The Roles of RNA in the Assembly and Disassembly of Single-stranded RNA Icosahedral Viruses

Robert John Ford

Submitted in accordance with the requirements of Doctor of Philosophy

University of Leeds

Institute of Molecular and Cellular Biology

September 2012

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

Acknowledgements

Overall, I would say that I have had a blast studying for my PhD. After a rocky start, I have learned an awful lot about myself in times of adversity, and the meaning of knuckling down when times are tough. There are so many people that I need to thank, because without them, there is no way I would have made a success of my PhD. First and foremost, I would like to thank my supervisor, Peter Stockley. He put his faith in me, gave me a 'kick up the arse' when I needed it, and never lost his confidence in my ability. I would also like to thank David Bunka, who was always there (quite literally, always there) both professionally and personally. Dave, you went above and beyond mate. Thanks a lot. I would also like to thank Simon White for reading loads of my thesis, providing excellent feedback, and for at least trying to remain interested throughout. I would also like to thank you Si for making me look strong in the gym (weakling!). A ma-hussive thanks also goes to Amy Barker for performing the AUCs, synthesising the RNA oligos and for being the Queen of cake club. Without you, Mondays would suck! Thank you Saskia Bakker for introducing me to TCV, with whom I've shared many a nice evening with, carrying out STNV-B3 VLP crystallisation and for collecting the data at Diamond Light Source. On that note, thank you Arwen Pearson for taking me to Diamond, for her kindness and for buying me lunch on several occasions. Thanks to George Lomanossoff and Keith Saunders for providing the TCV-infected leaves, Robert Coutts for the STNV-C clone, to Peiyi Wang for TEM training and Neil Ranson for his help with figures. Thanks also to Alex Borodavka for providing the clones for MS2 RNAs, for help with loads of stuff in the lab, and for generally being very generous with his coffee. I would also like to thank Ruiyang Yan for helping me improve my rating at chess, and for being the only one with talent in our new band 'We

really are Scientists'. Thanks also to Nicola Derbyshire for her warmth, kindness and for letting me share my write up year with her. Thanks also to past members of the lab, in particular to Olga Platonova for her no-nonsense approach in making me clean up after myself in the lab and for useful discussions in regards to protein purification. Thanks to my Undergraduate and Masters students for their help in the lab, in particular Emma Wroblewski for helping push forward the TCV project. On a more personal level, I would also like to thank my mum and dad for being great role models, and for helping me to fulfil my potential. I'd also like to thank my brother, Dan, for sharing many a time down the pub and letting me beat him many times at pool and darts. Thanks also to my friends, too many to thank individually, but without whom I may have gone insane. Thanks also to my lovely girlfriend, Jo, for being patient during her time as a 'thesis widow' and for preparing countless lovely teas in Tupperware's, many of which have spent most of the year on my desk! After all these thanks, I feel like I'm picking up an Oscar. I hope I haven't missed anyone out, although I'm sure I will have. As a backup, my final thanks go to all you people I have forgotten to thank individually, and to anyone who enjoys reading my thesis.

Abstract

Single-stranded, positive-sense icosahedral viruses are major pathogens in every kingdom of life. Despite this, their capsid assembly and uncoating mechanisms remain poorly understood. This work describes these processes in two model systems; satellite tobacco necrosis virus and turnip crinkle virus. For satellite tobacco necrosis virus, several aptamers were previously raised against the coat protein, where each aptamer folded into stem-loops displaying the motif AXXA. Aptamer B3 contained the strongest sequence similarity to the cognate genome, including a 10/10 contiguous stretch. Capsid assembly using the purified coat protein shows that RNA is critical for capsid assembly, and that stem-loops displaying the motif AXXA can efficiently trigger this process. There is a clear preference for this loop motif, which is unaffected by the sequence of the base paired stem. The structure of the B3-encapsidated virus-like particle has been solved by X-ray crystallography to 2.3 Å, together with a lower resolution map encompassing the RNA. The presence of B3 results in an extension of the N-terminal helices by roughly one and a half turns, such that residues 8-11 that are disordered in all previous X-ray structures are now visualised, including R8 and K9. The binding of B3 facilitates charge neutralisation and trimer formation in the coat protein, resulting in the assembly of a *T*=1 capsid. This assembly mechanism is consistent with additional assembly studies using longer RNAs, in which the first step in assembly is genomic compaction. This compaction event is driven by multiple binding events of coat proteins with packaging signals in the form of stem-loops displaying the preferred loop sequence.

In turnip crinkle virus, a putative disassembly mechanism has been suggested. Expansion and proteolysis mediates extrusion of the viral genome, such that the formation of “striposomes”, which are thought to be polysomal arrays of ribosomes on extruding RNA, can be visualised by TEM. Purification of proteolysed capsids revealed that the cleaved coat proteins become dissociated and the remaining protein shells lose their icosahedral symmetry, often appearing to begin release of RNA from unique sites in the absence of ribosomes. These results explain why coat proteins are essential for wild-type infections because they facilitate a ribosome-mediated uncoating mechanism avoiding host RNA silencing. The results in this thesis suggest new paradigms for capsid assembly and uncoating, which may be exploited by other members of the same family of viruses, especially those having similar coat protein folds.

Contents

Figures

Tables

Abbreviations

(+)ssRNA	positive-sense single-stranded RNA
A	adenine
Å	Ångström (10^{-10} m)
APS	ammonium persulphate
bp	base pair
BPMV	bean pod mottle virus
C	cytosine
CCD	charge-coupled device
CCMV	cowpea chlorotic mottle virus
CNV	cucumber necrosis virus
CP	coat protein
CPMV	cowpea mosaic virus
cryo-EM	cryo-electron microscopy
ddH_20	double distilled water
DNA	deoxyribonucleic acid
ds	double-stranded
EDTA	ethyldiaminetetraacetic acid
FHV	flock house virus
G	guanine
HIV	human immunodeficiency virus
IPTG	isopropyl-β-D-thiogalctoside
MOPS	3-[N-morpholino]propanesulphonic acid
mRNA	messenger RNA
nt	nucleotide

PAGE	polyacrylamide gel electrophoresis
PaV	pariacoto virus
RCNMV	red clover necrosis mottle virus
RdRp	RNA-dependent RNA polymerase
RNA	ribonucleic acid
S	svedberg
SBMV	southern bean mosaic virus
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SELEX	systematic evolution of ligands by exponential enrichment
ss	single-stranded
STMV	satellite tobacco mosaic virus
STNV	satellite tobacco necrosis virus
T	thymine
T	triangulation number
TBE	tris-borate EDTA
TBSV	tomato bushy stunt virus
TCV	turnip crinkle virus
TEM	transmission electron microscope
TEMED	N,N,N',N'-tetramethylethylenediamine
TR	translational repressor
Tris	tris (hydroxymethyl) amino methane
TYMV	turnip yellow mosaic virus
U	uracil

VLPs	virus-like particles
w/v	weight/volume
w/w	weight/weight
α_2M	α_2 Macroglobulin

1. Structure and function of ssRNA icosahedral viruses

Single-stranded positive-sense RNA viruses are organised infectious agents and obligate intracellular parasites which infect host organisms from every kingdom of life[1]. They do not encode their own replication machinery within their genome, so therefore rely on their host for viral progeny production, often to the detriment of their host. At their simplest, they are composed of one molecule of nucleic acid and multiple protein subunits arranged into a protective container. It is this protective container which helps keep the genome intact and transports the genome between host organisms.

Virus research has been at the vanguard of molecular biology for some time. Many of the tools of molecular biology (such as restriction enzymes[2]) and basic understanding of host cell biochemistry have been gained from virus research. Underpinning the relevance virus research has to the fields of molecular biology and medical biochemistry are the technologies of phage display[3], heterologous gene expression[4], targeted drug delivery[5] and vaccine development[6], to name but a few. Virus assembly and disassembly are events in the virus life cycle where their understanding should greatly contribute to the elucidation of new drug targets against viruses. In addition, the self-assembling properties of viruses are of great interest in the fields of bionanotechnology[7].

This thesis aims to increase the understanding of the processes of virus assembly and disassembly in two virus model systems: Satellite Tobacco Necrosis Virus (STNV) and Turnip Crinkle Virus (TCV). These are model systems for the family of positive sense, single stranded ((+) ss) RNA icosahedral viruses.

1.1 Model systems for (+) ssRNA virus research

According to the International Committee on the Taxonomy of Viruses (ICTV), ssRNA icosahedral viruses represent ~15% of the viruses in the biosphere[8]. Of these viruses, several are responsible for the diseased states of animals and humans such as the common cold, gastroenteritis, hepatitis, polio, foot and mouth disease, yellow fever and rubella. Pathogenic viruses represent a serious health hazard to any researcher investigating them. This potential hazard can be removed by working with viruses and virus-like particles that are similar in characteristics but are non-infectious to humans. Detailed information concerning (+) ssRNA icosahedral viruses that infect plants and bacteria has been ascertained, on their structures, life cycles and virus assembly/disassembly characteristics. However, much of this research has been directed towards the protein capsid, neglecting the many roles of the RNA genome. To this end, these viruses represent ideal model systems for further study.

1.2 Virus structure

Viruses differ vastly in architecture and size, ranging from filamentous or rod shaped to spherical, and may be enveloped or non-enveloped. The largest virus known is called *Megavirus chilensis* (or just Megavirus)[9], which has a total diameter of ~680 nm. The smallest virus isolated to date is satellite tobacco mosaic virus (STMV) with a diameter of 16-17 nm. The vast difference in sizes between these two viruses also translates to their complexities (see Figure 1.1 and Table 1.1). Megavirus is composed of a multitude of different proteins, contains hundreds of surface 'hairs', and has the largest known genome of all viruses (>1.2Mbp) which encodes seven aminoacyl-tRNA synthetases[9]. STMV,

by contrast, has a genome size of 1058 nt which encodes just one protein; that of its capsid protein[10].

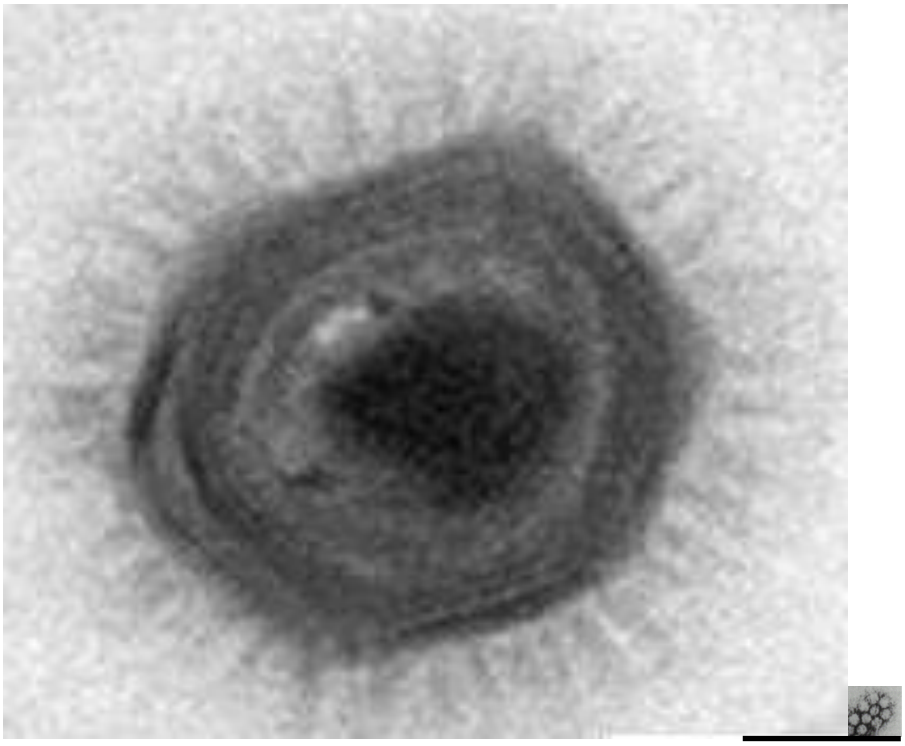


Figure 1.1: Electron micrographs illustrating the size difference between Megavirus and STMV.
The scale bar (300 nm) is shown on the far right to emphasise the differences in size.

Despite these striking differences, all viruses do share common features, such as a protecting shell known as the capsid, a nucleic acid genome, and their inability to replicate away from their hosts. An exhaustive review of the differences in virus structure is not relevant here. Instead, the overall principles of virus structure will be discussed, the methods used to characterise these structures, and the symmetry of icosahedral viruses.

Table 1.1: Differences between viruses at the opposite ends of their complexities

	Megavirus chilensis	STMV
Capsid diameter (nm)	440	16-17
Genome size	1,259,157 bp	1058 nt
Total genes	~1120 proteins	One protein
Complexity	Contains several hundred surface spikes; codes for 7 aminoacyl-tRNA synthetases; contains two lipid membranes; contains genes for sugar, fatty acid and protein synthesis	Has no additional features

1.2.1 Principles of virus structure

The principles of virus structure have arisen from the requirements of the virus imposed on them from their evolutionary constraints. What are these requirements? The first and simplest of these is the general formation of a shell which encompasses the genome with protective functionality as it passes from host-to-host. The second is one of simple economics: the information which codes for the capsid must not exceed the information present within the genome. The third is that they must encode strategies for entry and exit from their host. The final requirement is that of the principles of protein folding, conformational switching and thermodynamics. Viruses must encode all the information necessary for a variety of complex tasks. These include packaging their genomes, strategies for entry and uncoating, presenting their genome to the host's replication machinery, and also to allow the virus to exit the host cell to infect another. The progress made on the elucidation of virus structure in atomic detail now allows us a tantalising glimpse of intrinsic similarities across

various families of viruses. In the next subsection, the methods of how viruses are characterised will be discussed as well as the language needed to describe them.

1.2.2 How virus structure is studied

Electron microscopy is easily the most efficient way of determining the general morphology of a virus particle. However, the coarseness of staining methods used limits the resolution which can be observed. Negative staining of viruses with uranyl acetate, potassium phosphotungstate or other related electron dense compounds embeds the virus particles. The electron beam destroys the particle itself, but the stain leaves a ‘cast’ around the virus particle. If the particle is fully covered by stain, that is, on both the upper and lower surfaces of the particle, then contrast of the finer details can be difficult to distinguish[11].

In addition to negative stain, viruses can be preserved by rapidly freezing in liquid nitrogen or helium. This permits the visualisation of electron scattering directly from the structures in the particles themselves and not just from the cast created from a negative stain. Furthermore, quantitative methods for image analysis have been applied to this method. This technique is called cryo-electron microscopy (cryo-EM)[12]. An advantage of using cryo-EM is that regular images can be selected from a field of heterogeneous particles and aligned to produce an image reconstruction of a virus particle. Image reconstructions are possible by the averaging of hundreds or thousands of particles that are all uniform. Below is a gallery of different virus structures solved by cryo-EM which illustrates the diversity of icosahedral virus structures.

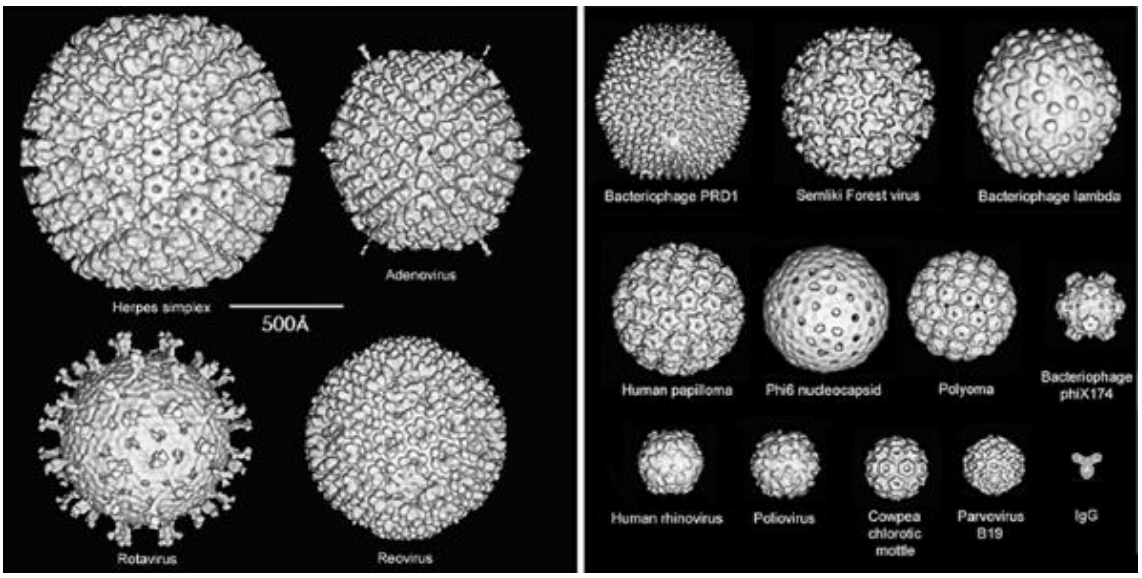


Figure 1.2: A gallery of image reconstructions of icosahedral virus particles by cryo-EM. Images illustrate the variety of sizes and features of viruses, even though the particles are highly regular. The particles are labelled; Herpes simplex virus is nucleocapsid only. An image of an immunoglobulin has been included for scaling. Scale bar = 50 nm. Figure taken from Field's Virology, 2006[13].

When particles of an asymmetric nature are viewed, cryo-EM cannot be employed to generate an image reconstruction using icosahedral averaging. Instead, a tomographic tilt series of images can be processed to generate an image. Tomographic reconstructions generally do not provide high resolution information due to electron damage occurring whilst generating the data[14; 15].

Information obtained from electron microscopy methods are generally not comparable to atomic resolution (~5-20 Å). Higher resolution images can be achieved by X-ray crystallography, if purified crystals of the virus particles can be prepared. Virologists have been purifying crystals of viruses for the analysis by X-rays since the 1930's. The first virus structure to be elucidated was Tobacco Mosaic Virus (TMV), where the crystal structure showed that it was composed from multiple subunits of a repeating monomeric protein[16]. These data sets, as well as others, led to Watson and Crick hypothesizing that most

simple viruses are composed of one or a few identical protein subunits[17]. X-ray crystallography has since been used to characterise a large number of different viruses, revealing both similarity and differences in architectures and size between different viruses. Addressed in the next subsection are the similarities in symmetry and the nomenclature used to address this symmetry.

1.2.3 Symmetry in viruses

The simplest type of shell is one that is composed of multiple identical subunits which can assemble spontaneously. The observation of self-assembly was first elucidated with *in vitro* assembly of TMV[18]. This led Crick and Watson to hypothesize that the structures of viruses would be highly symmetrical[17]. The arguments that contribute towards this hypothesis are simple: Genetic efficiency i.e. multiple identical subunits coming together to form a well defined shape; and *in vitro* assembly from these identical subunits resulting in particles of a defined size. Together, these observations suggested that the interactions between subunits are specific and well ordered. In general, identical subunits with specific interactions produce symmetrical objects.

But what exactly is meant by symmetry in this context? The precise definition of symmetry being discussed here involves an operation, such as rotation, which brings the object of study into view on another plane. For example, if the 'three-legs of man' (the national symbol for the Isle of Man, shown in Figure 1.3) was rotated 120° or 240°, it would not be possible to distinguish one leg from another. The full symmetry of an object is dependent on a full collection of symmetries applying to it. In the case of virus capsids, these operations can be rotations, translations or combinations of both. A symmetry axis with a 180°

rotation is called a two-fold symmetry axis; three 120° rotations are called a three-fold symmetry axis; and five rotations of 108° are called five-fold symmetry axes. The distinction between shape and symmetry is that a shape describes the geometry of its outline, whereas symmetry describes the rotational operations of the shape. For example, a triangle and the three-legs of man have the same symmetries but are of course completely different shapes (see Figure 1.3).

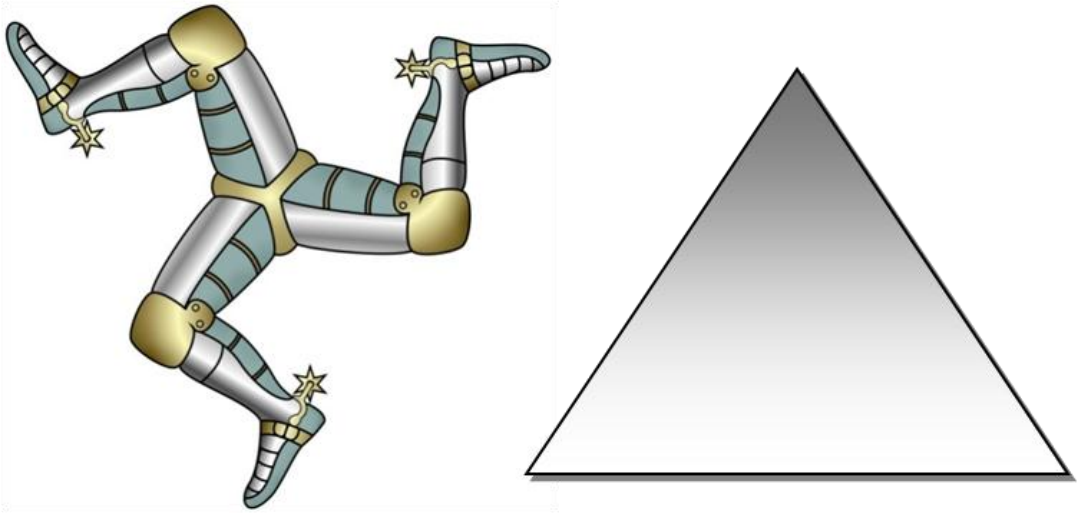


Figure 1.3: Three-fold symmetry
The three-legs of man and an equilateral triangle both share three-fold symmetry.

