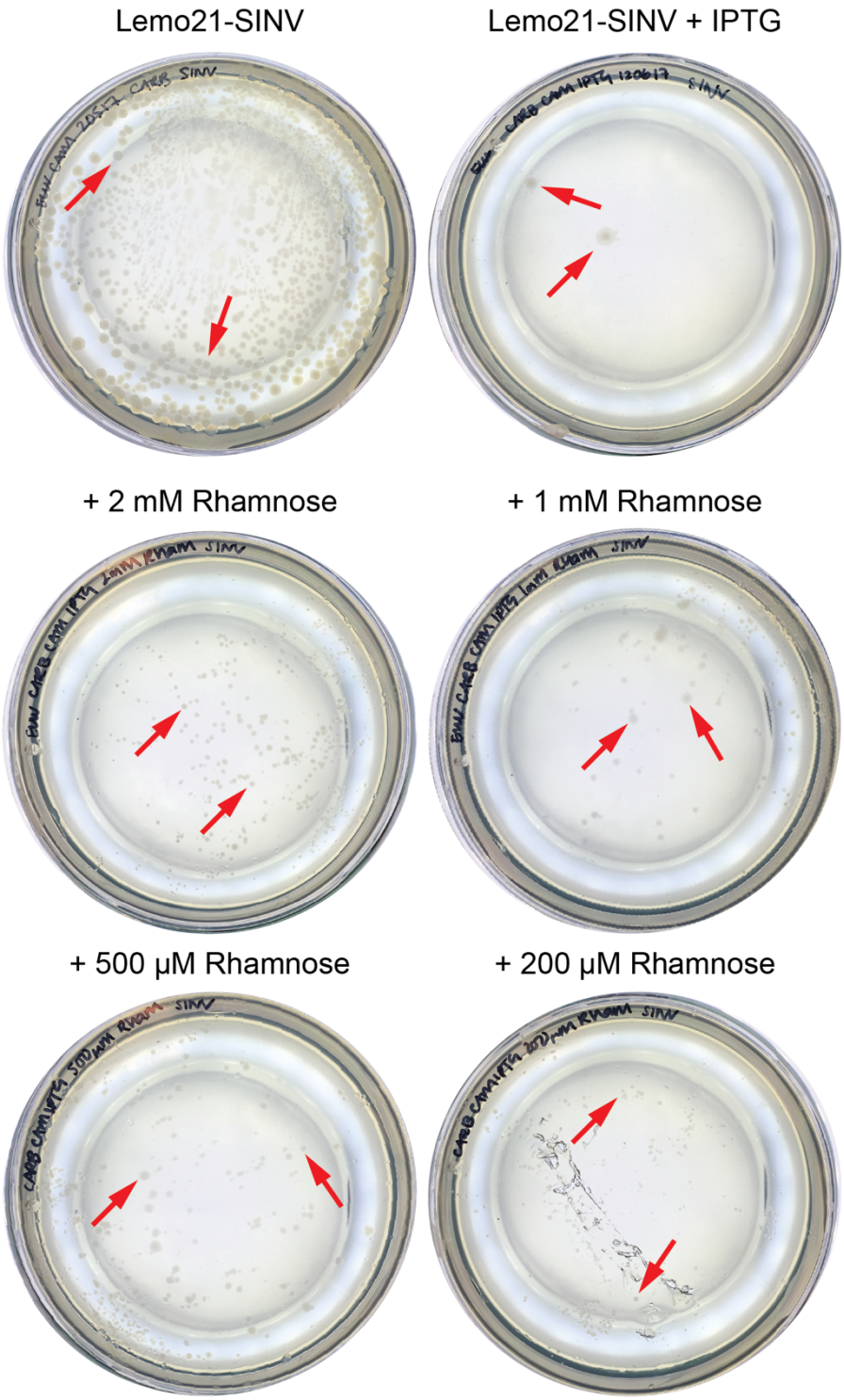


Figure E.3: Lemo21 expression of SINV in the presence and absence of Rhamnose. (Top) Photographs of Lemo21-SINV in the absence and presence of IPTG. **(Middle and bottom)** Lemo21-SINV in the presence of IPTG and indicated rhamnose concentration.