Closed shells that are composed of identical subunits which interact specifically and which are roughly spherical in shape can only have one of five geometries: the cube, the tetrahedron, octahedron, dodecahedron or the icosahedron. Of these geometries, the icosahedron is the most efficient of the designs for spherical virus architecture, i.e. they contain the largest number of subunits to make a container of a given size, and thus require the minimum coding genetic capacity. With these observations, it is not surprising that the extensive use of the icosahedron in virus architecture is seen in nature.

1.2.4 Icosahedral symmetry

The icosahedrons, shown in Figure 1.4, look down the five-fold, three-fold and two-fold symmetry axes, respectively. Placement of an asymmetric object on one face of the icosahedron leads to the generation of 59 others. One such unit, one sixtieth of the entire shell, can therefore be described as an icosahedral asymmetric unit. In other words, the asymmetric unit is the fundamental jigsaw piece used to construct an icosahedron.

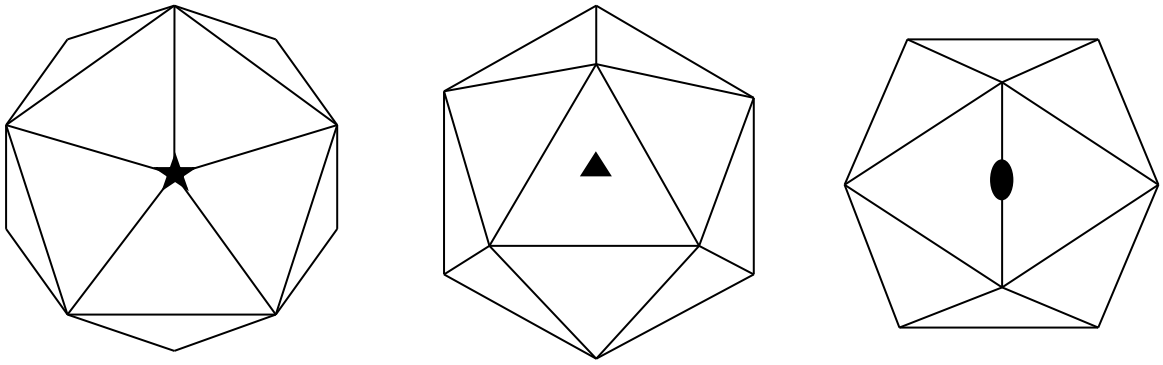


Figure 1.4: Icosahedral symmetry
Looking down the symmetry axes of an icosahedron. The star represents the five-fold axis, the triangle represents the three-fold axis and the ellipse represents the two-fold axis.

The asymmetric unit, in the context of building a capsid, is the coat protein. In the simplest icosahedral viruses, such as the plant satellite viruses, the capsid is composed of 60 identical coat proteins which form capsids of a perfect icosahedron. But what about larger icosahedral viruses observed in nature? X-ray crystallography confirmed that these viruses adhered to icosahedral symmetry but were not necessarily composed of just 60 subunits. In order to explain this phenomenon, Caspar and Klug defined the theory of quasi-equivalence[19]. They predicted that capsids could be composed of an ever increasing number of subunits and that these subunits could adopt similar yet distinct conformations. The protein contacts between each of the subunits need

not be identical so larger capsids can be made by increasing the number of identical proteins in the capsid adhering to icosahedral symmetry. Caspar and Klug defined the icosahedral organisation of capsids with more than 60 protein subunits by the triangulation number, *T*. *T* is defined by the equation: $T = h^2 + hk + k^2$ where *h* and *k* are integers, which define the position of the five fold axes relative to one another on the surface of a lattice (see Figure 1.5)[19; 20]. Hexamers are generally considered to be planar, whereas pentamers are considered convexed, propagating curvature into the hexagonal lattice when they are inserted. The formation of an icosahedron is generated by inserting 12 pentamers into relevant positions on the hexagonal lattice. The theory of quasi-equivalence has since become a paradigm for predicting virus structure[20].

1.2.5 Structure of icosahedral viruses

The first icosahedral virus structure to be elucidated at atomic resolution was Tomato Bushy Stunt Virus (TBSV)[21]. This structure adhered to Caspar and Klug's theory of quasi-equivalence, producing a capsid of *T*=3 icosahedral symmetry with 180 copies of identical coat protein subunits arranged in three different conformations, termed A, B and C. Since then, the structures of a multitude of different viruses have been elucidated by X-ray crystallography, and also many from cryo-EM, where some of which have revealed similarities in virus structures infecting hosts from every kingdom of life.

Historically, protein-protein contacts were thought to be the principle driving force behind ssRNA virus assembly supported by the spontaneous generation of capsids devoid of any nucleic acid[22; 23; 24]. Now there appears to be an accumulation of evidence which suggests that the RNA genome plays multiple

roles in a virus infection and is not just a passenger going along for the ride (reviewed in [1]). Therefore the next few subsections address these observations, beginning with RNA structure *in capsa*.

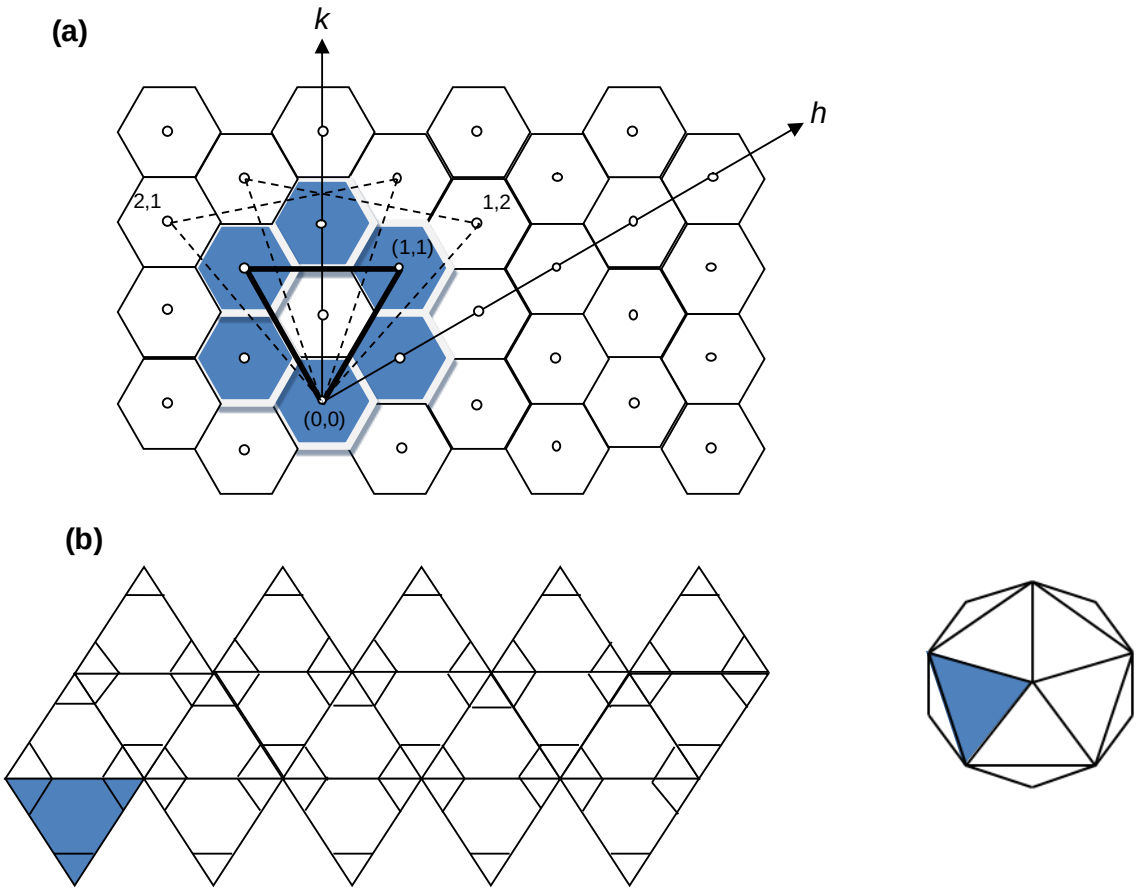


Figure 1.5: Geometric principles for generating icosahedral quasi-equivalent surface lattices.
(a) Hexagons are considered planar when on a surface lattice (they form a flat sheet, as shown above) and pentamers are considered convex, introducing curvature into the lattice whenever they are inserted. The icosahedral shell is composed of hexamers and pentamers, and is generated by inserting 12 pentamers into an array of hexamers within defined positions on the surface lattice. To construct a model of a particular quasi-equivalent surface lattice, one face of the icosahedron is generated in the hexagonal net. The origin of that is replaced by a pentamer and the (h,k) hexamer is also replaced by a pentamer. The third hexamer replaced with a pentamer is identified by its three-fold symmetry. The icosahedral face shown in bold defines a *T*=3 lattice (h=1, k=1). Two alternative equilateral triangles are shown as dotted lines for *T*=7 lattices, where h=2, k=1; or h=1, k=2
(b) The hexagonal lattice of a *T*=3 particle from (a) (outlined in blue) and the *T*=3 lattice face coloured blue. The ASU is one third of this face, with the icosahedron being composed of 20 of these triangles, as shown.

1.3 Structure of RNA within virions

Even though techniques for generating high resolution images have improved in recent years, the roles of RNA in virus assembly are still poorly understood. Underpinning this is that there is no high resolution structure available for the complete RNA genome *in capsa* from any virus. Predominantly, the focus has been on how protein-protein interactions, typically of hundreds of subunits, give rise to stable virus particles. The principal reason for this is that experimental techniques for determining protein structure have been well characterised, whereas information on the nucleic acid component is more difficult to obtain. This has created a highly protein-centric viewpoint to virus structure and assembly[25; 26].

However, as knowledge of specific virus structures have increased over the years, the focus has turned to exploring the contribution of the genome during virus assembly. There are now several examples of viruses where the packaging of a subsection of their genomes can be visualised at high or moderate resolution[27; 28; 29; 30; 31]. Furthermore, biochemical and genetic analyses have given glimpses of specific RNA: protein interactions that are required to initiate the assembly process[32]. The next few subsections will therefore place the emphasis on these observations, the structural studies of viral RNA genomes and the roles of the genome in virus assembly.

1.3.1 Genome packaging of icosahedral (+) ssRNA viruses

A significant proportion of all viruses known have icosahedral symmetry with a (+) ssRNA genome, and include important human pathogens such as the *Picornaviridae*, the *Caliciviridae*, *Astroviridae*, as well as members of the

Togaviridae and *Flaviviridae* families[1]. Upon infection and production of viral progeny, several stages are undertaken. These include the disassembly of the nucleocapsid *in vivo*, replication of their genome, expression of their coat protein monomers, and recognition of their replicated genome before self-assembly. Many factors play a role in specific recognition and packaging of viral RNA. These include the sequence and the secondary structure of viral RNA which serve as recognition signals and the specific, high affinity interactions that the capsid protein has with these recognition signals. Recognition sequences and their adopted secondary structure for many viruses have been determined[33; 34; 35; 36], although the structural domains that are involved in RNA binding and packaging remain unclear. Many viral coat proteins contain a run of basic residues towards the N- or C-terminus that are believed to be of importance in RNA packaging. Furthermore, host cell macromolecular complexes are also likely to play a part.

1.3.2 Fundamental problems of visualising packaged RNA in high resolution

The amount of RNA that can potentially fit into a capsid is intrinsically governed by the amount of space available in the interior of each respective viral capsid. The fraction of space the RNA genome takes up varies greatly between different (+) ssRNA viruses.

General properties of (+) ssRNA genomes can be observed at a low or medium resolution by using techniques such as solution X-ray scattering[37], neutron diffraction[38], Raman spectroscopy[39] and Cryo-EM[40].

Structural studies by X-ray crystallography of viruses often do not reveal structural information of the RNA. This is because crystal structures are governed by the surface contacts of the protein capsid and these occur independently of RNA encapsidation. Furthermore, equivalent icosahedral viruses have 60-fold symmetry; each protein subunit exists in an identical surface environment, even in the presence of differing orientations of interior RNA[1]. Therefore, a typical virus crystal may not contain an ordered structure of the genome. Moreover, data from RNA structure is often lost due to the averaging process in structural determination. Regardless of these fundamental problems, there has been progress made due to both advancements in technology and to serendipity[1]. Examples of such are to be discussed in the forthcoming subsections.

1.3.2.1 Bean Pod Mottle Virus

Bean pod mottle virus (BPMV) is a member of the *Comoviridae* family, which are a group of viruses with icosahedral symmetry. It contains a bipartite (+) ssRNA genome and its capsid contains pseudo *T*=3 quasi-equivalent symmetry. The two (+) ssRNA strands, RNA1 and RNA2, differ in size of 6kb and 3.5kb, respectively. The protein capsid of BPMV is composed of two lots of 60 protein subunits; one of the subunits is smaller and contains one jelly roll domain whereas the second subunit is larger and contains two jelly roll domains.

BPMV was the first virus particle that when visualised by X-ray crystallography, a portion of the packaged genome was identified[27]. A portion of RNA2 was observed that revealed density of six well-ordered ribonucleotides about the three-fold axis of the virion (see Figure 1.6). The conformation of these ordered

structures approximates that one strand is arranged as an A-form RNA helix where the bases are stacked on top of one another. However, because the electron density is icosahedrally averaged, the actual structure that the RNA adopts is unknown. Under recent scrutiny of the helix structure, it suggests that the sequence is non-random and may correspond to residues A-Pu-Py-Py-Py-X[41]. This suggests that there may be multiple copies of this consensus sequence distributed along the length of RNA2. Analysis by bioinformatics suggests that this is the case. Almost 200 sequences appear to fit the consensus sequence in RNA2, albeit with no regular spacing or patterning between them[1].

The interactions between the RNA and the capsid appear to be mediated by contact with residues on the N-terminus of the subunit. These interactions are mostly governed by water molecules and include contacts to both bases and the sugar phosphate moieties on the RNA molecule. The RNA-capsid interactions appear to give order to the first 19 residues on the L subunit, which are disordered on capsids lacking RNA[41].

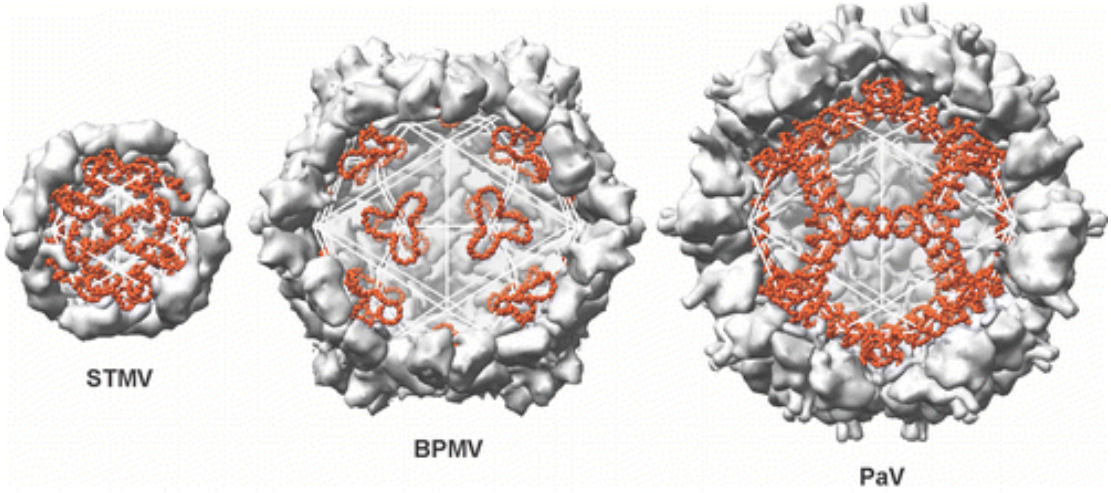


Figure 1.6: Cut-away view of STMV, BPMV and PaV capsids give rise to icosahedrally ordered RNA genomes. The capsid is represented in grey whereas the RNA is in red. Taken from Schneemann *et al*[1].

Further to the six well resolved RNA nucleotides on RNA2, five additional nucleotides could also be modelled into the electron density, albeit with less confidence. With these additional bases, this generates a loop structure of 11 nucleotides, three of which form a trefoil structure at each of the 20 three-fold contacts on the virus particle (see Figure 1.6).

The total amount of the BPMV genome that can be visualised by X-ray crystallography is about 660 nt, which corresponds to about 20% of the genome[27]. X-ray crystallography of BPMV particles containing just RNA 1 revealed an identical structural arrangement, thus it may be that the RNA trefoils represent an important structural arrangement, where this structure initiates the assembly of the virion by acting as specific signals for RNA packaging[1].

1.3.2.2 Satellite Tobacco Mosaic Virus

Satellite tobacco mosaic virus (STMV) has a *T*=1 icosahedral capsid arranged from 60 identical subunits of a 17.5 kDa coat protein. The diameter of the virus particle is approximately 16-17 nm and as a result is more restricted in its ability to package its (+) ssRNA genome.

The three-dimensional structure of STMV was solved to a high resolution of 1.8 Å by X-ray crystallography[10; 29]. The resulting electron density revealed density consistent with double helical segments of RNA at the two-fold contacts of the virion (see Figure 1.6). Each segment contains nine base pairs, which have been modelled as complementary A:U pairings, and an unpaired stacked nucleotide at the 3' end. In addition, a nucleotide not part of the segment but in

close propinquity was detected. These helical segments of dsRNA lie parallel to the 30 edges of the capsid and resemble an A-form RNA conformation[1].

The densities of the helical segments detected are strongest towards the two-fold axis but then weaken towards the three-fold axis of symmetry. This suggests that the RNA takes different paths into the interior of the STMV particle at these points, and therefore becomes invisible as they no longer conform to the icosahedral symmetry of the virus particle.

The double helical segments interact intimately with the interior of the STMV capsid. A dimer of STMV subunits supports the RNA, effectively functioning as a dsRNA binding protein. These interactions are primarily governed by non-specific interactions with the ribophosphate moieties of the RNA, with many contacts being bridged by water molecules. It is interesting to note that unencapsidated STMV RNA has a secondary structure of 'tRNA-like' conformations at the 3' end governed by around 200 nucleotides[42]. This structure does not seem to be present after encapsidation, suggesting that RNA is refolded in the packaging process.

Recently, Schroeder *et al* used chemical probing and computational modelling to generate secondary structure predictions that are consistent with the crystallography data[43]. These results showed all possible hairpins that are consistent with the experimental data. From these results, a co-transcriptional and assembly hypothesis was generated, where the combinations of hairpins that were most consistent with the experimental data were used as the most representative structure of the ensemble. Multiple packaging solutions could

provide an evolutionary advantage to viruses; therefore it may be that the RNA structure can be found in multiple orientations *in capsa.* Harvey and co-workers have since determined an all-atom model using this data and that of the crystal structure as constraints for the model. To date, this is the first all-atom model for a complete virus to be elucidated, albeit by using a coarse grained approach with the RNA genome a beads-on-a-string model[44].

1.3.2.3 Pariacoto Virus

Pariacoto virus (PaV) is an insect virus in the *Nodaviradae* family, which is a group of viruses with a bipartite (+) ssRNA genome and a *T*=3 icosahedral capsid. The three-dimensional structure of PaV has been solved to a high resolution of 3 Å by X-ray crystallography[45]. From the X-ray crystal structure, a dodecahedral cage of duplex RNA can be observed (see Figure 1.6), which indicates that the (+) ssRNA bipartite genome is reorganised upon encapsidation[45; 46]. These RNA duplexes contain 24 base paired segments, where the ends of the RNA duplexes extend towards the three-fold contacts of the virion. The total RNA that can be observed corresponds to approximately 35% of the total viral genome[45]. Similarly to STMV, the RNA density is weaker towards the three-fold contacts, indicating that RNA duplexes drop into the interior at some of these positions.

The coat protein interactions with the duplex RNA are mainly of an electrostatic nature, which involve residues from all three quasi-equivalent subunits. Although the structure of each subunit is chemically identical, they form different contacts in their quasi-equivalent locations (see Figure 1.7)[45].

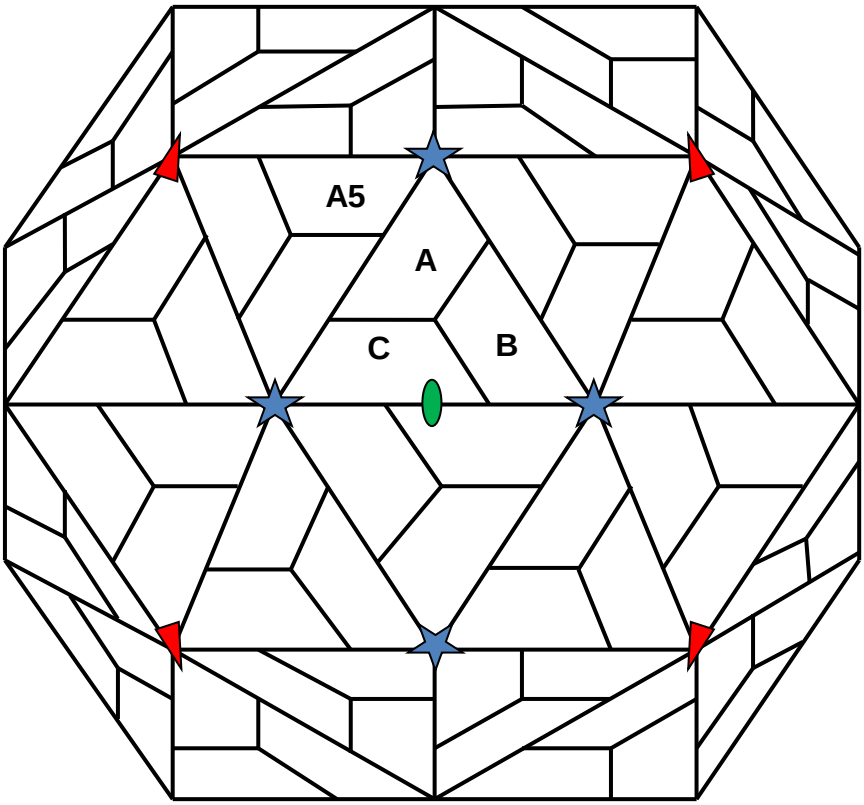


Figure 1.7: Schematic representation of the *T*=3 quasi-equivalent lattice. The positions of the icosahedral two-fold (green oval), three-fold (red triangles) and five-fold (blue stars) axes are indicated. Subunits A5 to A related by icosahedral five-fold symmetry.

The interactions are most extensive with the N-terminus of the A subunits, which form pentagons at the five-fold axis of symmetry (see Figure 1.8)[1]. The N-terminus of the A subunits extend all the way round the edge of the icosahedron, where it inserts itself into the major groove of the duplex RNA[45]. Numerous positively charged amino acids such as arginine and lysine form a positively charged surface, which functions to partly neutralise the duplex RNA backbone.

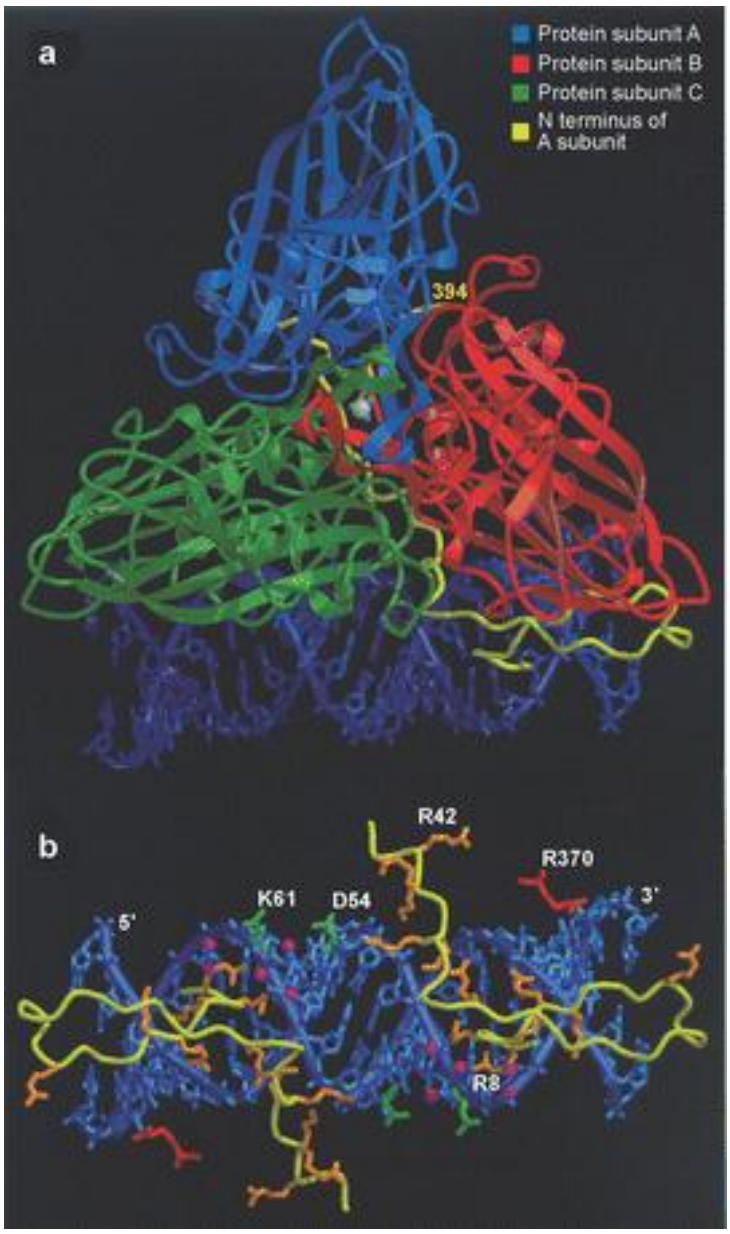


Figure 1.8: (a) Structure of the asymmetric unit of PaV. The ordered duplex RNA is represented as a stick model and comprises 24 base pairs modelled as A:U.

(b) Close-up view of ordered duplex RNA and its interaction with side chains from the three quasi-equivalent coat protein subunits and their two-fold related partners. Most contacts are with amino acid side chains from the N terminus of the A subunit. In addition, there are two side chain contacts from the B subunit and one side chain contact from the C subunit. Figure taken from Schneemann *et al*[1].

The mechanism of PaV *T*=3 viral assembly, like other Nodaviruses, involves the duplex RNA to function as a molecular switch[45; 47; 48]. This molecular switch operates at the two-fold contacts between the A subunits, where the N-terminal arms force the interacting coat proteins to form a flat contact. However, the duplex RNA and the ordered peptide arms are absent at the quasi-twofold contacts, allowing these subunits to form bent contacts. This exact alternation between bent and flat conformations forms the basis of correct *T*=3 capsid formation.

1.3.3.4 MS2

MS2 is a bacteriophage that infects *E. coli* and is part of the *Leviviradae* family. The three-dimensional structure of MS2 was characterised by X-ray crystallography and refined to a resolution of 2.8 Å[49; 50]. The capsid contains 180 proteins arranged as 90 non-covalent dimers with quasi-equivalent *T*=3 icosahedral symmetry. The capsid proteins are arranged as three conformers, A, B and C, which arrange into dimers of A/B and C/C, respectively. The main conformational difference between the three different subunits is in the FG loop (a 15 residue loop in between the F and G β-strands), which mediate either a five-fold or quasi six-fold contact

Despite high resolution determination of the MS2 capsid, RNA structure inside the MS2 bacteriophage was not seen by X-ray crystallography. To obtain information about the RNA component, the lower resolution technique of cryo-EM was employed[40], and refined[51]. In the refined cryo-EM model, approximately 90% of the genome was observed (see Figure 1.9).

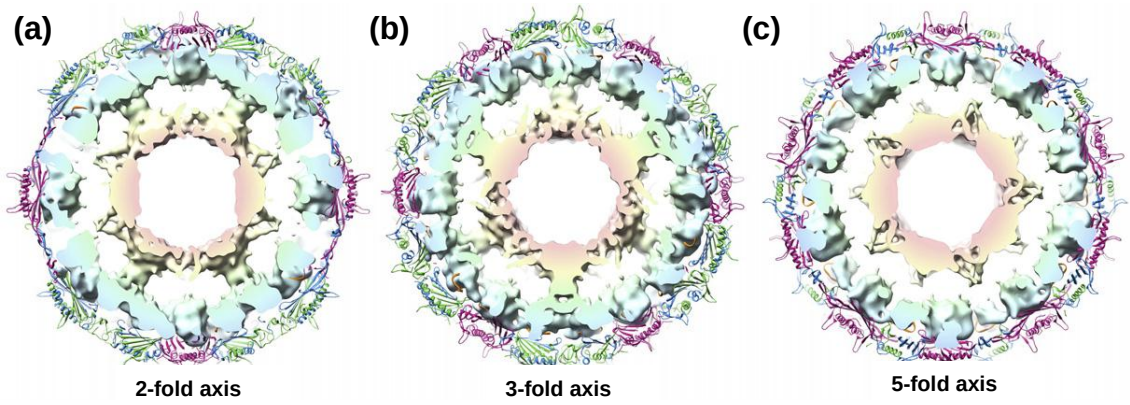


Figure 1.9: The organisation of the MS2 genome visualised by cryo-EM
A cartoon representation of the MS2 coat protein is shown mapped around the density of the RNA which is coloured radially from pale blue to pale red. Representations are shown at the 2-fold (a), 3-fold (b) and 5-fold (c) for comparison. This figure is taken from Torapova *et al.* [51].

From this model, the genome is organised into two concentric shells, with additional density connecting the two shells in columns of density located at the five-fold contacts, where the size of these columns of density are consistent with double-stranded RNA. The outermost shell of density resides immediately beneath the capsid, forming a network which maps out the RNA binding sites of the inner surface of the capsid. Underneath the A/B dimers, extended density is observed which protrudes towards the five-fold axes of symmetry and forms a ring of continuous density around the five-fold axes. Density contributing to RNA is also observed beneath each of the C/C dimers, however this density is dissimilar to the density observed in respect to underneath the A/B dimers. These observations suggest that RNA interactions between these sets of dimers are different, which is consistent with a model of RNA-induced conformational switching in MS2 assembly.

This so-called dimer switching model assumes that RNA is organised in the capsid to maximise the contacts between the RNA and the 60 A/B CP_2 dimers. This requires that the RNA contacts the A/B positions once at every position,

forming the vertices of the outer polyhedron of density[52]. The path the RNA can take in making these contacts can be modelled mathematically using a Hamiltonian path. In total, there are 40,678 possible combinations this path can take which are consistent with the density from the cryo-EM map[52]. MS2 virions also contain 1 copy of maturation (A) protein, which is essential for infection[53]. The A protein is known to bind to specific sequences close to the 5' and 3' ends of the RNA molecule, essentially circularising the RNA[54]. Using this information, only a total of 66 Hamiltonian paths in the dimer switching model can be realised. Combining this information with data from mass spectrometry[55] and kinetic modelling[56] shows that only 3 of these 66 configurations are energetically favourable, with one of these 3 being most favourable energetically, and therefore most likely[52].

1.3.3 RNA folding *in capsa* is independent from sequence and length

Encapsidated RNAs are involved in many secondary structure interactions. Whether these interactions are governed by the primary sequence of nucleotides or facilitated by other factors is of debate. *In vitro* folding studies of RNA have shown that the polynucleotide chain tends to collapse into a collection of many non-native states, often with a significant energy barrier between them[57]. This indicates that RNA is guided into the correct conformation by exogenous factors as folding prior to virus assembly would result in many non-native RNA conformations, which is not what is observed in virus particles. It is most likely that some sort of assistance is required to chaperone the folding process, most probably the coat protein. Further evidence of this comes from the fact that recombinant PaV viral capsids can encapsidate random sequences of RNA instead of their wild-type genome[46; 58]. A heterologous expression

system was used to generate random sequences of differing length and subjected to cryo-EM and image reconstruction[46; 58]. The resulting virus-like particles with randomly encapsidated RNAs contain a virtually identical double stranded dodecahedral cage structure to that of the wild-type. These results suggest it is the coat protein that governs encapsidated RNA structure and spontaneous folding does not occur prior to assembly. The sequence and the length of the RNA chain do not appear to play a role in RNA structure. Instead, the RNA appears to adopt this conformation by being condensed with assistance to encapsidation by the PaV coat protein dimers[46; 58].

Although virus particles can package random sequences of RNA, it is likely that viral RNAs have evolved to be preferentially packaged over non-viral RNAs. Evidence for this hypothesis is supported from two studies using another Nodavirus, flock house virus (FHV). Firstly, FHV particles containing random RNAs have different biophysical properties when compared to wild-type FHV particles. An example of this is that 'particle breathing' can be observed, which is detected by a greater susceptibility to protease digestion[59]. This is probably the result of inefficient protein-RNA contacts which give rise to this decrease in structural stability. Secondly, FHV viral coat protein preferentially packages genomic RNA over other cellular RNA in competition experiments[1]. Preference is probably governed by three main factors; firstly, there may be a particular sequence of RNA that viral coat proteins bind with high affinity. Secondly, the sequence of RNA may be optimised with viral RNAs so that the correct conformation of RNA can be adopted within the viral capsid. Finally, viral RNA has a greater propensity to 'collapse' compared to other RNAs when presented to their complement coat protein[60]. These factors, as well as others, probably

contribute to whether a virus particle can assemble or not. The next subsection discusses virus assembly in general and common characteristics in virus assembly, giving specific examples where necessary.

1.3.4 *In vitro* virus assembly

The earliest capsid assembly experiments were performed with Tobacco Mosaic Virus (TMV), which is the classical rod-shaped RNA virus that appears in virology text books. It has a diameter of about 18 nm and is around 300 nm in length, composed from ~2130 identical disc-shaped subunits stacked in a helix around its 6395 (+) ssRNA genome.

The *in vitro* self assembly of TMV has been well characterised and consists of two separate events: nucleation and elongation. TMV nucleation begins by the specific recognition of the central pore of the disk monomer to a stem loop structure on the TMV (+) ssRNA genome. This in turn melts the stem loop structure, thus more of the ssRNA genome is available for binding. This event mediates a change in the conformation of the disk protein to form a protohelix, which allows further binding of disk proteins to both the top and the bottom, thus completing the nucleation reaction. The resulting nucleated structure then allows elongation reactions to occur as the RNA is threaded through the central pore until stacking of the 5' end is complete.

The hairpin structure is located ~1 kb towards the 3' of its ssRNA genome and its specificity is achieved through a repeated G-x-x motif, which is essential for coat protein binding[61]. After this initial stage, disk proteins interact in an unspecific manner, although an apparently preferred guanine residue is placed

throughout the genome. The positioning of guanine in this way gives an added stability by adding two extra hydrogen bonds per trinucleotide-disk protein interaction. Surprisingly, it has been demonstrated that TMV coat protein can package any ssRNA as long as the initial RNA stem loop initiation signal is present in the sequence. This confirms that further RNA-protein interactions are not base specific and not sequence driven.

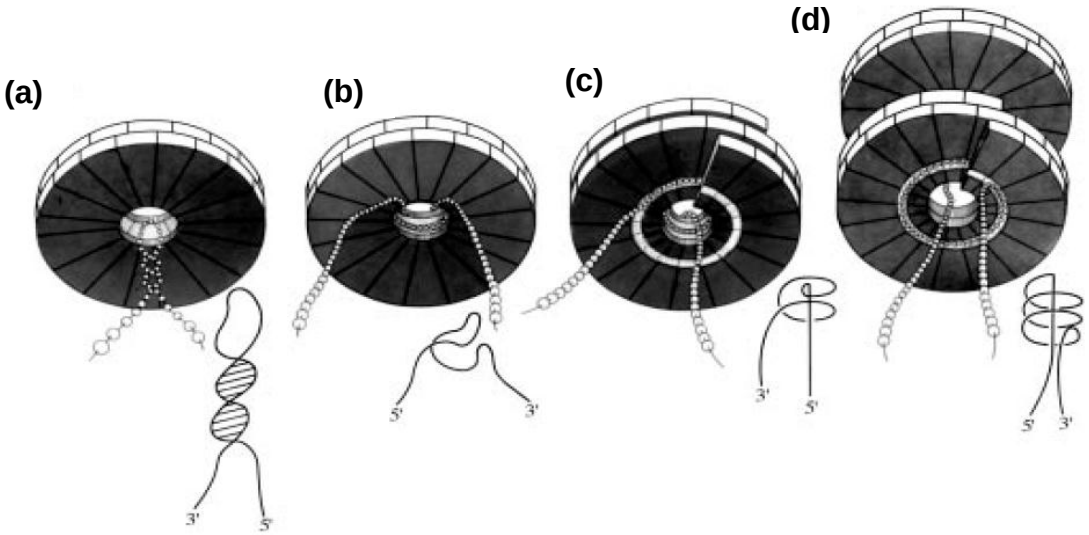


Figure 1.10: Diagrammatic representation of TMV assembly
(a) TMV coat protein disk recognises stem loop
(b) Secondary structure of ssRNA is unwound
(c) Conformational change in coat protein monomer, completes nucleation event to form protohelix structure
(d) Elongation of the protohelix with more subunits using the genome as a template until the genome is completely encompassed. Figure adapted from Butler *et al.* [61]

The nucleation event in the example of TMV is not unique to TMV capsid assembly, and neither is the ability of viruses to package other heterologous RNAs. Several other sequence specific protein: RNA interactions have since been identified in the assembly of ssRNA viruses; examples include MS2[35; 62; 63; 64; 65], poliovirus[66], alphaviruses[67], turnip crinkle virus (TCV)[33], human immunodeficiency virus (HIV-1)[68] and flock house virus (FHV)[69]. These specific sequences of RNA are known as packaging signals, and are usually stem loop

structures. In this way, it seems the tools necessary for preferential packaging of the genome from a milieu of RNA within the infected cell are provided by both the RNA and the coat protein.

In the example of MS2, the packaging signal has been identified as a short translational operator of 19 nt arranged as a stem-loop (TR) which encompasses the start codon of the replicase gene. This operator has been implicated *in vitro* as having sub-nanomolar affinity for CP_2, where *in vivo* it initiates the repression of the replicase cistron by displacing replication complexes from the RNA[70]. The molecular detail of this recognition event has been extremely well characterised both biochemically and structurally. The TR-CP interaction had been accepted as being the sole basis for initiation of RNA-CP *in vivo*, but recently it has been observed that there may be multiple RNA packaging sequences of lower affinity in the MS2 genome (Dykeman *et al*, submitted). This has implications for how the RNA genome controls capsid assembly in a cooperative manner and supports the dimer switching model described previously.

Some ssRNA viruses can assemble *in vitro* into structures resembling that of their native capsids without the need for RNA. These observations do not invalidate the importance of RNA in capsid assembly, but do suggest that the information necessary to construct a capsid can be found within the protein of these viruses. However, because RNA is required for virus integrity, having RNA involved in nucleation is reasonable. It is therefore not surprising that early *in vitro* assembly experiments suggested that assembly was catalysed in the presence of RNA at low concentrations in conditions where capsids would not

usually form with protein alone[71; 72; 73]. In addition, many viruses actually require RNA for assembly, implicating RNA in important role(s) during the complete life cycle of the virus. The next subsections give some background to the model system of STNV, the structures of recombinant STNV and the wild-type virus. In addition, the recombinant expression system and the sequence specificity of RNA to STNV CP will also be discussed.

1.4 Satellite tobacco necrosis virus (STNV)

Satellite Tobacco Necrosis Virus (STNV) is one of the smallest known viruses; it has a diameter of 17 nm and is the satellite virus of Tobacco Necrosis Virus (TNV) on which it relies for its replication[74; 75]. *In vivo*, STNV infects the root cells of tobacco plants, transmitted by the soil fungus *Olpidium brassica*. Like other plant satellite viruses, the coat protein of STNV shows no immuno-cross-reactivity with its helper virus and displays very little sequence similarity as observed both at the amino acid and nucleotide level. The only sequence similarity lies in 100-150 nucleotides towards the end of the 3' RNA genome. This is the region where the replicase of TNV is thought to bind initially to the STNV genome[76].

The genome is 1239 nucleotides in length, where the initiation AUG codon of the coat protein starts at position 30-32, close to the 5' terminus of the RNA genome. This translates a 195 amino acid, 21.54 kDa coat protein monomer, 60 equivalent copies of which, arrange into a *T*=1 icosahedral capsid.

1.4.1 STNV structure

STNV readily crystallizes; indeed, it was even observed to crystallize *in vivo*[76]. This property of STNV has enabled the three-dimensional structure to be determined to a high resolution by X-ray crystallography[77; 78].

The coat protein monomer has a distinctive jelly-roll β-barrel fold, which despite being the same topology as many other plant satellite viruses, is otherwise striking in its dissimilarity to other plant satellite viruses[79]. The crystal structure of STNV-1 shows that residues 1-11 were disordered; otherwise the electron density can be followed from residues 12 completely to the C-terminal residue[77; 78]**.** Residues 12-24 fold into an α-helix with the remainder of the coat protein forming a β-barrel composed of two four-stranded anti-parallel β-sheets. The first β-sheet, termed BIDG due to the relative positions in the polypeptide chain, has a twist of about 5° per strand. At the 5 fold axis of the icosahedron, the strands of the two β-sheets are connected by four short loop regions, whereas the three-fold axis is connected by two larger loop regions. Another loop region is found connecting the β-strand of βE and βF of the CHEF β-sheet (see Figure 1.11).

1.4.2 STNV stability, calcium ion binding and capsid swelling

STNV is a very stable virus particle. It is resistant to both high and low ionic strength solutions, as well as being stable between pH 2 to pH 11, and is resistant to both nuclease and protease digestion[80]. Despite being stable in a wide variety of harsh conditions, wild-type STNV is still capable of releasing its RNA genome when it enters the plant cell cytoplasm.