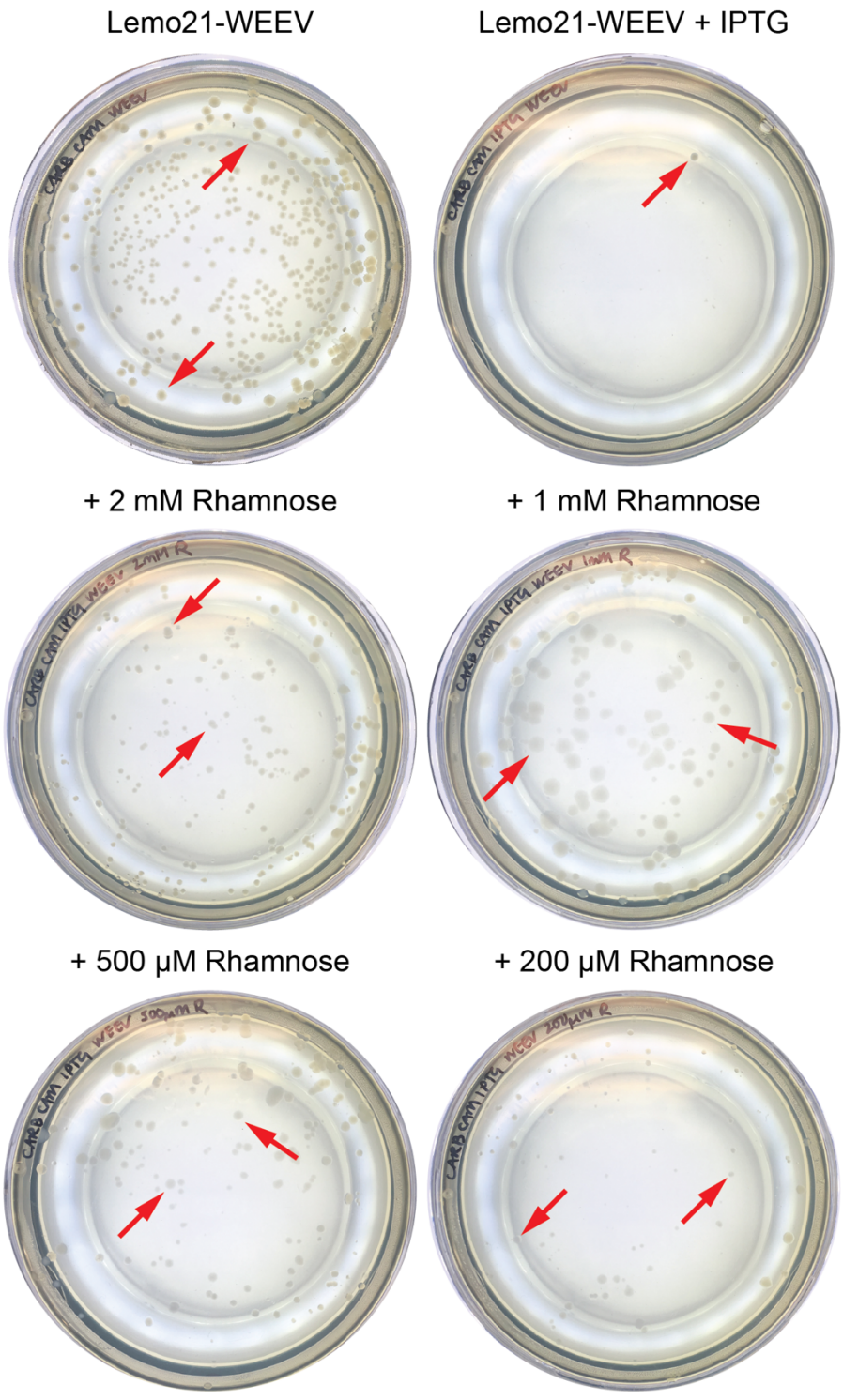


Figure E.4: Lemo21 expression of WEEV in the presence and absence of Rhamnose. (Top) Photographs of Lemo21-WEEV in the absence and presence of IPTG. **(Middle and bottom)** Lemo21-WEEV in the presence of IPTG and indicated rhamnose concentration.