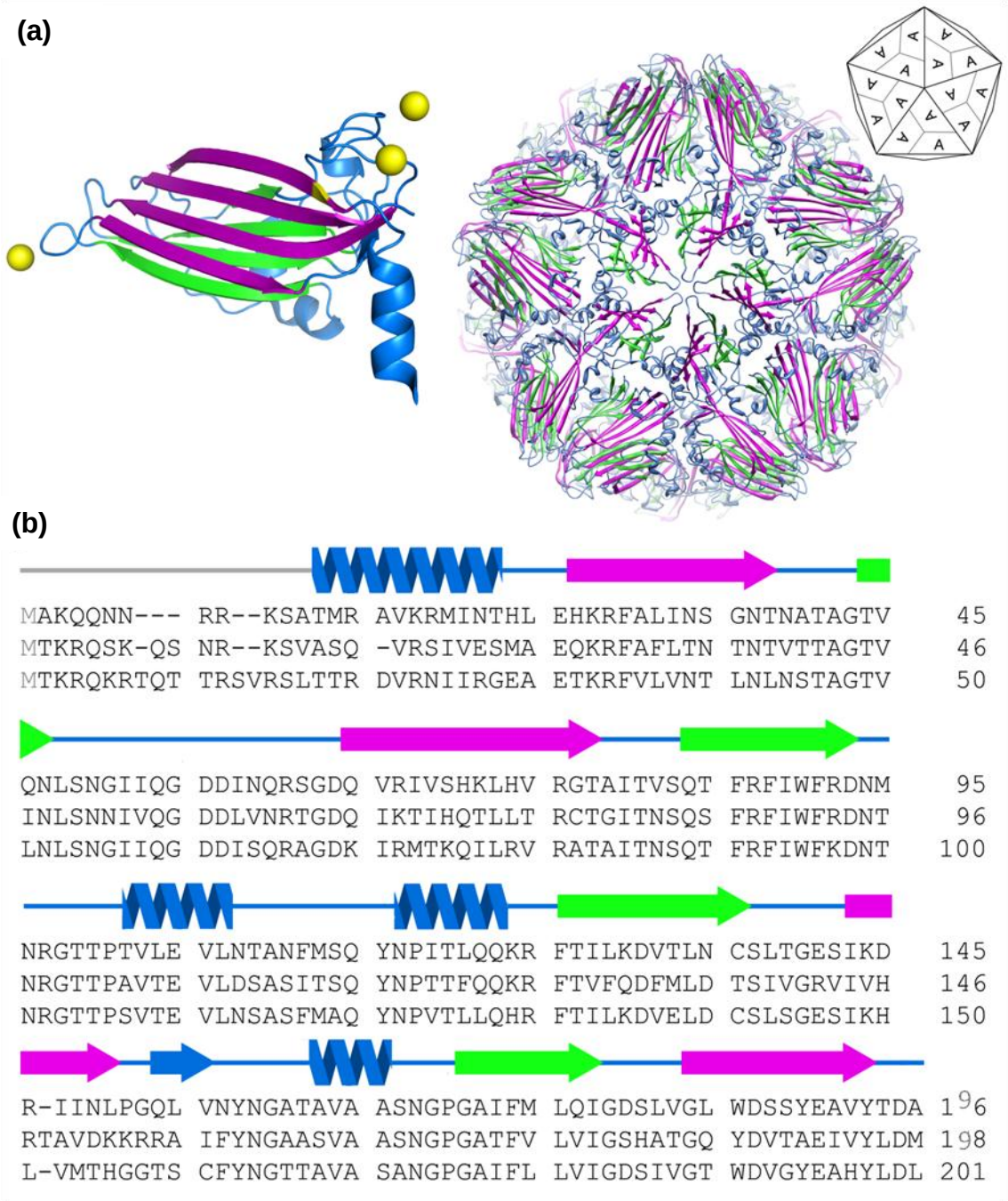


Figure 1.11: Architecture of STNV. (a) Shows the secondary structural elements of a single subunit of STNV (Protein Data Bank ID 2BUK), with the anti-parallel β-sheets shown in magenta (BIDG) and green (CHEF). The helical regions are shown in blue. The three classes of associated calcium ions are shown as yellow spheres. The amino acid highlighted in yellow is Arg 28, a known trypsin cleavage site. The X-ray structure of the STNV *T*=1 capsid is shown alongside with an icosahedron as a guide to its symmetry. (b) Shows the amino acid sequences as single-letter codes of the STNV-1 (upper), STNV-2 (middle) and STNV-C (lower) CPs. Shown above the sequences are the positions of the secondary structure elements in STNV-1 from (a). Figure taken from Bunka *et al* [81].

The proposed mechanism of virus disassembly has not been well characterised, but is thought to involve the removal of calcium ions - which function to stabilise the capsid - upon entry of the virus capsid into the low calcium environment of the plant cell cytoplasm[82]. Further investigation regarding STNV VLP disassembly is discussed in Chapter 5.

Between each subunit of the STNV coat protein, calcium ions are found at three different sites (shown as yellow spheres in Figure 1.11). The calcium ion at site one is at a general position close to the three-fold axis, and is coordinated by the carbonyl oxygen atoms from residues 61 and 64, a carboxylate group from Asp194 from the same subunit and another carboxylate group from Glu25, a residue found on the adjacent subunit. An octahedral arrangement is completed by the presence of two water molecules[83; 84]. Another calcium ion binding site is present at the three-fold axis termed site two, where the calcium ion is coordinated by three carboxylate oxygen atoms of Asp55 from the three surrounding subunits. The final calcium ion binding site is located at the five-fold axis, where the calcium ion is coordinated by the five carbonyl atoms of Thr138 and two water molecules.

In view of the presence of these stabilising calcium ions in the STNV capsid, research was carried out by the lab of Liljas, Montelius and Unge to simulate the conditions of the plant cell cytoplasm. Treatment of STNV with divalent cation chelating agents, such as EDTA, caused a pH-dependent swelling of the virion as well as subsequent sensitivity to nucleases and proteases[80]. A low-resolution crystal structure was determined for this slightly expanded STNV structure, which showed that the calcium ions present at sites 1 and 2 in the

native crystal structure were absent from the EDTA-treated crystal structure, whereas the calcium ions present on the five-fold axes of the native crystal structure were still present in this expanded form[83]. The calcium ions in the five-fold axes can be removed by increasing the pH to pH 8[84].

1.4.3 RNA structure in STNV

At first look, satellite tobacco necrosis virus (STNV) is similar in size, shape and architecture to STMV. This close relationship however begins to buckle when compared biochemically, structurally[79] and when the RNA structure is analysed *in capsa*[85].

Extensive structural analysis has been carried out on STNV using various biophysical techniques, such as neutron diffraction[85], X-ray small-angle diffraction[86] and X-ray crystallography[78], where the three-dimensional structure has been resolved to 2.5 Å[77]. Continued crystallographic attempts to get the encapsidated RNA visualised at high resolution resulted in no detailed structural information of the RNA conformation[83; 84]. Instead, it was the low resolution studies of neutron diffraction which led to a better view of the encapsidated RNA conformation[85].

1.4.4 Recombinant expression and structural characterisation of STNV VLPs

Recombinant STNV coat protein was first expressed in *E. coli* in 1984, where the coat protein was rendered almost completely insoluble, extensive proteolysis was detected and no appreciable self aggregation was observed[87]. However, previous success was achieved using a truncated version of the

Sesbania Mosaic Virus CP expressed in *E. coli*, where it spontaneously assembled into *T*=1 VLPs and the crystal structure was characterised[88]. Encouraged by these observations, Lane *et al* designed a synthetic, codon-optimised variant of the STNV-1 CP gene to maximise its expression in *E. coli*[89]. Following expression, the STNV CP spontaneously assembled into particles resembling that of their wild-type *T*=1 capsids as analysed by analytical ultracentrifugation (AUC) and transmission electron microscopy (TEM). The material was purified and shown to contain a similar RNA content to wild-type virus as shown by UV absorbance[89].

Large crystals of the purified VLPs grew readily under conditions described previously[77; 78]. From these crystals, X-ray diffraction data were collected at high resolution and low resolution to obtain as much information as possible about the structure of the protein capsid and the RNA structure inside the VLP.

The high resolution structure was determined by molecular replacement with the coordinates of the wild-type (PDB ID: 2STV) used as a model. The electron density from the VLP was present in almost the entire polypeptide chain, other than residues 1-11, as is the case in the wild-type virus[77; 78]. The overall density showed a high correlation between that of the recombinant VLP and the wild-type. However, the N-terminal helix was less well defined, with density becoming progressively weaker towards the N-terminus (see Figure 1.12).

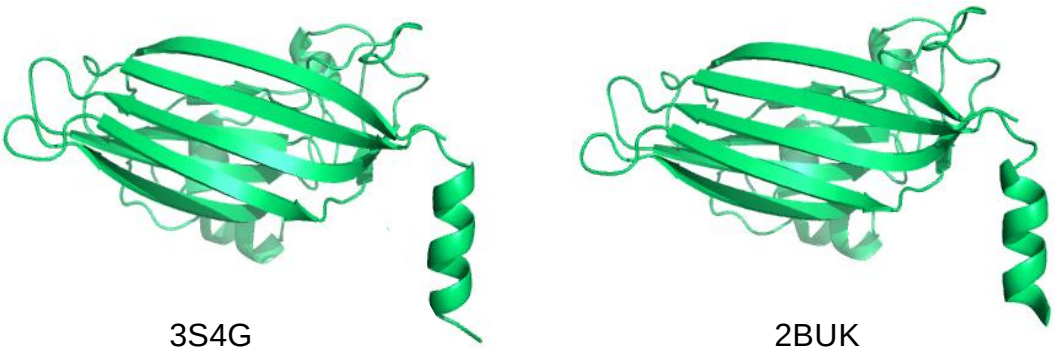


Figure 1.12: Comparison between recombinant STNV CP and wild-type STNV CP
The density from the N-terminal helix is less well defined in the recombinant STNV CP (left) compared to the wild-type (right). Residues 1-11 are disordered, and not observed in either crystal structure.

The high resolution crystal structure showed no interpretable density corresponding to RNA, suggesting that there is no repeating structure obeying icosahedral symmetry from the RNA. A crystal structure determined at 16 Å with H_2O/ D_2O contrast matching did however reveal RNA density running along the 5-fold symmetry axis between the clusters of N-terminal helices[85]. The resolution of this density was however too low for any interpretable RNA structure to be resolved. Although the recombinant VLPs lack the wild-type genome, the $A_{260:280}$ ratio suggests that recombinant VLPs package a similar amount of RNA.

Encouraged by the success of the neutron diffraction study on wild-type virus, Lane *et al.* used low resolution X-ray diffraction data collected between 140-6 Å in an attempt to visualise the RNA component. The high resolution model was used as the basis for rigid body refinement and 60-fold icosahedral averaging. The density from the high resolution model fits the averaged electron density of the capsid from the low resolution model well, but leaves a large area of electron density unaccounted for (see Figure 1.13). From this density, a short 3

bp stretch of A-form double-stranded RNA was modelled with arbitrary sequence, and includes an unpaired base at the 3' end of the stem.

Each RNA fragment fits near the inner surface of two adjacent CP subunits and close to a neighbouring RNA fragment. These RNA fragments are related to the two-fold and five-fold axes, and are arranged as rings under the pentamer vertices of the capsid (see Figure 1.14). The protein surfaces facing the RNA molecules are highly positively charged, where three basic side chains, Arg 14, Lys 17 and Arg 18 lie along one side of the helix, within easy reach of the RNA.

The 3' unpaired base at the top of the RNA duplex lies in a cavity below the surface of the protein, which could plausibly accommodate a loop connecting the RNA duplex. This region does not abide by icosahedral symmetry well enough to be enhanced by icosahedral averaging in the density map. Although other orientations of RNA are possible, the stem loop motif is by far the most plausible. Larson *et al* also observed an unpaired base at the 3' end of a duplex from electron density observed with STMV. This was also interpreted as the start of a loop region joining the stem. In addition, three further basic residues, Arg 66, Arg 91 and Lys 123, lie on the inner surface of the neighbouring protein subunit, and lie in close enough proximity to make RNA contacts with any potential loop (see Figure 1.14a).

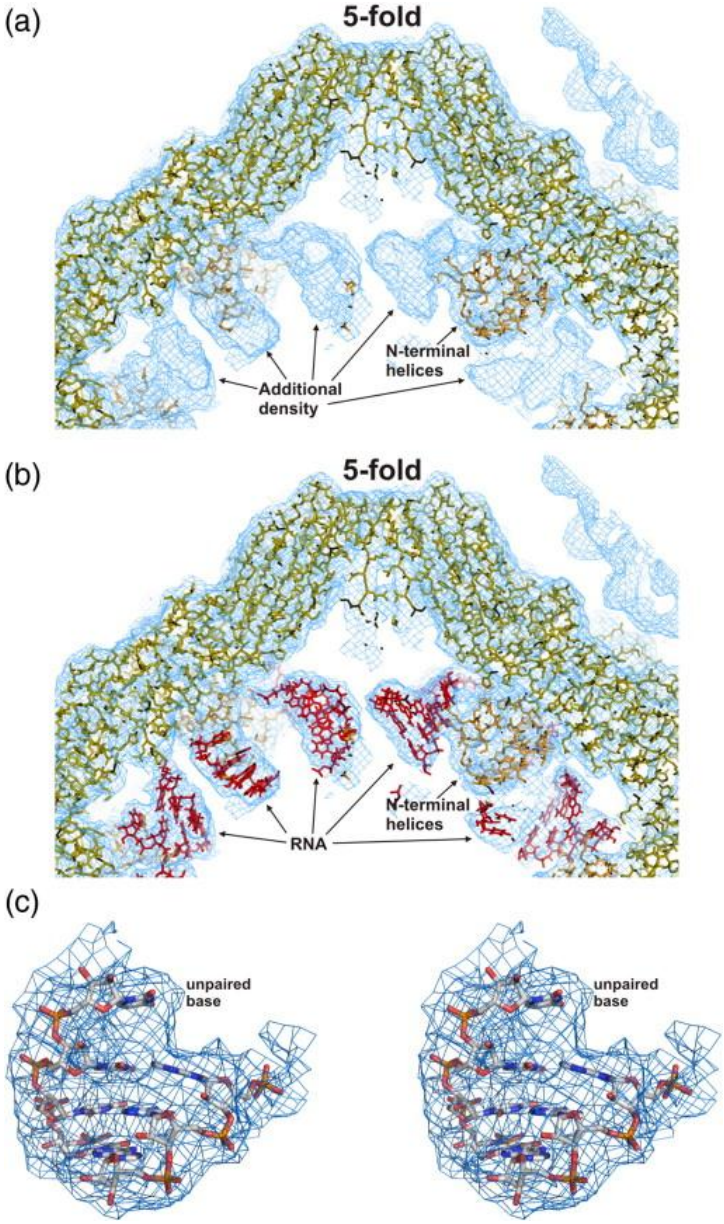


Figure 1.13: Additional density after modelling of low-resolution averaged electron density map.
(a) The CP is shown in stick representation and fits the electron density for the capsid and helical protrusions. Strong additional density is clearly visible between helix clusters.
(b) RNA fragments modelled into electron density. No RNA density observed at the 5-fold vertices or in the centre of the particle.
(c) Stereo view of modelled electron density.
Figure taken from Lane *et al*. [89]

There is also a cavity underneath the five-fold axes. However, there was no density that was attributable to RNA observed in this cavity. If the duplex RNAs are extended, then they would not be able to avoid interfering with an adjacent fragment around the five-fold axes. One plausible explanation is that the duplex RNAs that lie at the two-fold axes are each connected by single stranded regions that do not adhere to icosahedral symmetry at the five-folds.

Although the recombinant VLP does not encapsidate the STNV genome, the mRNA is similar in sequence and secondary structure to the wild-type genome[81]. In addition, it has been observed in other viruses, such as Flock House Virus[58] and Pariacoto Virus[46], that the capsid can impose an architecture on heterologous RNA similar to that of the wild type. In order to better understand the relationship between the STNV VLPs and the packaging of its mRNA ‘genome’, further studies to assess the nature of sequence specificity of this phenomenon were undertaken. The next subsection describes these studies in detail.

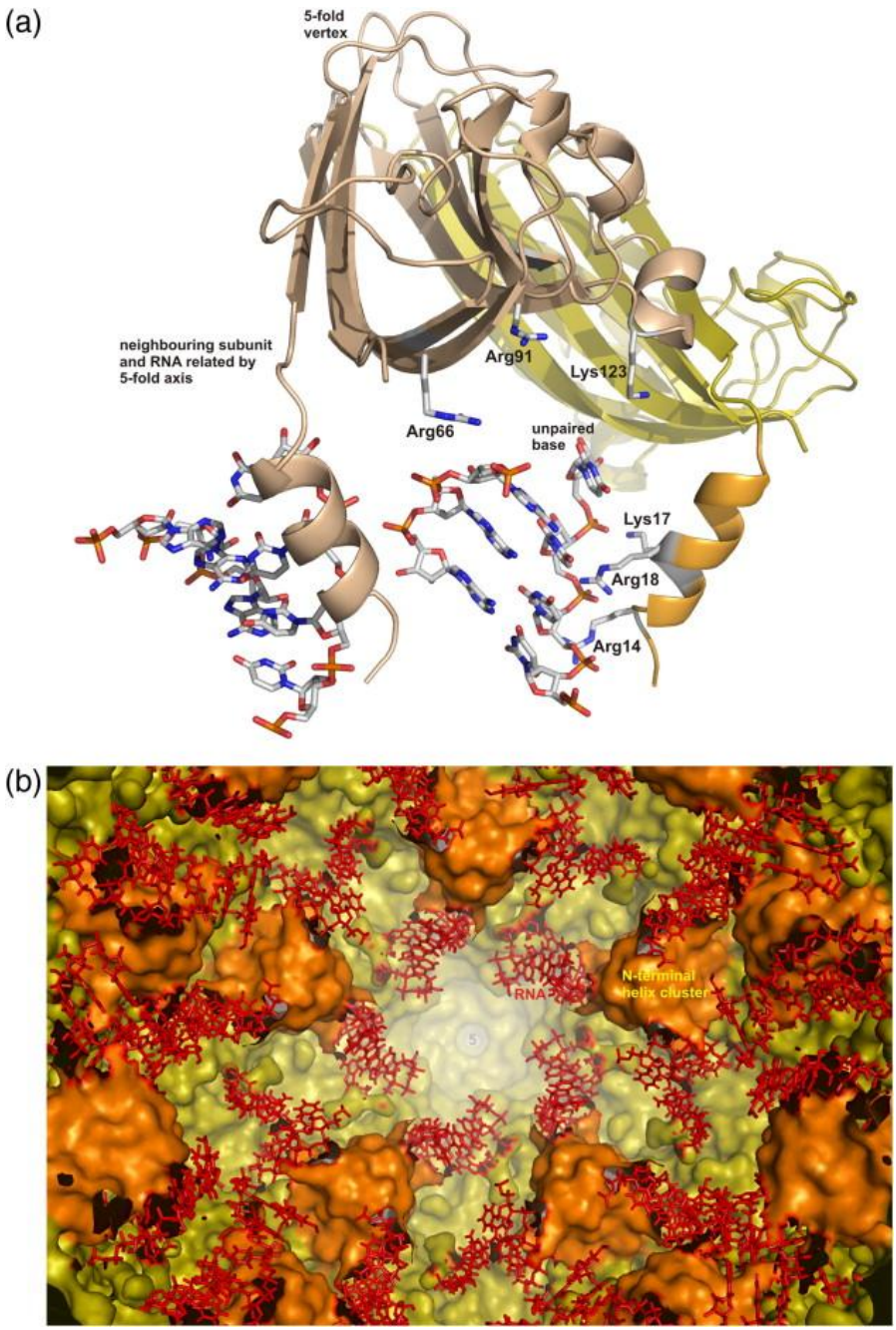


Figure 1.14: RNA: protein interactions in recombinant STNV and a stereo view inside the capsid

(a) CP subunit and its specific interactions with modelled RNA duplex. Three basic residues (Lys 17, Arg 18 and Arg 14) make good hydrogen bond contacts with three phosphate groups on the RNA duplex. Three basic residues (Arg 66, Arg 91 and Lys 123) on the inside face of a neighbouring CP subunit also make good hydrogen bond contacts with the adjacent RNA duplex. The neighbouring subunit is related to the other in relation to its five-fold axis.

(b) View of the inner surface of the capsid looking from the centre of the particle down its five-fold axis. RNA fragments form a ring around the five-fold axis, each associated with an N-terminal helix. Figure taken from Lane *et al.* [89].

1.4.5 Degenerate RNA packaging signals in the genome of STNV

It has previously been established that the expression of STNV CP in *E. coli* results in the production of *T*=1 STNV VLPs *in vivo*. It was known from phenol: chloroform RNA extraction and RT-PCR that a significant proportion of this population encapsidates the STNV CP mRNA. This indicates a degree of sequence specificity towards the 'mRNA genome' and the STNV CP. The STNV VLP crystal structure described in Chapter 1.4.4 shows that there must be multiple protein: RNA contacts to account for the density observed[89].

In order to determine whether mRNA packaging is sequence-driven, SELEX experiments were performed against the RNA binding sites of the STNV CP[81]. In SELEX experiments, it is important to carefully define the selection target. For viruses such as MS2 and Turnip Crinkle Virus, the building block for capsid formation and therefore the selection target, should be a CP dimer. In the case of STNV, it was not exactly known what this selection target should be, as the basic capsomere for STNV assembly was not well defined.

To overcome this problem, Bunka *et al* devised a generic strategy which, in principle, should work for any given virus RNA-binding selection target (see Figure 1.15). The surface-exposed lysine residues on STNV VLPs (e.g. Lys 27) were biotinylated using EZ-link® sulfo-NHS-LC-LC Biotin (Pierce Biotechnologies). The biotinylated capsids were then bound to MyOne™ streptavidin-coated Dynabeads® beads (Invitrogen) and disassembled by exposing to disassembly buffer. The beads were then washed before unmodified, disassembled STNV VLPs were added to the immobilised CP in reassembly buffer in a sub-stoichiometric ratio. The RNA from the disassembled

VLPs was therefore included in the reaction mixture. The rationale behind this methodology was that this would, in principle, enable the formation of several partially formed capsids on the beads. These partially formed capsids would have some RNA bound to them, but should also display some 'empty' RNA binding sites that are hidden in the VLPs. In addition, it is likely that during selection, the formation of tighter binding RNA ligands would displace the bound RNA, especially given that only the selected sequences are recovered and amplified during each round. These partially assembled targets were therefore used for 10 rounds of selections. To remove aptamers directed against epitopes on the surface of the VLPs, unmodified VLPs were also added in 10-fold molar excess in the selection rounds.

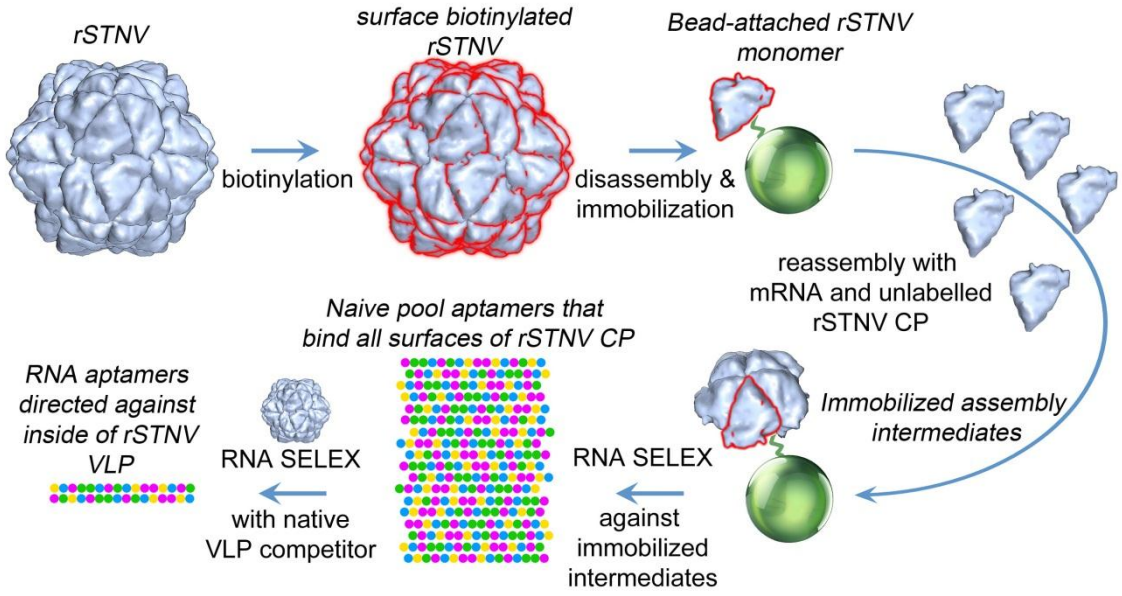


Figure 1.15: Schematic representation of aptamer selection process.
STNV VLPs were biotinylated by amine modification before being immobilised on streptavidin beads and disassembled by exposing to disassembly conditions. The particles were then extensively washed. Unmodified disassembled VLPs were then added to the immobilised CP in moderation and subjected to reassembly conditions. Because the concentration of unmodified CP was controlled, the immobilised CP subunits were not able to fully assemble into VLPs, therefore displaying a range of partially assembled intermediates. Ten rounds of SELEX were then performed in the presence of unmodified VLP competitor to remove any aptamers directed against epitopes on the outer surface of the VLPs. Figure provided courtesy of Dr Neil Ranson.

After 10 rounds of selection, the resulting clones were sequenced, with 17 aptamers identified (see Table 1.2).

Table 1.2: Sequences of the 10th round STNV-1 CP aptamers

Sequence name	Sequence	Number of residues
A1	CGTCATGAATAGAGTTTTTTGCGGTCAGCA	30
A2	GTGTTTATCTTGTCACTAGGTACGGGTATG	30
A5	CGCGACCATTGAAATCACTTCTGGGGTTAA	30
B1	GTGTAATTTGGTTTATTCTTTCAATGGTGA	30
B2	CATTACAACTATCACTTCTCCTC	23
B3	CCTTTTCAAGACATGCAACAATGCACACAG	30
B4	CGTTTTAGTTTTCGGGTATTAATAGGCGTG	30
B5 (& F2)	TATGAGTTGATTTGTGGTTTTTGGTAACGT	30
C1	CTTGATTAACTAAAGGCTAGCAAACAGCG	30
C2	AACATTTGGATGTATGCCTTAGTTACCTAG	30
C3	CACAGGTAAGAATCCAGCCCTATACTAACA	30
C4	TGGACACCCTAAAAATATACGGCTCCCTGG	30
C5	ACATGCCTTTATTTTGTCTTGGATCTTGAC	30
D1	GTCTCCATGAACAATATTGTAGCTACAGGTC	30
D5	CTGTGTAGTTTGAAATCATTATCGCCAACA	30
E5	GAGCAATCGGGATACCTTCACATGATGTGC	30

Out of these 17 aptamers, 16 were unique sequences, with aptamers B5 and F2 being identical sequences. When each of the aptamer sequences were compared using the Alibee multiple alignment software (http://www.genebee.msu.su/services/malign_reduced.html), a number of sequence motifs were identified, but there were no clear consensus sequence. When these sequences were compared to the genomes of STNV-1, several statistically-relevant matches on different regions of the STNV-1 genome were

identified. One of these aptamers, termed B3, contained 16 nt that matched the STNV-1 genome from positions 57-81 (see Figure 1.16a). In other words, that is a match from 16/25 positions, including a 10/10 contiguous stretch. This is unlikely to have occurred by chance, suggesting that this sequence is a preferred STNV CP binding site[81]. MFold analysis of this motif, as well as the STNV-1 genomic region and the STNV CP mRNA where this aptamer matches, suggest that this sequence folds into a stem loop structure[81], a known coat protein binding motif[62] (see Figure 1.16b). The nucleotides immediately adjacent to the matching sequences form base pairs, increasing the significance of the match. Density corresponding to RNA in the crystal structure from the recombinant VLP is consistent with a stem loop occupying the CP binding site[89].

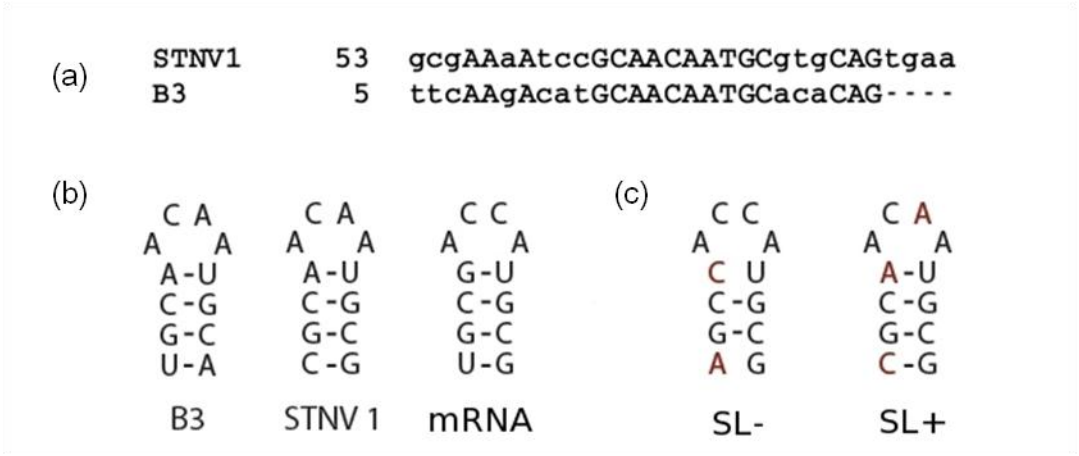


Figure 1.16: Identification of a common structural motif
(a) Sequence alignment of B3 and STNV-1 genome. The matches are shown as capital letters whereas the unmatched nucleotides are shown in lower case.
(b) Analysis of the stem loop structures from aptamer B3, the STNV-1 genome and the mRNA.
(c) The stem loop mutations from the mRNA stem loop to weaken the base pairing (SL-) and to strengthen the base pairing (SL+). Figure taken from Bunka *et al*. [81]

To assess whether aptamer B3 was functional *in vitro*, RNase protection assays were performed on the aptamer in the presence and absence of STNV CP. The

results obtained were consistent with the predicted secondary structure of aptamer B3, with the stem loop being protected from nuclease digestion in the presence of STNV CP, but digested in the absence of CP. Capsid assembly studies with B3 are discussed in Chapters 3 and 4.

In order to explore the significance of the B3: STNV CP interaction *in vivo*, a series of stem loop mutations were made to disrupt (SL-) or to mimic (SL+) the stem loop from the wild-type genome, but without changing the amino acid sequence of the coat protein (see Figure 1.16c). These mutants were used to assess how these mutants affect the production of VLPs and the $A_{260/280}$ ratios relative to the VLPs described in Lane *et al*[89]. The results from these studies showed that both the yield and $A_{260/280}$ ratios from each case were identical to the original particles. These results suggest that this B3-like stem loop is not a unique initiation site in the STNV genome, and that many equivalents of this stem loop may exist. This may explain the multiple sequences selected in the aptamer selection experiment[81].

To test this hypothesis, the other aptamers were scrutinised both for their ability to form stem loop structures, and whether they contain sequence similarity to the STNV genome. The next best alignment was with aptamer D1 with two matches on the STNV genome, one with 17/23 identical matches and another with 18/23, both with a contiguous stretch of 5/5 residues. MFold analysis from this aptamer predicted that it formed a stem loop structure also displaying the motif AXXA, albeit in the context of a 5 nt loop (AACAA). The other aptamers were also screened for their ability to produce a stem loop in either a 4 nt loop or a 5 nt loop. Four aptamers (A2, B3, B4, and C3) were predicted to form 4 nt

stem loops, and an additional five aptamers (A5, C1, D1, D5 and E5) were capable of forming 5 nt stem loops. It was therefore postulated that this AXXA motif was important for STNV CP binding. Further analysis from the rest of the aptamers was carried out, which included the fixed flanking regions of the aptamers. In addition, aptamers were screened against producing a stem loop with six nucleotides in the loop. These analyses identified an additional 5 aptamers capable of forming these structures, with only aptamers B1 and B2 incapable of doing this. One of these aptamers (B1) was much shorter than the rest[81].

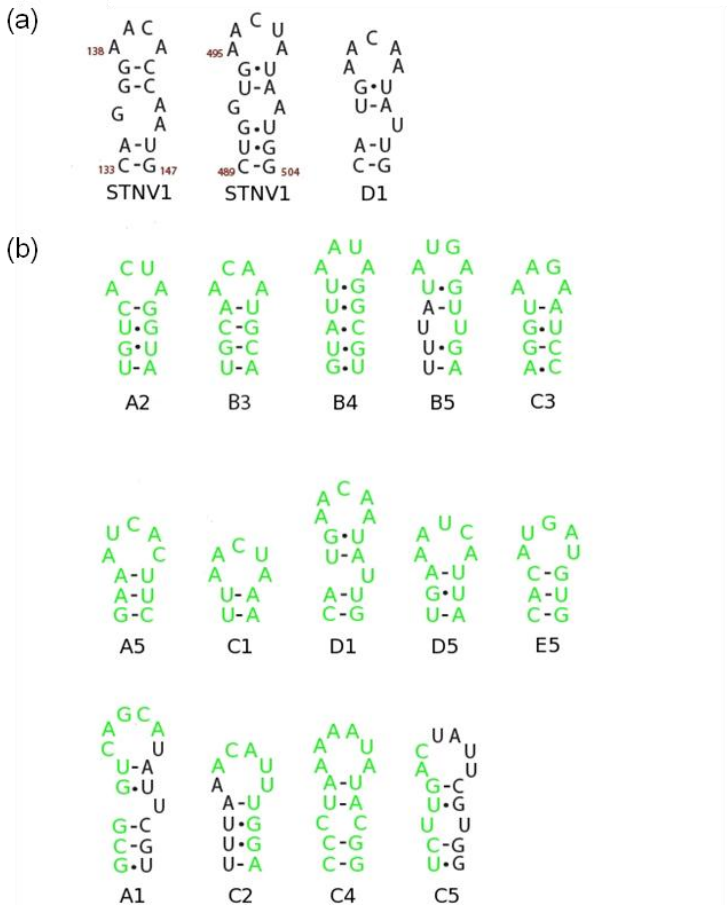


Figure 1.17: Secondary structure predictions of selected aptamers
(a) The secondary structure of the next best aptamer match (D1) to two portions of the STNV-1 genome, all of which can be folded into stem loops displaying the AXXA motif.
(b) MFold secondary structure Figures displaying the other aptamers capable of forming either a 4, 5 or 6 nt stem loop displaying AXXA. The selected regions are shown in green, whereas the fixed regions are shown in black. Figure taken from Bunka *et al.* [81]

In order to probe the STNV genome for stem loop sequences containing 4 to 6 nucleotides in the loop and allowing for non-Watson-Crick base pairing in the stem, a series of six degenerate sequences were set up. This analysis identified 30 potential stem loops[81], enough to satisfy the density corresponding to the RNA binding sites found in the crystal structure of the STNV VLP[89].

1.5 Uncoating and disassembly

Despite the multitude of different molecular mechanisms that viruses utilise for infection, all viruses share the same goal: to transport their viral genome from an infected cell to a non-infected cell, delivering it to either the cytoplasm or the nucleus, in a replication-competent manner. In order to achieve this goal, the virus has to overcome several barriers depending on the host it would like to infect. The target of virus infection can either be a neighbouring uninfected cell, or a cell in another organism. Uncoating is an integral part of this process, where the virus is generally destabilised or modified, disassembled and the genome is presented to the replication machinery. The progressive success of a virus through these stages of infection is completely dependent on the processes of the host cell. This 'Trojan horse' strategy is necessary because viruses are simple; they are generally incapable of having multiple independent functions. Another problem the virus must overcome is that of performing a balancing act: viruses must be stable enough to protect their genomes but dynamic enough to dissociate and deliver the genome to the host's replication machinery.

Strategies by which viruses enter the host cell differ depending on the structure of the virus. Enveloped viruses tend to enter the cell by membrane fusion, either by direct means or after endocytosis[90]. For non-enveloped viruses, this mode of penetration is not possible; therefore they have evolved other ways of entering. Many animal viruses, for example, enter the cell by receptor mediated endocytosis. Endocytosis is a process which mediates the controlled internalisation of large macromolecules that would normally be incapable of passing the membrane by diffusion or facilitated through ion channels[91]. It can come in many different flavours, such as clathrin-mediated endocytosis (CME), macropinocytosis, caveolar/lipid raft-mediated endocytosis, or as yet unknown, poorly characterised methods[92].

The most understood process that viruses utilise is the CME pathway. CME is an endocytic process driven by the accumulation of clathrin and the formation of a clathirin coat on the cytoplasmic leaf of the lipid bilayer. The production of primary endocytotic vesicles requires a large number of cellular factors, some of which are exploited by viruses. A subset of these are known as attachment factors and receptors; including the transferrin receptor, low-density lipoprotein receptors or carbohydrate moieties of proteoglycans, glycolipids and glycoproteins[92].

The most commonly exploited attachment factors are the glycosaminoglycans (GAGS), particularly heparin sulphate[92]. Upon virus binding to these attachment factors, clathrin on the inside of the membrane cause the membrane to curve and pinch off into the cell, resulting in the formation of an early endosome[93]. Once the early endosome has formed, it is organised into a sorting centre called

the endosomal network, where the resulting endosome can either be sent back to the cell surface in a recycling endosome, or be targeted for degradation to the late endosome and the lysosome. The pH of the early and late endosome is lower than that of the cytoplasm. Viruses can therefore take advantage of this change in pH, driving conformational changes in the coat protein which enables the escape of the particle[91].

The first virus to be identified using this pathway was the semliki forest virus[94]. Since then, a multitude of different viruses have been identified, such as adenovirus 2 and 5, hepatitis C virus, dengue virus and influenza A virus. In addition to CME, some viruses may use additional entry mechanisms at the same time, such as those described for influenza A[95]. Adenovirus undergoes a conformational change upon acidification, but the DNA genome is not released into the cytoplasm. Instead, a modified, uncoating intermediate is released and exploits the cellular transport machinery in order to be transported to the nucleus[96].

Plant viruses, on the other hand, need to employ different strategies to overcome the major barrier of the cell wall. Unlike DNA bacteriophages, which operate complex injection mechanisms to insert their genome into the cytoplasm of the bacterial cell, plant pathogens rely on physical damage to the cell wall[97]. This damage may have arisen through seed or vegetative propagation, or in other examples, by using insects or fungi as agents that damage or weaken the cell wall[97]. The relationship between the virus and its vector is often very specific (and complex), and accounts for the specificity of the plant pathogen to that of its host range[98].

The uncoating of some plant viruses have been shown to occur co-translationally. This process is mediated by the ribosome, where translation of the viral genome often results in the formation of striposomes; this has been observed both *in vitro*[99] and *in vivo*[100]. The first striposomes observed were from TMV. In 1984, Wilson and co-workers found that by treating TMV under mildly alkali conditions, some of the subunits of the TMV dissociated from the 5' end of the RNA. Exposure of these modified TMV particles to cell-free expression systems resulted in the production of viral proteins. Further analysis showed that TMV RNA can be attached to as many as 24 ribosomes[101].

1.6 Host antiviral mechanisms

Over the course of evolution, the virus and host have coexisted, imposing an evolutionary pressure on both[102]. On the one hand, the host has developed mechanisms to prevent virus infection and to eliminate virally infected cells. On the other hand, the virus has evolved an array of immune evasion techniques. As viruses infect cells in every kingdom of life, antiviral defence mechanisms can be found in all organisms. For example, bacteria have evolved by expressing restriction enzymes[103]. Restriction enzymes recognise specific sequences and denote them as foreign; these sequences are then destroyed by the restriction enzyme, leaving the host DNA untouched. Bacteriophage have evolved to express proteins which interfere with this function, by mimicking the shape and the charge distribution of DNA and blocking the restriction enzyme's active site[103].

Higher organisms have evolved more elaborate defence mechanisms, such as the innate/ adaptive immune systems and RNA interference. Vertebrates have an extremely intricate immune system to protect themselves against viral infection. Details of this are complex, and way beyond the scope of this thesis. Instead, a brief overview will be described and a few examples given as to how viruses are able to evade the immune system.

The immune system is generally split into two separate categories for convenience and thought of as two separate entities; innate immunity and adaptive immunity. This can often be misleading. In reality, adaptive and innate immunity is inextricably linked, where the mechanisms of each enhances the others ability to recognise and destroy foreign, infectious microorganisms. The innate immune response is also sometimes referred to as the non-specific immune response, because unlike acquired immunity, it protects non-specifically against pathogens and is the first line of defence. The components of the innate immune system include phagocytes, mast cells and natural killer cells amongst others. Adaptive immunity on the other hand, recognises foreign invaders specifically and has an immunological ‘memory’ of when it last encountered the foreign invader. When this memory event is triggered, it leads to a cascade of events which results in the secondary immune response being activated. This secondary immune response is fierce, and becomes more ferocious every time the pathogen is encountered. The key players involved in this are the memory B- and T-cells, which are the main targets during immunization strategies.

Viruses have adapted to the immune system in different ways depending on the coding capacity of the virus. Small viruses, such as the *Picornaviridae* and retroviruses, change their immunodominant antigens so the immune system does not recognise them. Larger viruses can express a large array of proteins with specific effects on immune recognition and effector functions. In general, viruses utilise three main generic strategies to evade the immune system:

(1) Rapid mutation or 'hiding' of their immunodominant epitopes to avoid detection by the humoral immune response. Examples of viruses utilizing these evasion techniques are Influenza A (mutates haemagglutinin/ neuraminidase antigens by antigen drift/ shift[104]) and HIV (as well as rapid mutation, coats its immunodominant antigens with carbohydrate, resulting in a 'masking' effect from neutralising antibodies[105]).

(2) Interfere with the functioning of the cellular immune response. Examples include human cytomegalovirus (HCMV) and herpes simplex virus (HSV) (downregulate the production of natural killer cells and express a major histocompatabilty complex (MHC) class 1- homologue, so the immune system recognises the infected cell as 'self', and binds inhibitory receptors of the immune system for protection).

(3) Interfere with immune effector functions, such as inhibiting inflammation, activation of complement, apoptosis and signalling. Examples include HCMV, HSV and Adenovirus (they each express multiple homologues to subvert these immune functions). For a review, please refer to Vossen *et al* [102].

The other defence mechanism involved is called RNA interference. This mechanism was first recognised in plants, but has subsequently been found in all eukaryotic organisms[106]. The primary roles of RNA interference is for the regulation of gene expression and for protection against invading viruses and transposons. Its utility is dependent on the generation of short fragments (~21-26 bp) of complementary RNA[106]. The production of these complementary fragments is manufactured from precursor dsRNA molecules by an enzyme called Dicer. The dsRNA molecule is then separated into two strands; the guide strand binds either an RNA-induced silencing complex (RISC) or a RNA-induced transcriptional silencing complex (RIST), while the complementary strand gets degraded[107]. RISC is responsible for the degradation of the mRNAs or viral RNAs in question, whereas RIST induces methylation of the chromatin to downregulate transcription[107; 108]. Despite dsRNA molecules being the strongest triggers for RNA silencing mechanisms, ssRNA or ssDNA can also activate silencing, mediated by a dsRNA intermediate generated by the RNA-dependent RNA polymerase[107].

Plant viruses have evolved ways to circumvent this silencing. Evidence suggests that in some plant viruses, the capsid proteins directly interfere with the silencing mechanisms described above. This suggests that the capsid protein contains multiple functions, further illustrating how genetic economy is exploited within this class of viruses[109]. One such plant virus, turnip crinkle virus, is described in the next subsection.

1.7 Turnip Crinkle Virus

Turnip crinkle virus (TCV) is a member of the Carmovirus genus and is a family member of the *Tombusviridae*, the family being named after its founder Tomato Bushy Stunt Virus (TBSV). Experimentally, TCV has a very broad host range, but in nature its host specificity is determined by its insect vector, small jumping flea beetles. There are 9 species belonging to the genus *Phyllotreta* and a further 2 species belonging to the genus *Psylliodes*. When the beetles feed on the plant, the cell wall becomes damaged, which enables TCV to enter the plant cell and undergo its life cycle.