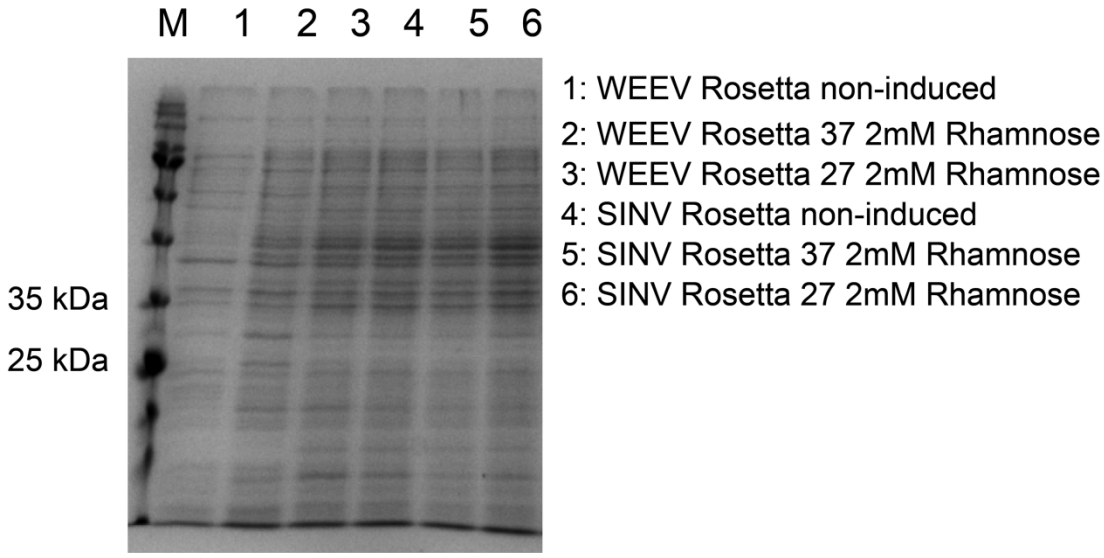


Figure E.5: Expression of SINV and WEEV core proteins using Rhamnose. SDS-PAGE of SINV and WEEV Cp expression in Rosetta cells at 37°C and 27°C with 2 mM Rhamnose. No protein expression is visible in the samples. M = BLUeye protein ladder (Geneflow).

Table E.1: Concise results from trypsin digest mass spec

Protein Group	-10lgP	Description
1	300.45	Structural polyprotein OS=Chikungunya virus (strain S27-African prototype) PE=1 SV=3
1	293.83	Structural polyprotein OS=Chikungunya virus (strain 37997) PE=1 SV=1
2	292.93	Structural polyprotein OS=Chikungunya virus (strain Nagpur) PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain 55989 / EAEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O81 (strain ED1a) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O157:H7 GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain UTI89 / UPEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain K12 / DH10B) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O8 (strain IAI1) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O1:K1 / APEC GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O9:H4 (strain HS) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain K12) GN=rplP PE=1 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O157:H7 (strain EC4115 / EHEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=rplP PE=3 SV=1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=rplP PE=3 SV=1

3	162.4	50S ribosomal protein L16 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rplP PE=3 SV =1
3	162.4	50S ribosomal protein L16 OS=Escherichia coli (strain SE11) GN=rplP PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain K12 / DH10B) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rpsD PE=3 SV= 2
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O157:H7 (strain EC4115 / EHEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain SE11) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O157:H7 GN=rpsD PE=3 SV=2
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O81 (strain ED1a) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain 55989 / EAEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O1:K1 / APEC GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain UTI89 / UPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O9:H4 (strain HS) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain K12) GN=rpsD PE=1 SV=2
4	127.37	30S ribosomal protein S4 OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=rpsD PE=3 SV=1
4	127.37	30S ribosomal protein S4 OS=Escherichia coli O8 (strain IAI1) GN=rpsD PE=3 SV=1

6	90.47	Thiol:disulfide interchange protein DsbG OS=Escherichia coli (strain K12) GN=dsbG PE=1 SV=2
6	90.47	Thiol:disulfide interchange protein DsbG OS=Escherichia coli O157:H7 GN=dsbG PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain K12) GN=rplR PE=1 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O157:H7 GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O81 (strain ED1a) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain UTI89 / UPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rplR PE=3 SV =1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain K12 / DH10B) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain SE11) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O9:H4 (strain HS) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O157:H7 (strain EC4115 / EHEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O8 (strain IAI1) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain 55989 / EAEC) GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli O1:K1 / APEC GN=rplR PE=3 SV=1
13	66.68	50S ribosomal protein L18 OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=rplR PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O8 (strain IAI1) GN=acpP PE=3 SV=1