The three-dimensional structure of TCV has been solved by X-ray crystallography to a resolution of 3.2 Å[110]. The structure of TCV is very closely related to southern bean mosaic virus (SBMV) and TBSV, where the latter was used to elucidate the structure by phase refinement[110]. Despite these striking structural similarities, the coat proteins of these viruses do not share significant sequence similarity at the amino acid level. This is most evident by the domain organisation of the coat proteins, suggesting that similarities exist in the mechanisms of capsid assembly and uncoating.

The diameter of TCV is ~330 Å and is composed of 180 copies of a single polypeptide chain. The conformations of this polypeptide differ in a symmetry-related fashion, where these differences create 3 different quasi-conformers termed A, B and C. These conformers form non-covalent dimers (CP_2) via interactions with their C-terminal projecting (P) domains. These connect via short hinge regions to globular domains termed the shell (or S) domains, which creates the shell of the capsid. The capsid is composed of 90 CP_2, 60 of which

exist in the A/B conformation and 30 of which exists in the C/C conformation. These differ in hinge angles between the S and P domains, but are principally differentiated in their N-terminal domains. The N-terminal (R) domain forms an extended region in the polypeptide chain. In the C/C dimers, this region forms an additional β-strand structure at the edge of the S domain which extends to make contacts with the symmetry-related R domains to form an additional structure termed the β-annulus at the three-fold axis. The equivalent structures in the A/B dimers are disordered, where the electron density is not seen in the X-ray crystal structure. The N-terminal regions of the R domains in both the A/B dimers and the C/C dimers disappear towards the centre of the particle, where they presumably interact with RNA via positively-charged amino acids in the N-terminal regions.

1.7.1 *In vitro* disassembly and reassembly of TCV

Like a lot of other plant viruses, TCV contains calcium ions which function to stabilise the capsid of TCV. Conditions for the dissociation of TCV *in vitro* have been established, where incubation on ice with buffers containing EDTA, high ionic strength and alkali pH results in fully disrupted virus particles[111]. Dissociated particles can be separated into two distinct species as judged by size exclusion chromatography[34]. One of these species, termed the RP-complex, elutes in the excluded volume whereas the other species, CP_2, is separated. The RP-complex has been well characterised, and contains 1 molecule of RNA, 3 CP_2 and a unique entity known as p80[34]. The p80 protein is a covalent dimer of the TCV CP, and insensitive to separation under reducing conditions, SDS, urea, boiling and mildly acid and alkali conditions. The exact nature of the chemical cross-link in p80 has not been determined[112].

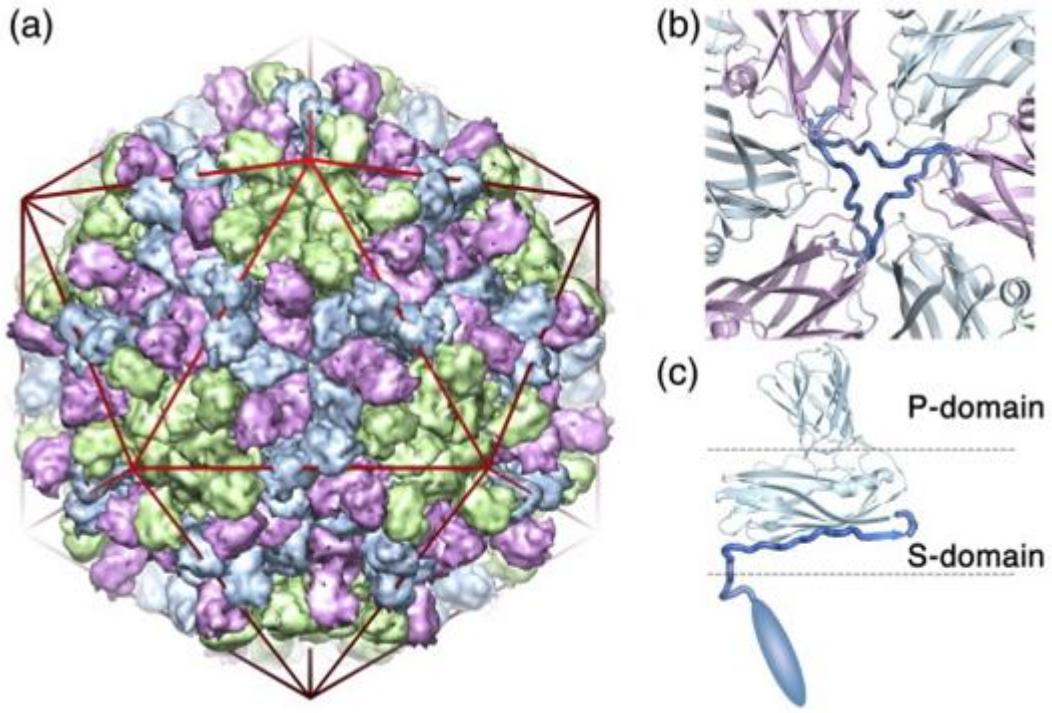


Figure 1.18: Architecture of TCV
(a) TCV capsid
TCV is organised as a capsid with *T*=3 architecture, containing 180 CP arranged into 3 conformers termed A,B and C. The A conformers (green) are packed around the 5-fold symmetry axes, whereas the B (blue) and C (pink) conformers interdigitate around the 3-fold axes. The icosahedron drawn around the capsid indicates the orientation, where it's viewed along its 3-fold axis.
(b) Structure of the β-annulus
At the 3-fold symmetry axes, the N-terminal arms from the C conformers are partially ordered and intertwine to form a structure termed the β-annulus. This structure is highlighted in dark blue, where the protein subunits are shown in cartoon representation.
(c) The domain organisation of the TCV CP
The P domain forms the surface spikes which are typical of the *Tombusviridae*, whereas the S domain forms the continuous shell of the capsid. The N-terminal regions are largely disordered, where they are proposed to bind RNA via their basic amino acids. Figure taken from Bakker *et al.* [113]

Reassembly of dissociated TCV *in vitro* can be performed using two methodologies, termed pH jump and sequential dialysis[34]. Sequential dialysis into buffers first containing high ionic strength low pH, and then low ionic strength low pH were used when the concentration of TCV was low, or when the dissociated virus was separated by size exclusion chromatography[34]. The pH jump methodology involves dilution of concentrated dissociated virus into a

buffer at physiological pH. It was shown that TCV can either be assembled from free CP_2 by using either the RP-complex or by using free RNA[34]. TCV can be reassembled by heterologous RNAs, but assembly has a strong preference for TCV RNA when subjected to other RNAs in competition experiments[34]. This selectivity was shown to be independent of p80, so therefore should be mediated by binding of 3 CP_2 to a region on the TCV genome[34].

1.7.2 The RP-complex

Formation of an RP-complex and free CP_2 during disassembly is indicative of high affinity interactions between 3 CP_2 on a unique site of RNA. As p80 is not important for capsid assembly, it is possible that it is bound to another position on the RNA molecule. In addition, it is plausible that the CP_2 would have the same conformation in the capsid as in the growing shell. The two dimer conformers in TCV come as either A/B or C/C. The conformation of the dimers is dependent on the N-terminal R domains, on whether they are in the folded conformation (C) or the extended conformation (A/B). It is unlikely that the 3 CP_2 in the RP-complex to be in the A/B conformation because conditions for RP-complex generation involve the removal of Ca^{2+} ions from the 3-fold axis, so it is unlikely that these are present to stabilise A/B dimer formation[34]. It is more likely that these 3 CP_2 are found in the C/C dimer conformation to mediate the formation of the first β-annulus structure[34].

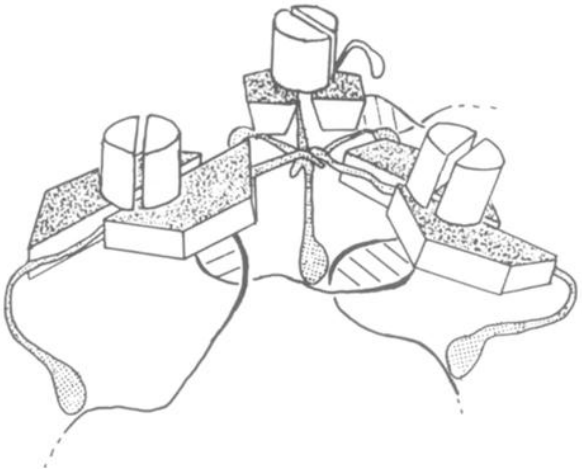


Figure 1.19: Proposed structure of the RP- complex
The model depicts three CP_2 in the C/C conformation joined by 1 set of arms which intertwine to form the β-annulus. This structure also interacts with several sites on the TCV genome. RNase treatment of the RP-complex revealed portions of RNA that were protected from digestion, revealing an RNase footprint. These areas of protection are discussed next. Figure modified from Wei *et al.,* [34].

1.7.3 Specific RNA: CP_2 interactions in TCV

Dissociated TCV can be fractionated into 2 distinct species, where one of these is the RP-complex[111]. Subsequent RNase digestion of the RP-complex revealed 2 regions of RNA protection, presumably protected by the specific interactions of the 3 CP_2 on the unique binding sites of the RNA genome[114]. These undigested sequences were then determined, and found to be on 2 general locations on the TCV genome. One of these sites, Fa, was located downstream of the amber codon on the replicase gene. The position of the binding site suggests that TCV CP binding may regulate the read-through event, thereby affecting the expression of the replicase. The second region of CP binding was located on the CP gene, where analysis of the sequences found two regions (Fd and Fe) which are separated by about 200 nt. Each of these has the potential to form stem loops secondary structures.

These two sites were further investigated in regards to TCV capsid assembly by *in vivo* deletion mutations in the polymerase gene and the CP gene[33]. Using these experiments, it was found that a 186 nt portion from the FdFe region at the 3' end of the CP gene was critical for capsid assembly, and that a bulged hairpin loop of 28 nt was the most vital element[33]. When this 186 nt portion was inserted into the CP gene of TBSV RNA in the presence of TCV CP, it was found that TBSV RNA was encapsidated. These results implicated this region of RNA essential for TCV assembly[33].

1.8 Overview of thesis

This work describes the *in vitro* assembly and disassembly of two virus systems; recombinant satellite tobacco necrosis virus-like particles (STNV VLPs) and turnip crinkle virus (TCV). In addition, the three dimensional X-ray crystal structure of STNV encapsidating the aptamer B3 is described, as well as how RNA sequence specificity affects the capsid assembly of STNV VLPs *in vitro*.

In Chapter 2, the Materials and Methods of the general techniques used with Chapter specific details are provided as appropriate. Chapter 3 details the recombinant expression and purification of recombinant STNV VLPs and the role of the basic N-terminus during capsid assembly. The RNA-dependence of capsid assembly is described and the purification of STNV CP from STNV VLPs is explained, including how proteolysis was controlled. Purified STNV CP was also extensively characterised biochemically.

Chapter 4 describes an analytical ultracentrifuge sedimentation velocity (svAUC) assay which helps characterise the efficiency of capsid assembly through STNV CP titrations. These titrations are performed using a fixed concentration of RNA, where several RNAs were examined which differed in sequence and length. TEM images from each titration complement the results from the svAUC assay. Furthermore, large-scale production and purification of STNV-B3 VLPs were characterised by X-ray crystallography. The X-ray crystal structure of STNV-B3 is described and a capsid assembly mechanism is proposed based on this structure.

Chapter 5 describes evidence for the uncoating mechanisms of TCV and STNV, respectively, and Chapter 6 is a general discussion which puts this work into context with the literature already known for this class of viruses.

2. Materials and methods

2.1 Materials

2.1.1 General Chemicals

All chemicals were purchased from Sigma Aldrich Co, Fisher Scientific Inc., or Ambion.

2.1.2 Molecular Biology Reagents

All restriction enzymes were manufactured by New England Biolabs. DNA size markers were purchased from Invitrogen. Prestained protein markers were purchased from New England Biolabs. Lysozyme, DNase I and Streptavidin were purchased from Sigma Aldrich Co.

2.1.3 Competent cells

Protein expression was performed with transformed BL21 (DE3) Gold pLysS (Stratagene) competent cells. Plasmid preparations were performed using HiSpeed Plasmid Maxi Kit (QIAGEN) on transformed XL-1 Blue competent cells (Stratagene). See Appendix, Table 9.1 describing genotypes of these strains.

2.1.4 DNA primers

All DNA primers were purchased from Eurofins MWG.

2.1.5 RNA oligonucleotides

All RNAs of ≤ 30 nt were synthesized by Amy Barker as described in Chapter 2.4.4.

Table 2.1: Buffer composition

All buffers were autoclaved or filtered through a 0.2 µm filter prior to use.

Buffer	Recipe
1 X STNV native buffer	50 mM HEPES, 200 mM NaCl pH 7.5
1 X STNV disassembly buffer	50 mM HEPES, 1 M NaCl, 3 mM EDTA pH 8.5
10 X STNV disassembly buffer	500 mM HEPES, 100 mM EDTA, pH 8.5
1 X STNV reassembly buffer	50 mM HEPES, 3 mM $CaCl_2$ pH 7.5
10 X SDS-PAGE running buffer	30.3 g Tris, 144 g Glycine, 10 g SDS, make up to 1 L with ddH_2O
Coomassie brilliant blue	30 g $(NH_4)_2SO_4$, 10 % v/v Acetic acid, 0.5 g Coomassie stain (G-250), make up to 1 L with ddH_2O
2 X Laemmli sample loading buffer	0.625 mL 1M TRIS-HCl pH 6.8, 2 mL 10 % SDS, 1 mL Glycerol, 0.5 mL β- mercaptoethanol, 1 g Bromophenol blue, 0.875 mL ddH_2O
SDS-PAGE stacking buffer	12.5 mL 1 M Tris-HCl, pH 6.8, 1 mL 10 % SDS, 76.5 mL ddH_2O.
SDS-PAGE resolving buffer	94 mL Tris-HCl pH 8.8, 2.5 mL 10% SDS, 3.5 mL ddH_2O
10 X MOPS buffer	400 mM MOPS free acid, 100 mM NaOAc, 10 mM EDTA, pH 7
Luria Broth (rich)	20 g tryptone, 10 g NaCl, 10 g yeast extract, adjusted to pH 7 with NaOH
LB agar	1 L LB media and 20 g agar
5 X TBE buffer	54 g Tris, 27.5 g Boric acid, 20 mL of 50 mM EDTA solution, adjust to pH 8 and top up to 1 L with ddH_2O
1 X TCV native buffer	10 mM sodium phosphate, 10 mM $MgCl_2$ pH 7.5
10 X TCV expansion buffer	1 M Tris, 50 mM EDTA pH 8.5
2 X TCV disassembly buffer	200 mM Tris, 10 mM EDTA 2 M NaCl pH 8.5

2.2 General methods

2.2.1 Electrophoresis

2.2.1.1 SDS-PAGE gel preparation

Polyacrylamide gels were made using resolving buffers and stacking buffers, respectively, detailed in Table 2.1. Resolving gels (15 %) were prepared using 4 mL resolving buffer, 1 mL ddH_2O and 5 mL 30 % acrylamide: bisacrylamide (37.5:1) (SequaGel, National Diagnostics). Polymerisation was initiated by adding 20 µL tetramethylethylenediamine (TEMED) (Lancaster Synthesis) and 100 µL ammonium persulphate (APS) (Fisher Scientific). Gels were poured into 0.75 mL spacer plates (Biorad) and the interface was made by pouring a minimal volume of 50 % (v/v) isopropanol (Fisher Scientific) on to the top of the gel before the gel polymerised. Stacking gel (3 %) was prepared with 4.5 mL stacking buffer and 0.5 mL 30 % acrylamide: bisacrylamide (37.5:1) (SequaGel). Polymerisation was initiated by adding 10 µL TEMED and 50 µL APS. The isopropanol was removed from the resolving gel and stacking gel was poured to the top of the gel. A 10-well or a 15-well gel comb (Biorad) was added before the stacking gel was allowed to set.

2.2.1.2 SDS-PAGE

Samples (15 µL) were prepared by adding an equal volume of 2 X Laemmli sample loading buffer (Sigma Aldrich) and boiling for 5 mins. 10-15 µL of sample was loaded into each well of the SDS-PAGE gel before being run at 150 V for 1.5-2 h in 1 X SDS-PAGE running buffer. The gel was removed from the gel plates and the stacking gel was removed. The gel was fixed in 12.5 % trichloroacetic acid (TCA, Fisher Scientific) before being stained by Coomassie

brilliant blue G-250 (Fisher Scientific). Coomassie stain was removed after 2-24 h and the gel was washed with deionised H_2O before being de-stained in deionised H_2O until the background was removed. Gel images were captured using the Syngene Gel Doc.

2.2.1.3 Denaturing RNA agarose gel electrophoresis

RNA was separated by electrophoresis on a 1 % (w/v) denaturing RNA agarose gel. The agarose gel was prepared by melting 0.5 g agarose in 36 mL of distilled water before 5 mL of 10 X MOPS buffer and 9 mL 37 % (v/v) formaldehyde (Sigma) was added to the molten agarose. RNA samples (0.5 µg) containing 0.1 ng ethidium bromide were prepared in 2 X gel loading buffer (Ambion) and boiled for 1 min at 95 ºC. The samples were run at 1-5 V/cm in 1 X MOPS buffer for ~1 h before visualising using the transilluminator filter on the Syngene Gel Doc.

2.2.1.4 Denaturing RNA polyacrylamide gel electrophoresis

Denaturing PAGE gels (10 %) were prepared using the 29:1 UreaGel system (SequaGel, National diagnositics). The gels were pre-run for ~2 h in TBE buffer at 15 W before the RNA samples were prepared with 2 X gel loading buffer (Ambion) and incubated at 95 ºC for 2 mins. The sample wells were rinsed with syringing buffer into the wells before the samples were loaded, including a 10 bp ladder (NEB). The gel was run at 15 W for ~1 h in 1 x TBE buffer and stained with ethidium bromide solution for ~20 mins. The gel was rinsed in deionised H_2O before being visualised using the transilluminator filter on the Syngene Gel Doc.

2.2.2 Preparation of grids for transmission electron microscopy (TEM)

Each grid (Agar Scientific) was irradiated with UV for 15 mins. A 5 µL droplet of sample was added onto the carbon-coated grid and allowed to adsorb onto the grid for 30 s. After this incubation time, the grid was dabbed onto filter paper (Whatman No. 1®) to remove excess unbound protein and the grid was stained by pipetting approximately 30 µL of 1 % (w/v) uranyl acetate onto the grid surface and the excess removed by flicking onto filter paper. This was performed three times before dabbing off further excess onto filter paper. The grids were then left to air-dry for 5 mins before they were visualised.

The images were captured using a JEOL 1200 EX transmission electron microscope (TEM) operating at 80 keV; a Phillips CM 10 operating at 80 keV or a Technai F12 Spirit operating at 120 keV. Images were taken using an exposure time ~1-2 s, automated based on the brightness of the beam. Film was captured using Kodak ISO film (JEOL 1200 EX and Phillips CM 10) or CCD camera (Technai F12 Spirit). Film was developed by Martin Fuller and digitised using an Imacon FlexTight 848 scanner.

2.2.3 Phenol: chloroform extraction and ethanol precipitation of RNA

STNV VLPs (0.5 mg) were added to 500 µL of acidic phenol: chloroform (25:24:1) (Applied Biosystems) and vortexed until the solution went cloudy. The sample was centrifuged at 16,300 g for 1 min and the upper aqueous layer removed from the phenol. This step was repeated before loading and vortexing with an equal amount of chloroform: isoamyl alcohol (24:1) (Applied Biosystems) and centrifuged at 16,300 g for 1 min. This step was repeated to ensure that the phenol was removed from the RNA sample. Two volumes of

ethanol and a tenth volume of 3 M sodium acetate pH 5.2 were added and the sample was incubated at -80˚C for 3-16 h to precipitate the RNA. The precipitate was pelleted by centrifugation at 16,300 g for 30 mins at 4ºC and the ethanol removed whilst being careful not to dislodge the pellet. The pellet was washed carefully with 70 % (v/v) cold ethanol and vacuum dried before being resuspended in 20 µL of nuclease free water. The RNA concentration was calculated by its UV absorbance using the equation:

Equation 2.1: µg/mL RNA = 40 X A_{260}.

2.2.4 Transformation

A 50 µL aliquot of competent *E. coli* cells (BL21 (DE3) Gold pLysS (Statagene) for protein expression; XL-1 Blue (Stratagene) for plasmid purification) was thawed on ice. Upon thawing, 1 µL (150 µg/mL) of plasmid (pET22b vector containing the synthetic STNV gene; pUBS-STNV-C; pSMART-5'-MS2) was mixed and incubated on ice for 30 mins. The cells were heat shocked by transferring to a 42°C water bath for 90 s before being incubated on ice for a further 5 mins. The heat-shocked *E. coli* were transferred to 0.5 mL of sterilized Luria Broth (LB) and the cells were incubated for 1 hour at 37°C. After incubation, 20 µL from the transformed *E. coli* cells were plated out onto LB agar containing 100 µg/mL ampicillin and incubated at 37°C overnight.

2.3 Methods for Chapter 3

2.3.1 STNV VLP protein expression

Single colonies of STNV-transformed *E. coli* were picked and grown in 5 mL of LB-ampicillin (100 µg/mL). The inoculation was incubated overnight at 37°C,

160 rpm. The initial 5 mL cultures were used to inoculate 500 mL flasks of LB containing 100 µg/mL ampicillin and the suspension was shaken at 160 rpm and at a temperature of 37°C in an orbital inhibitor. Once the cultures reached an OD_{600} of 0.6, the cultures were subjected to a heat shock for 30 mins at 50°C. The cultures were cooled to 24°C for 20 mins before being induced with isopropyl-β-D-thio-galactopyranoside (IPTG) to a final concentration of 1 mM. The transformed cells were incubated at 24°C for a further 24 h before being harvested by centrifugation at 4°C and 5251 g. A time course was prepared to detect the level of protein expression over time, where 1 mL of culture was taken every hour after IPTG induction, starting at *T=0*. The OD_{600} was measured before the cells were pelleted by centrifugation at 16,300 g. A volume of Laemmli sample loading buffer was added according to the equation:

Equation 2.2: *OD_{600} x 50 = µL Laemmli sample loading buffer*

The pellets from each time point were resuspended and 3.5 µL of each were analysed by 15 % SDS-PAGE as described in Chapter 2.2.1.2.

2.3.2 Cell harvest and lysis

Cell pellets were resuspended in 100 mL STNV native buffer and homogenised by pipetting, ensuring that there were no large ‘clumps’. The homogenised solution was incubated at room temperature for 30 mins in the presence of 1 mg lysozyme (Sigma) prior to five cycles of 30 s on / off sonification. After sonification, 1 mg DNase I (Sigma) was added to the cell suspension and incubated at 4°C for 1 h. The cell suspension was centrifuged at 15,000 rpm (26,892 g) for 30 mins at 4°C to remove any cell debris from the supernatant.

2.3.3 Ammonium sulphate precipitation

The supernatant was ammonium sulphate precipitated by slowly dissolving ammonium sulphate until reaching a final concentration of 1.75 M (23 % w/v). The precipitate was centrifuged at 15,000 rpm (26,892 g) in a Sorvall SS-34 rotor to pellet. The pellet was resuspended in 50 mL STNV native buffer and dialysed against 5 L of STNV native buffer to remove the ammonium sulphate.

2.3.4 Q-Sepharose anion exchange chromatography

The dialysed sample was loaded onto an XK 26/60 pre-equilibrated Q-Sepharose^{TM} anion exchange column at a flow rate of 7 mL/min and the 'flow through' was collected. The fractions absorbing at 260 and 280 nm were analysed by 15 % SDS-PAGE, and the fractions containing STNV were pooled and concentrated by ammonium sulphate precipitation as described in 2.3.3.

2.3.5 Size exclusion chromatography

The ammonium sulphate precipitated sample was resuspended in 5 mL of native buffer and dialysed extensively against native buffer to remove the ammonium sulphate. The sample was loaded onto a pre-equilibrated Superdex-200 (Pharmacia^{TM} HiLoad 26/60) column at a flow rate of 2 mL/min. The fractions that absorbed at 260 and 280 nm were analysed by 15 % SDS-PAGE and TEM.

2.3.6 Sucrose density gradient sedimentation

The precipitate obtained from ammonium sulphate precipitation was centrifuged in a Sorvall SS-34 rotor at 15,000 rpm and the pellet was resuspended in 12-18 mL of 50 mM HEPES 200 mM NaCl pH 7.5. The ammonium sulphate was

removed by extensive dialysis against native buffer before carefully pipetting 2-3 mL on each gradient and centrifuging using a SW32ti rotor at 18,500 rpm on a 15-45 % (w/v) sucrose density gradient for 21 h at 4°C. Fractions were collected across the gradient and peaks that absorbed at 260 and 280 nm were analysed by 15 % SDS-PAGE. The fractions containing STNV were pooled and dialysed extensively against native buffer to remove the sucrose.

2.3.7 One-pot disassembly and reassembly of STNV

Disassembly was achieved by dialysis of 1 mL of recombinant STNV VLPs (1 mg/mL) into 3 L of disassembly buffer for 3 h at 4ºC. Reassembly was achieved by taking the disassembled particles and dialysing in 3 L of reassembly buffer for 18 h at 4ºC. Characterisation of native recombinant STNV, disassembled recombinant STNV and reassembled recombinant STNV was carried out by analysing each sample by Superdex-200 10/300 GL gel filtration chromatography (flow rate 0.5 mL/min). The peak fractions from the chromatogram (absorbing at 280 and 260 nm) were analysed by SDS-PAGE and visualised by transmission electron microscopy (TEM).

2.3.8 One-pot disassembly/ reassembly of STNV in the presence of α,2-macroglobulin (α_2M)

Disassembly was achieved by dialysis of 0.5 mL of recombinant STNV VLPs (2 mg/mL), in the presence of 1 unit of α_2M, into 3 L of disassembly buffer for 6 h at 4ºC. Reassembly was achieved by dialysis into 3 L of reassembly buffer for 18 h at 4ºC. Characterisation of native recombinant STNV, disassembled recombinant STNV and reassembled recombinant STNV was carried out by analysis by Superdex-200 10/300 GL size exclusion chromatography (flow rate

0.5 mL/min). The peak fractions from the chromatogram (absorbing at 280 and 260 nm) were analysed by SDS-PAGE. An aliquot of each sample was taken prior to size exclusion chromatography and visualised by transmission electron microscopy (TEM) as described in Chapter 2.2.2.

2.3.9 STNV coat protein (CP) purification

STNV VLPs (12 mL, 1 mg/mL), 100 µL complete protease inhibitor (Roche) and 10 µg Pepstatin A, were mixed and incubated for 30 mins. The STNV VLPs were disassembled by adding 1/5th volume of 10 X disassembly buffer and incubated by agitation for 5 mins at room temperature. The sample was loaded onto a pre-equilibrated 1 mL Q-Sepharose FF and SP Sepharose^{TM} FF ion exchange columns in series at a flow rate of 0.2 mL/min. The sample was flushed through with 15 mL of 1 X disassembly buffer minus NaCl to ensure all the sample was loaded onto the columns. The Q-Sepharose^{TM} FF anion exchange column was removed and the SP Sepharose^{TM} FF-column was washed with 20 column volumes of disassembly buffer minus NaCl, before STNV CP was eluted with a 20 mL gradient of 2 M NaCl in 1 X disassembly buffer. STNV CP was collected in 2 mL fractions and were analysed by SDS-PAGE and UV spectroscopy. The protein concentration was calculated by $A = \varepsilon cl$, where ε = 16,960 M^{-1} cm^{-1} and the molecular weight is 21,584 Da.

2.3.10 STNV assembly in the presence/absence of aptamer B3

STNV CP was purified according to Chapter 2.3.9. Aptamer B3 was synthesized by Amy Barker, according to Chapter 2.4.4. STNV CP (0.5 mL, 0.5 mg/mL) was incubated against reassembly buffer in the presence (80 µg/mL) or absence of aptamer B3 using 0.1-0.5 µL Slide-A-Lyzers (3 kDa MWCO)

(Pierce). An aliquot was taken from each sample for TEM as described in Chapter 2.2.2. Each sample was analysed by size exclusion chromatography (S-200 10/300 GL) at a flow rate of 0.5 mL/min. Fractions with a UV absorbance were collected in 2 mL fractions and concentrated by Amicon Ultra® spin concentrators (Millipore®) before analysis by SDS-PAGE according to Chapter 2.2.1.2.

2.4 Methods for Chapter 4

2.4.1 Preparation of template DNA by restriction digest

Plasmids pET22b-STNV[89], pUBS-STNV-C[115] and pSMART-5'-MS2[116] were linearised (10 µg) by incubation at 37ºC for 1 h with 2 units of restriction enzymes HindIII (for pET22b-STNV and pSMART-5'-MS2) and XhoI (for pUBS-STNV-C) (NEB), respectively. The linearised plasmids were analysed by 1 % native agarose gel electrophoresis. Restriction enzymes were heat-inactivated by incubation at 65ºC for 30 mins. Plasmid DNA concentration was determined using Equation 2.3.

2.4.2 Preparation of template DNA by PCR

MS2 iRNA template was produced from pSMART-5'-MS2 by PCR amplification with primers 1419F_T7 and 2346R as described in [116]. PCR products were analysed by 1 % native agarose gel electrophoresis. DNA concentration was determined using Equation 2.3.

Equation 2.3: [DNA] (µg/mL) = A_{260} x 50

2.4.3 *In vitro* transcription of STNV CP mRNA, STNV-C, MS2 5'RNA and MS2 iRNA

1 µg of DNA template was used in each transcription reaction using MEGAScript *in vitro* transcription kit (Ambion) following manufacturer's guidelines. The DNA templates were digested after transcription using turbo DNase I obtained with the kit (Ambion). These transcripts were purified by phenol: chloroform extraction and ethanol precipitation as described in Chapter 2.2.3. Nucleotides were removed by Illustra G-25 microspin desalting columns® (GE Healthcare) using the manufacturer's guidelines. All RNAs were analysed by denaturing agarose gel electrophoresis as described in Chapter 2.2.1.3 and the concentration was determined by UV absorbance as described in Equation 2.4:

Equation 2.4: [RNA] (µg/mL) = 40 x A_{260}

2.4.4 RNA oligonucleotide synthesis

RNA oligonucleotides were synthesised by Amy Barker. Aptamer B3 sequence (5'-CCU UUU CAA GAC AUG CAA CAA UGC ACA CAG-3'); Aptamer B3 4U sequence (5'-CCU UUU CAA GAC AUG CAU UUU UGC ACA CAG-3'); B3 short sequence (5'-CGU GCA ACA AUG CAC G -3'); MS2 TR sequence (5'-ACA UGA GGA UUA CCC AUG U-3'), were synthesized using *N*-benzoyl-protected adenosine, *N*-dimethylformamidinyl-protected guanosine and *N*-acetyl-protected cytosine; no protective group required for uracil. The 2' hydroxyl group was protected by t-butyldimethylsilyl (TBDMS) and the 3' was activated by 2-cyanoethyl-(*N,N'*-diisopropyl)-phosphoramidite (Link Technologies). Ammonia-saturated methanol was used to remove protecting

groups and to cleave RNA from controlled pore glass (CPG) resin at room temperature for 24 hours. Methanol was removed under vacuum and the pellet was resuspended in anhydrous dimethyl sulfoxide (DMSO). One volume of triethylamine trihydrofluoride (TEA.3HF) was added and incubated at room temperature to remove TBDMS from the 2' hydroxyl groups. RNAs were butan-1-ol precipitated and resuspended into diethylpyrocarbonate (DEPC) treated water (Severn Biotech). Each RNA was purified by ion exchange high performance liquid chromatography (HPLC) and analysed by negative mode electrospray ionisation mass spectrometry and denaturing PAGE as described in Chapter 2.2.1.4.

2.4.5 STNV capsid assembly titrations

All RNAs were prepared according to Chapter 2.4.3 or 2.4.4, respectively. RNAs (20 µg/mL) were kept at a fixed concentration and STNV CP was either omitted or added at 1:1, 2:1, 5:1 or 10:1 (w/w) ratios in 500 µL total volume. Each reassembly reaction was extensively dialysed in reassembly buffer overnight at 4ºC using 0.1-0.5 mL Slide-A-Lyzers (3 kDa MWCO) (Pierce).

2.4.6 Analytical ultracentrifugation – sedimentation velocity (svAUC)

2.4.6.1 Background

Molecules in solution can be separated and characterised based on their rates of sedimentation under an applied centrifugal force. This type of analysis can be performed by analytical ultracentrifugation. The analytical ultracentrifuge is similar to any other ultracentrifuge, other than it has a laser light detection system that allows the direct measurement of the macromolecules in solution, thus the rate of sedimentation can be followed. This can be achieved by

measuring the change in UV absorbance of the sedimenting macromolecules in solution over time.

Analytical ultracentrifugation can be used to determine the molecular weight, the gross shape and can also give information on the heterogeneity of the sample. Information regarding the latter can be obtained by even trace amounts of oligomers or aggregates, which is very useful in the characterisation of complexes in solution. Two basic types of analysis can be performed with the analytical ultracentrifuge, namely sedimentation velocity (svAUC) and sedimentation equilibrium (seAUC) experiments.

One of the fundamental advantages of sedimentation velocity experiments is that biological macromolecules of nearly any size can be investigated in solution, within a very broad range of buffer conditions. The sedimentation of macromolecules is dependent on three factors; gravitational force, buoyancy and the hydrodynamic friction. These parameters are described by the Svedberg equation:

Equation 2.5: $\frac{S}{D} = \frac{M(1-\bar{\nu}\rho)}{RT}$

2.4.6.2 Method

Sedimentation velocity was performed by Amy Barker. Each sample (0.32 mL) was placed in a 1.2 cm pathlength 2-sector miniscus-matching epon centrepiece cell constructed with sapphire windows. These samples were centrifuged at 15,000 rpm in an An50-Ti rotor in an Optima XL-1 analytical ultracentrifuge at 20ºC. Changes in absorbance at 260 nm in the solute were

detected by absorbance optics, with a total of 100 scans being taken in approximately 12 hours. Buffer densities and viscosities were calculated by Sednterp version 1.09[117]. Radial absorbance profiles were fitted using the program Sedfit version 12.1b using a continuous distribution c(s) Lamm equation model[118].

2.4.7 X-ray crystallography of STNV VLPs

2.4.7.1 Production and purification of VLPs for X-ray crystallography

Recombinant STNV VLPs (~12mg) were disassembled and STNV CP was purified by ion exchange chromatography using the methodology described in Chapter 2.3.9. The total amount of CP was determined by UV spectrophotometry and analysed by SDS-PAGE. STNV CP was added to each RNA (aptamer B3, B3 short and TR) at a ratio of 5:1 (w/w) before being extensively dialysed in reassembly buffer. Reassembled VLPs were concentrated to 500 µL by Amicon Ultra® spin concentrators (Millipore) and purified by size exclusion chromatography (S-200 10/300 GL) at a flow rate of 0.5 mL/min. Fractions containing purified VLPs were analysed by SDS-PAGE, TEM and sedimentation velocity to assess the quality of the sample.

2.4.7.2 Crystallisation conditions

STNV-B3 crystallisation and data collection was performed by Saskia Bakker. The VLPs were crystallised using the conditions previously established for recombinant STNV[89]. Reservoir solutions contained 50 mM phosphate at pH 6.2-6.8, 0.2-0.8 % (w/v) PEG 6000 and 0-1 mM $MgCl_2$. Drops were set up using 3 µl VLP (7 mg/mL) and 3 µl reservoir solution. Diamond-shaped crystals

appeared after 4 weeks of incubation at 25 °C and measured about 0.1 mm in their longest dimension.

Crystals were prepared for data collection by cryoprotection with glycerol before flash cooling in liquid nitrogen. Addition of glycerol was carried out in a stepwise fashion as follows, 60 s soak in 20 % (v/v) glycerol, 80 % (v/v) reservoir solution followed by a 20 s soak in 25 % (v/v) glycerol, 75 % (v/v) reservoir solution and finally a 1 s soak in 30 % (v/v) glycerol, 70 % (v/v) reservoir solution.

2.4.7.3 Data collection

Diffraction data were collected at Diamond Light Source on beam line I02 at a wavelength of 0.9795 Å to a maximum resolution of 2.29 Å. To enable examination of the less ordered internal structure of the VLP high quality low resolution data was required. Three sweeps of data were therefore recorded from the same region of the crystal with varied detector and beam stop positions with resolution ranges of 116-3.8 Å, 217-7 Å and 30-2.29 Å. 180° of data were recorded for each sweep with an oscillation angle of 1° for the very low resolution data and 0.5° for the medium and high resolution sweeps. All three sweeps of data were processed together using the 3dii option of the xia2 expert data reduction system.

2.4.7.4 Phasing and refinement

Phasing and refinement was carried out by Drs Saskia Bakker and Arwen Pearson. The data were isomorphous with recombinant STNV VLP. (PDB 1VTZ & 3RQV) and this structure was used to provide initial phases for map calculation. Rigid body refinement of the entire capsid was carried out using

REFMAC5. 60-fold averaged maps were calculated using COOT after which the coat protein structure was rebuilt and refined, using strict 60-fold NCS, with REFMAC5.

2.4.7.5 Examination of the internal structure

This was performed by Dr Arwen Pearson. 60-fold sigmaA weighted 2m|Fo|-D|Fc| and m|Fo|-D|Fc| averaged maps were calculated using the Uppsala Software Suite. A mask comprising a single coat protein and the corresponding internal segment of the VLP was created using MAMA. 60-fold averaging of the mask was carried out using MAVE.

2.4.7.6 Accession numbers

Accession numbers were assigned by Dr Arwen Pearson. The co-ordinates and structure factors for the protein capsid in the presence of the B3 aptamer RNA have been deposited with the PDB.

2.4.8 Purification of STNV-B3 'half capsids'

STNV 'half capsids' were prepared from 0.5 mg STNV CP (purified as described in Chapter 2.3.9) and 250 µg aptamer B3 (prepared as described in Chapter 2.4.4) at a 2:1 (w/w) ratio. Sample was dialysed extensively in reassembly buffer at 4ºC before being purified by 15-45 % (w/v) sucrose density gradient fractionation in 12 mL tubes (Seton). STNV half capsids were analysed by UV absorbance, visualised by TEM and characterised by Amy Barker using sedimentation equilibrium.

2.4.9 RNase A treatment of reassembled STNV VLPs to remove aggregates

100 µL STNV VLPs reassembled with STNV-C and MS2 5'RNAs (5:1 CP: RNA (w/w)), respectively, were treated with 1 µL (2.5 units) RNase A (Sigma Aldrich) and incubated at room temperature for 30 mins. The STNV VLPs were diluted two-fold before being prepared for TEM according to Chapter 2.2.2.

2.5 Methods for Chapter 5

2.5.1 Purification of Turnip Crinkle Virus

Turnip Crinkle Virus was purified from frozen leaves of *Nicotiana benthamiana* infected with 35S TCV (derived from TCV-M cDNA) in an *Agrobacterium tumifaciens* vector[110; 119]. Leaves were kindly provided by Dr Keith Saunders and Prof. George Lomonossoff, John Innes Centre, Norwich. The leaves (100 g) were homogenised by blending with distilled water on ice before being strained through muslin under vacuum. Celite (Supelco, 5g per 100 mL of sap) was added and the mixture was centrifuged at 15,000 rpm (26,892 g) for 30 mins using a Sorvall SS-34 rotor at 4 °C. The green pellet was discarded and 30% (w/v) PEG 6000 (Fluka) and 0.8 mL of 5 M NaCl (Fisher Scientific) were added to the supernatant and stirred for 10 mins at 4 °C. The mixture was centrifuged for 30 mins at 15,000 rpm (26,892 g) as above. The pellets were resuspended in minimal volume of native buffer (10 mM sodium phosphate, 10 mM magnesium sulphate at pH 7.4) and the extracts were centrifuged in a Beckmann preparative ultracentrifuge with SW55Ti rotor for 2 h at 35,000 rpm at 4°C. Pellets were resuspended in native buffer and 10 % (v/v) Bentonite (36 mg/mL) was added to resuspended virus to remove impurities which bind to the Bentonite. The suspension was centrifuged for 30 mins at 15,000 rpm (26,892

g) to remove the Bentonite. The supernatant was spun using a SW55Ti rotor for 2 h at 4°C and 35,000 rpm, pellets were resuspended in TCV native buffer and spun for 10 mins at 10,000 rpm in an SS-34 rotor at 4ºC.

These last two steps were repeated until the ultracentrifuge supernatant was clear. The virus was further purified on a 10-40% sucrose gradient, using a Beckmann preparative ultracentrifuge with an SW40Ti rotor, and spun at 35,000 rpm for 2 h at 4°C. The fractions containing TCV were pooled and dialysed using SnakeSkin^{TM} pleated dialysis tubing with a molecular weight cut-off of 3.5 kDa (Pierce Thermo Scientific) against 5 L TCV native buffer for 24-48 h, changing the buffer twice, until the sucrose is removed. Purified TCV was analysed by 15 % SDS-PAGE and the concentration was calculated. Concentration of the virus was determined by UV absorbance and calculated using Equation 2.6 [120].

Equation 2.6: [protein (mg/mL)] = (1.55 x A_{280} – 0.76 x A_{260}) x dilution factor

2.5.2 Expanded TCV chymotrypsin digestion

250 µL TCV (4 mg/mL) was expanded by adding 100 µL 10X expansion buffer and incubated at room temperature for 10 mins. TCV was subjected to protease digestion by addition of 2.5 µg of chymotrypsin and incubated for 30 mins at room temperature. The reaction was quenched by the addition of 100 µL of complete protease inhibitor cocktail (Roche) and 160 µg Pepstatin A (Roche). The subsequent proteolysed TCV was separated by size exclusion chromatography using 10/300 GL Superdex-200 column run at 0.5 mL/min and 1 mL fractions were collected. Fractions that contained UV absorbance were

concentrated (500 µL) by Amicon Ultra® centrifugal concentrators. The concentrated fractions were analysed by SDS-PAGE as described in Chapter 2.2.1.2. The eluted TCV VLPs were analysed by TEM as described in Chapter 2.2.2.

2.5.3 Expanded and proteolysed TCV striposome formation

The purified expanded, chymotrypsin treated TCV (10 µL, 4 mg/mL) was prepared as described in Chapter 2.5.2. The sample was incubated for 30 mins in the presence of wheat germ extract as per manufacturer's instructions (Promega). Translation was terminated by adding 1 mM cycloheximide final volume. The sample was centrifuged at 16,300 g for 10 mins at 4ºC and 5 µL was prepared for TEM as described in Chapter 2.2.2.

2.5.4 RNase A treatment of native, expanded, expanded and chymotrypsin treated TCV

The different states of TCV were prepared as described in Chapter 2.5.2. Each sample was incubated in the presence of 0.5 µL RNase A (1.25 units) (Sigma Aldrich) for 30 mins at room temperature before being quenched by 1 µL (40 units) RNase OUT^{TM} (Invitrogen). RNAs were purified from TCV treated with RNase A in each state by RNeasy® RNA extraction kit (QIAGEN) according to manufacturer's instructions. The RNA from untreated TCV was also extracted as a negative control. Each purified RNA was analysed by 1 % (w/v) denaturing agarose gel electrophoresis as described in Chapter 2.2.1.3.