5	64.07	Acyl carrier protein OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O9:H4 (strain HS) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain 55989 / EAEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain K12 / DH10B) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=acpP PE=3 SV=2
5	64.07	Acyl carrier protein OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain SE11) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli O81 (strain ED1a) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain UTI89 / UPEC) GN=acpP PE=3 SV=1
5	64.07	Acyl carrier protein OS=Escherichia coli (strain K12) GN=acpP PE=1 SV=2
5	64.07	Acyl carrier protein OS=Escherichia coli O157:H7 GN=acpP PE=3 SV=2
9	55.39	Pyruvate dehydrogenase E1 component OS=Escherichia coli O157:H7 GN=aceE PE=1 SV=2
9	55.39	Pyruvate dehydrogenase E1 component OS=Escherichia coli (strain K12) GN=aceE PE=1 SV=2
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O1:K1 / APEC GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O9:H4 (strain HS) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC) GN=rpsN PE=3 SV=1

14	52.3	30S ribosomal protein S14 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain UTI89 / UPEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rpsN PE=3 SV =2
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain K12 / DH10B) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain SE11) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli (strain K12) GN=rpsN PE=1 SV=2
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O157:H7 (strain EC4115 / EHEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rpsN PE=3 SV=1
14	52.3	30S ribosomal protein S14 OS=Escherichia coli O157:H7 GN=rpsN PE=3 SV=2
16	24.12	50S ribosomal protein L35 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rpmI PE=3 SV=1
total 105 proteins		

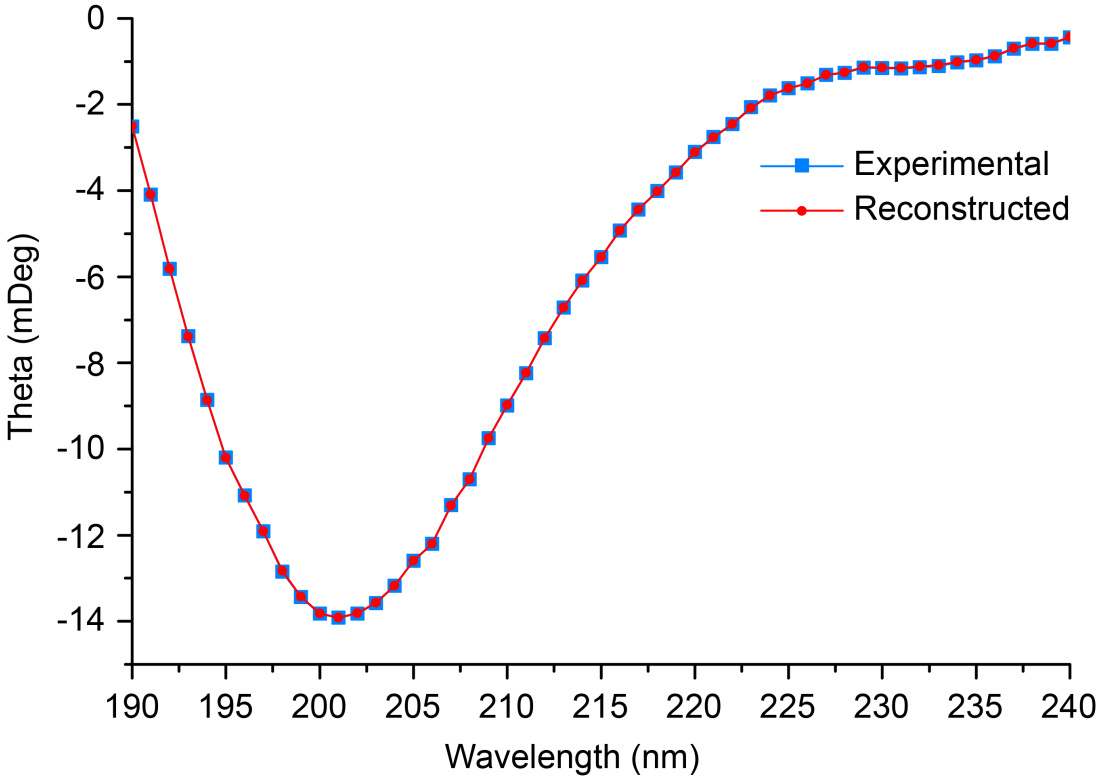


Figure E.6: CD analysis of recombinant SFV protein. Plot of experimental and reconstructed data in machine units using CDSSTR reference set 4 from 190-240 nm wavelengths. NRMSD = 0.015.

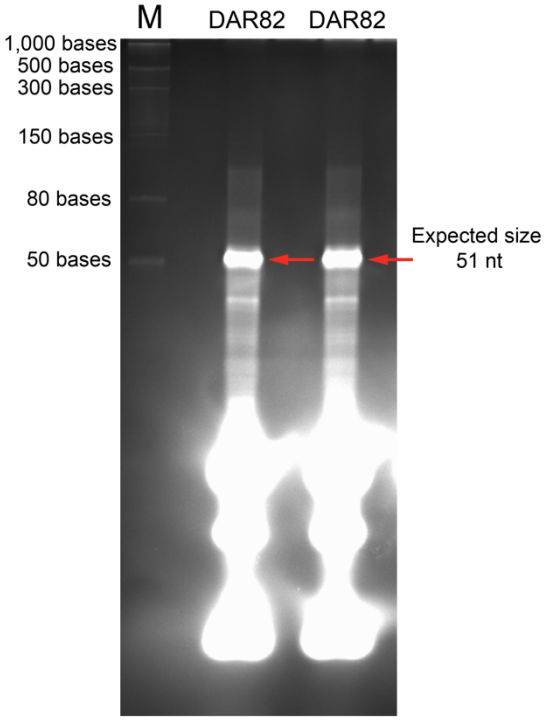


Figure E.7: Denaturing RNA PAGE of labelled DAR82. Denaturing PAGE for labelled DAR82. M = ssRNA ladder low range (NEB).

Appendix F: Fluorescence Measurements for Chapter 5

[This page has been left intentionally blank]

Table F. 1: Fluorescence measurements for 1:60 WT vs WT

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	9.19±0.16	9.08±0.18	48.96±2.40	61.06±1.65
R^2	0.9989	0.997	0.9996	0.9844
ƛ510 nm	8.9	0.9	14.8	2.8
ƛ617 nm	7.1	0.9	5.6	2.8

Table F. 2: Fluorescence measurements for 2:60 WT vs WT

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	8.72±0.16	8.53±0.17	48.30±2.36	59.08±1.32
R^2	0.9977	0.997	0.9951	0.9998
ƛ510 nm	5.2	2	24.3	0.9
ƛ617 nm	3.4	2	16	1

Table F. 3: Fluorescence measurements for 2:60 WT vs C4-WT

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	9.01±0.17	8.50±0.18	49.00±1.94	87.92±1.97
R^2	0.9978	0.995	0.9996	0.9997
ƛ510 nm	5.5	0.4	13.4	1.4
ƛ617 nm	15.7	8.8	33.5	11.3

Table F. 4: Fluorescence measurements for 2:60 WT vs ΔAXXA

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	8.85±0.16	8.59±0.18	48.36±1.76	197.25±16.89
R^2	0.9973	0.9939	0.9996	0.9997
ƛ510 nm	15.3	8.4	26.2	12.5
ƛ617 nm	3.2	2.4	18.5	3

Table F. 5: Fluorescence measurements for 2:60 WT-ΔAXXA vs ΔAXXA

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	8.99±0.17	8.48±0.18	48.85±2.97	246.70±4.54
R^2	0.9975	0.9943	0.9994	0.9931
ƛ510 nm	9.8	8.9	14	6.6
ƛ617 nm	4.8	4.4	8.4	5.3

Table F. 6: Fluorescence measurements for one shot WT vs C4-WT

	Peak 1	Peak 1 post Rnase	Peak 2	Peak 2 post Rnase
R_h	11.07±0.16	10.13±0.21	36.06±1.33	114.14±1.62
R^2	0.9983	0.9942	0.9996	0.9989
ƛ510 nm	12.5	3.1	21.6	11.3
ƛ617 nm	29.9	18.5	46.1	18.2