2.5.5 Chymotrypsin treatment and biochemical analysis of RP-complex

Purified TCV (200 µL, 75 mg/mL) was dissociated by adding 200 µL of 2X TCV disassembly buffer and 100 µL of complete protease inhibitor and incubated for 30 mins. The RP-complex was purified from free coat protein dimer by Superdex-200 10/300 GL size exclusion chromatography at a flow rate of 0.5 mL/min, where the RP-complex eluted in the void volume. The resulting fractions were concentrated to 400 µL by Amicon Ultra® spin concentrators (Millipore) and treated with chymotrypsin (Roche) at a final concentration of 1 µg/mL. The digested RP–complex was quenched by adding 1 µg of Aprotinin and 100 µL ($1/10^{th}$ of a tablet) of complete protease inhibitor (Roche). The digested RP-complex was separated by size exclusion chromatography using a Superdex-200 10/300 GL column, where fractions which absorbed at 260 and/or 280 nm were spin concentrated (Amicon Ultra®, Millipore) and analysed by 15 % SDS-PAGE as described in Chapter 2.2.1.2.

2.5.6 Chymotrypsin treatment of recombinant STNV VLPs

STNV VLPs (25 µL, 1 mg/mL) were chymotrypsin (0.5 µg) (Roche) treated and incubated for 1 hour at room temperature before being quenched by 10 µL Complete protease inhibitor tablets (1 tablet dissolved in 1 mL ddH_2O) (Roche). Chymotrypsin treated and untreated STNV VLPs were the analysed by SDS-PAGE as described in Chapter 2.2.1.2.

2.5.7 Chymotrypsin treatment and biochemical analysis of expanded STNV VLPs

STNV VLPs (250 µL, 4 mg/mL) were treated with EDTA (10 mM) and incubated at 4ºC for 24 h. Chymotrypsin was added (5 µg) and incubated for 30 mins at

room temperature before being quenched by 10 µg Aprotinin and 100 µL Complete protease inhibitor tablet (1 tablet dissolved in 1 mL ddH_2O) (Roche). The chymotrypsin treated, expanded STNV VLPs were separated by size exclusion chromatography using a Superdex-200 10/300 GL column, where fractions which absorbed at 260 and/or 280 nm were spin concentrated (Amicon Ultra®, Millipore) and analysed by 15 % SDS-PAGE as described in Chapter 2.2.1.2. The eluted STNV VLPs were analysed by TEM and incubated with wheat germ extract as described in Chapters 2.2.2 and 2.5.3, respectively.

3. The role of the N-terminal arm during capsid assembly

3.1 Introduction

Previously, the STNV coat protein (CP) gene was cloned into a pET22b expression vector that was codon optimised for efficient protein expression in *E. coli*. The expressed CP spontaneously assembled into virus-like particles (STNV VLPs)[89]. These STNV VLPs can be purified in milligram quantities with sufficient purity to be crystallised for structural studies, and can be disassembled and reassembled *in vitro*[81]. Furthermore, the sequence specificity of STNV assembly was investigated, where the generation of RNA aptamers against the RNA binding sites of STNV CP revealed structural motifs that are present in the wild-type genome[81].

Detailed in this Chapter is how improvements that have been made to the STNV purification protocol have contributed to furthering our understanding of STNV capsid assembly. These improvements enhanced yield, minimised proteolysis, and therefore maximised the efficiency of STNV capsid assembly *in vitro*. In addition, investigations into STNV CP purification from the VLP were examined to ensure *in vitro* capsid assembly experiments with heterologous RNAs could be carried out. Aptamer B3, an obvious RNA candidate based on the previous SELEX experiments[81], was used in capsid assembly studies.

.

3.2 Results

3.2.1 STNV CP expression

The pET22b-STNV[89] vector was transformed in BL-21 (DE3) Gold pLysS cells (Stratagene) as described in Chapter 2.2.4. This strain of bacteria was used to prevent the 'leaky' expression that can sometimes occur when using other strains of competent cells. The transformation efficiency routinely obtained was between 0.1-6 X 10^9 cfu/µL. Single colonies were then used to inoculate overnight cultures of LB-ampicillin as described in Chapter 2.3.1. These were used as starter cultures for inoculation of 500 mL flasks of LB-ampicillin, and incubated with shaking (160 rpm) at 37ºC until an OD_{600} of 0.4 - 0.6 was reached.

Upon reaching the required OD_{600}, the cultures were subjected to a heat shock. A sudden upsurge in temperature in *E. coli* induces the expression of heat shock proteins *in vivo* where molecular chaperones, such as DnaK, DnaJ, GrpE, GroEL and GroES, are expressed. These proteins constitute two of the major chaperone systems within *E. coli* and are important to prevent undesired aggregation of proteins as well as refolding proteins that have been misfolded[121; 122]. These properties are obviously advantageous when trying to induce the overexpression of a transgene whose product is prone to aggregation. Previous expression trials concluded that a 50ºC heat shock followed by incubation at 24ºC prior to IPTG induction produced the most soluble protein and therefore VLPs[123].

In order to harvest as many STNV VLPs as possible, an expression trial was performed where 1 mL of culture was taken every hour post induction. The

OD_{600} was measured and equal amounts of cells were loaded onto an SDS-PAGE, as described in Chapter 2.3.1.

The expression trials showed that STNV CP was not expressed before IPTG induction. After one hour, low levels of STNV CP expression are detected, increasing every hour until expression saturates after about 7 hours. The level of expression remained constant thereafter up until 24 hours (see Figure 3.1). Being as STNV particles are extremely stable[80], it was decided to harvest when the OD_{600} was highest (after ~24 hours). This ensured that STNV VLPs were purified from the most material present in the culture, before the bacterial cells reached the 'decline phase' of their growth.

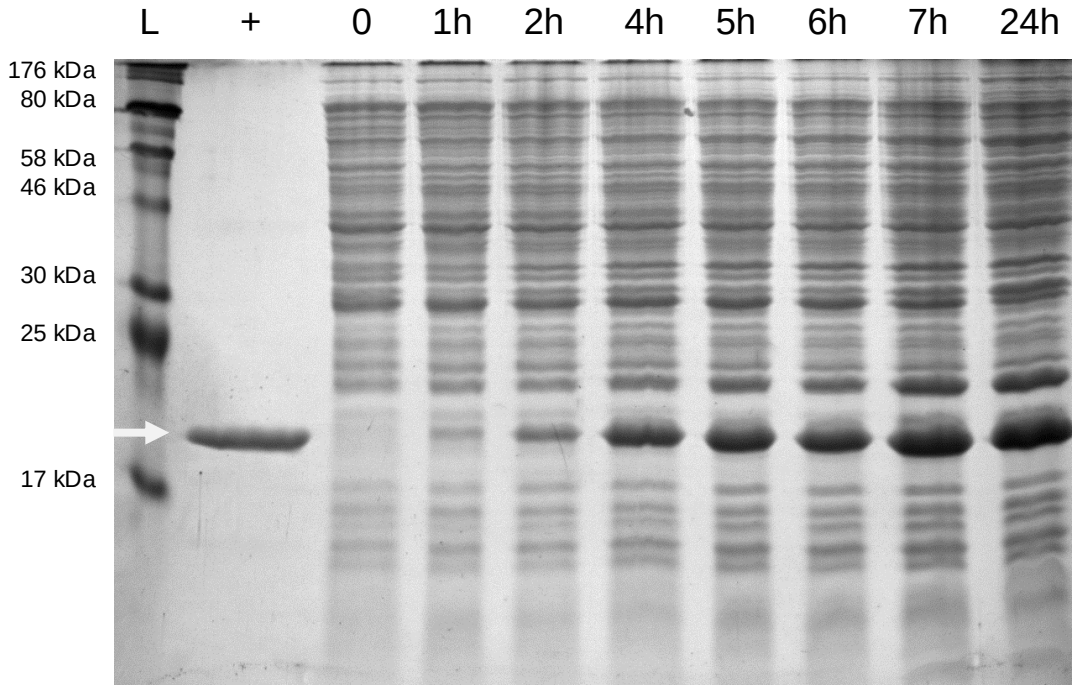


Figure 3.1: SDS-PAGE of STNV coat protein expression profiles over time
STNV CP expression was monitored by SDS-PAGE over time. These were immediately prior to induction, 0, after 1 h, 2 h, 4 h, 5 h, 6 h, 7 h and 24 h. Lanes labelled L = 7 – 176 kDa prestained ladder (NEB); + = STNV VLP control. The STNV CP levels increase following IPTG induction until saturation after approximately 7 h. No STNV CP is detected prior to IPTG induction. Arrow shows STNV CP.

The cells were pelleted by centrifugation at 4750 rpm and the pellets were frozen at -20°C. The pellets were thawed at room temperature before being extensively resuspended in 100 mL native buffer (50 mM HEPES 200 mM NaCl pH7.5) containing one tablet of complete protease inhibitor cocktail (Roche). This cellular suspension was treated with 1 mg of lysozyme (Sigma Aldrich) and left stirring for an hour at room temperature, before being sonicated and DNase I treated (Sigma Aldrich). The sample was centrifuged at 15,000 rpm (26892 g) using a Sorvall SS-34 rotor at 4ºC to remove cellular debris from the supernatant and the supernatant was ammonium sulphate precipitated (23% w/v) as described in Chapter 2.3.3. An aliquot was taken of both the supernatant and the cell pellet prior to ammonium sulphate precipitation and analysed by SDS-PAGE. Approximately 50 % of all STNV CP was obtained in the soluble fraction (see Figure 3.2). Aliquots were also taken after ammonium sulphate precipitation. The ammonium sulphate precipitated sample was centrifuged at 15,000 rpm (26892 g) using a Sorvall SS-34 rotor at 4ºC and an aliquot was taken of the supernatant for SDS-PAGE analysis. The pellet was resuspended in 50 mL native buffer and an aliquot was taken for SDS-PAGE. The SDS-PAGE showed that ammonium sulphate precipitation successfully precipitated all of the STNV CP from the supernatant and removed many proteins that did not precipitate (see Figure 3.2). Ammonium sulphate precipitation was therefore used as both a concentration step prior to further rounds of purification, and as an early purification step.

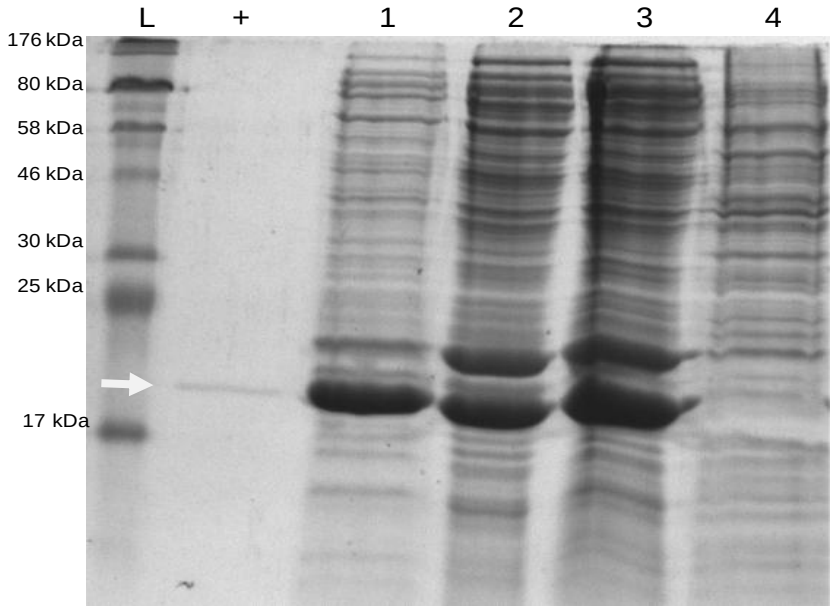


Figure 3.2: Monitoring STNV CP by SDS-PAGE
STNV VLPs are observed both in the supernatant and in the pellet in roughly equal amounts. STNV VLPs are efficiently ammonium sulphate precipitated away from other proteins in the supernatant. 1 = insoluble fraction/pellet from cell lysis; 2 = supernatant from cell lysis; 3 = resuspended ammonium sulphate precipitate; 4 = supernatant from ammonium sulphate precipitation; L = 7 – 176 kDa prestained ladder (NEB); + = STNV VLP control. Arrow shows STNV CP.

3.2.2 Strategies for STNV purification

Previously, recombinant STNV VLPs were purified from *E. coli* primarily for structural characterisation by X-ray crystallography[89]. The strategy for STNV VLP purification took advantage of both the difference in size of the VLPs and the difference in charge of the STNV CP relative to other *E. coli* proteins. The STNV VLP has a molecular weight of 1.5 MDa, so the VLP should elute in the void volume of most size exclusion chromatography columns. In addition, STNV CP has an isoelectric point (pI) = 10, therefore STNV CP shouldn't bind to a Q-Sepharose™ anion exchange resin, but should bind to SP Sepharose™ cation exchange resin at neutral pH. However, the VLP also contains encapsidated RNA. Thus in the VLP, the majority of the positively charged residues are making contacts with the phosphodiester backbone of the RNA[89]. In addition,

negatively-charged moieties on the STNV VLP are coordinated to 92 Ca^{2+} ions. These parameters would therefore affect the ability of the STNV VLPs to bind to SP resin, but the STNV VLPs shouldn't bind to Q-Sepharose™ resin. Thus the purification strategy was to purify STNV VLPs by Q-Sepharose™ anion exchange chromatography prior to size exclusion chromatography. STNV CP can then be purified from the STNV VLPs for capsid assembly experiments.

3.2.3 Q-Sepharose anion exchange chromatography

Immediately after resuspending the ammonium sulphate precipitate in native buffer, the STNV VLPs were dialysed extensively against native buffer to remove the ammonium sulphate. This sample was then passed down a Q-Sepharose anion exchange column in native buffer. This purification step was used primarily as a 'guard' column, where the majority of *E. coli* proteins bound to the column, whereas the STNV VLPs had little or no affinity (see Figure 3.3). The bound proteins were then eluted from the column by washing with 2 M sodium chloride (NaCl). The majority of *E. coli* proteins bound to the column, whereas ~50 % of all the STNV CP did not bind.

The fractions eluted from the Q-Sepharose anion exchange column which had 260 and 280 nm absorbance were analysed by 15 % SDS-PAGE. In general, one broad peak was eluted in the flow through (peak 1). A larger peak, which contained a shoulder, was then eluted upon washing with 2 M NaCl (peaks 2 and 3). SDS-PAGE analysis of the flow through (peak 1 in Figure 3.3) showed that STNV VLPs eluted predominantly in fractions from the front until the centre of the peak (lanes 1-4).

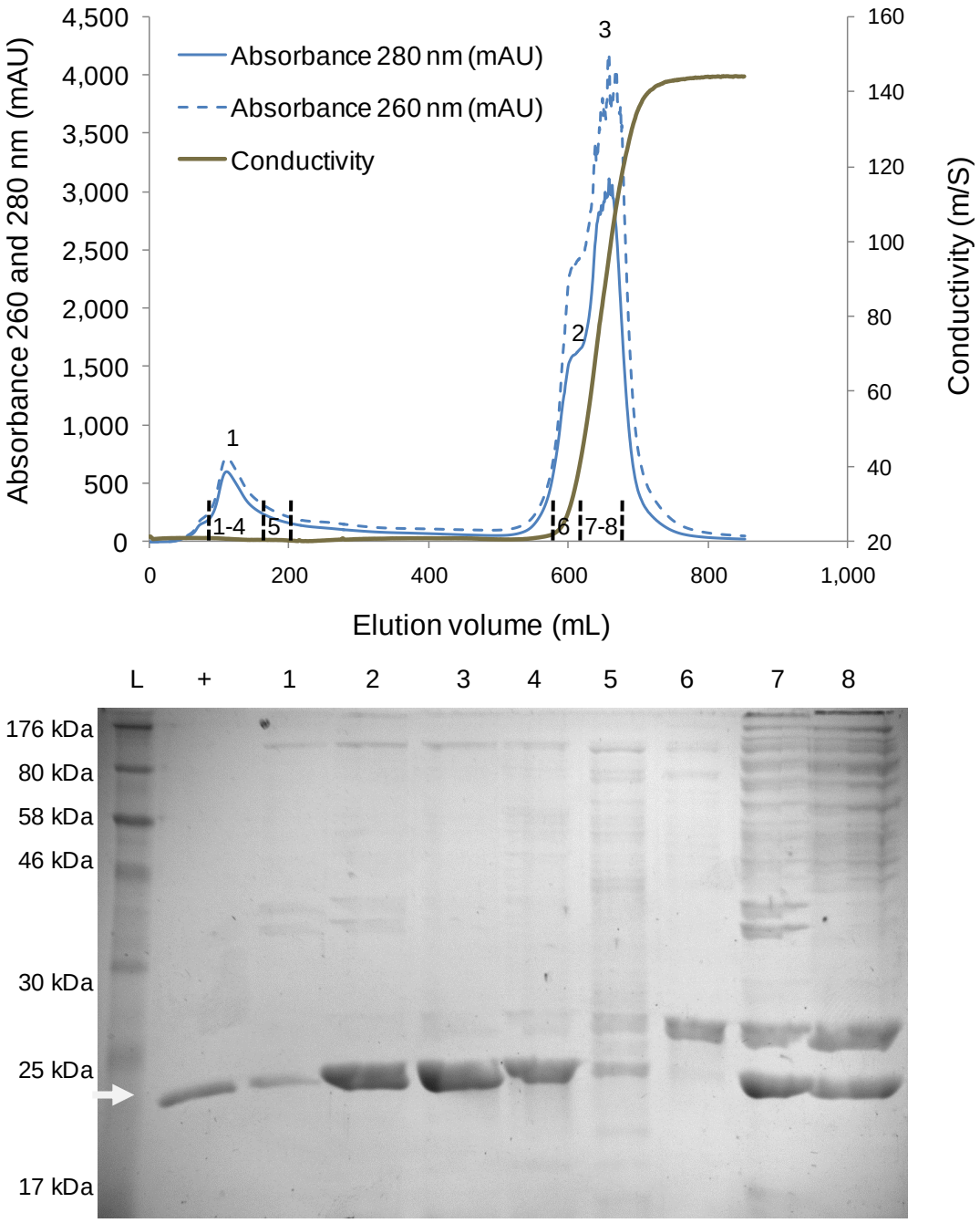


Figure 3.3: Q-Sepharose anion exchange and SDS-PAGE
The above chromatogram follows the A_{280} (solid blue line) and A_{260} (dashed blue line). The conductivity is monitored by the brown line. SDS-PAGE shows STNV VLPs are eluted in the flow through. Lanes on the gel indicated by fraction numbers, shown as vertical lines on the chromatogram. Lanes 1-5 = fractions from peak 1; Lane 6 = fraction from peak 2; Lanes 7-8 = fractions from peak 3; L = prestained 7-176 kDa ladder (NEB); + = STNV VLP control. STNV CP shown by white arrow.

The fractions analysed from the tail of peak 1 became more and more contaminated, until there was predominantly contaminating proteins (see lane 5, Figure 3.3). When the column was washed with 2 M NaCl, a shoulder from the

main peak (labelled 2, Figure 3.3) eluted a protein of ~25 kDa in size. This protein was not expressed prior to induction, but became more abundant the longer the cells were grown. This suggests that this protein, possibly phage shock protein, is a stress-related protein that *E. coli* expresses under certain conditions[124; 125], such as when exposed to increasing temperature. This protein was easily removed by Q-Sepharose™ anion exchange chromatography (see Figure 3.3).

Q-Sepharose anion exchange chromatography proved to be a very effective step in purification. A variety of differently sized Q-Sepharose™ columns with bed volumes of 5 mL, 20 mL, 120 mL and 330 mL were tried in this step of purification. The largest of these columns proved to be most effective at improving the purity of the sample. Smaller columns did not remove as many contaminating proteins as this larger column therefore the largest column was used. An additional advantage of using this large column was that more of the sample could be loaded onto the column at once. This ensured that all of the sample could be loaded in one injection, and that the sample was not excessively concentrated prior to purification.

Fractions were therefore collected from peak 1 (omitting fractions from the tail of the peak) and were pooled and ammonium sulphate precipitated (23 % w/v) as described in Chapter 2.3.3. Peaks 2 and 3 contained many contaminating proteins and nucleic acid, so these fractions were discarded.

3.2.4 Size exclusion chromatography

The ammonium sulphate precipitated sample was centrifuged at 15,000 rpm (26892 g) using a Sorvall SS-34 rotor at 4ºC to pellet. The pellet was resuspended in 5 mL STNV native buffer and extensively dialysed against native buffer to remove the ammonium sulphate. After dialysis, size exclusion chromatography was performed using a Superdex-200 26/60 column at a flow rate of 2 mL/min, where the UV absorbance was monitored at 260 and 280 nm and 3 mL fractions were collected.

The sample started eluting from this column after ~95 mL and stopped eluting after 140 mL (see Figure 3.4a). A shoulder from this peak eluted from ~95 mL to 100 mL, but no additional contaminating protein was detected by SDS-PAGE analysis. In addition, a smaller broad peak also eluted from ~175 mL to 200 mL, but was not analysed by SDS-PAGE due to the assumption that the STNV VLPs were eluted in the larger peak. The fractions from the shoulder of the peak were discarded in case of aggregation of the STNV VLPs and/or low level contamination not detected by SDS-PAGE. The fractions from the main peak were pooled and analysed by UV absorbance. The $A_{260/280}$ ratio was routinely calculated to be ~1.7, agreeing with values previously determined[89]. A small aliquot from the pooled fractions were diluted 10-fold to be visualised by TEM as described in Chapter 2.2.2, and confirmed to be STNV VLPs.

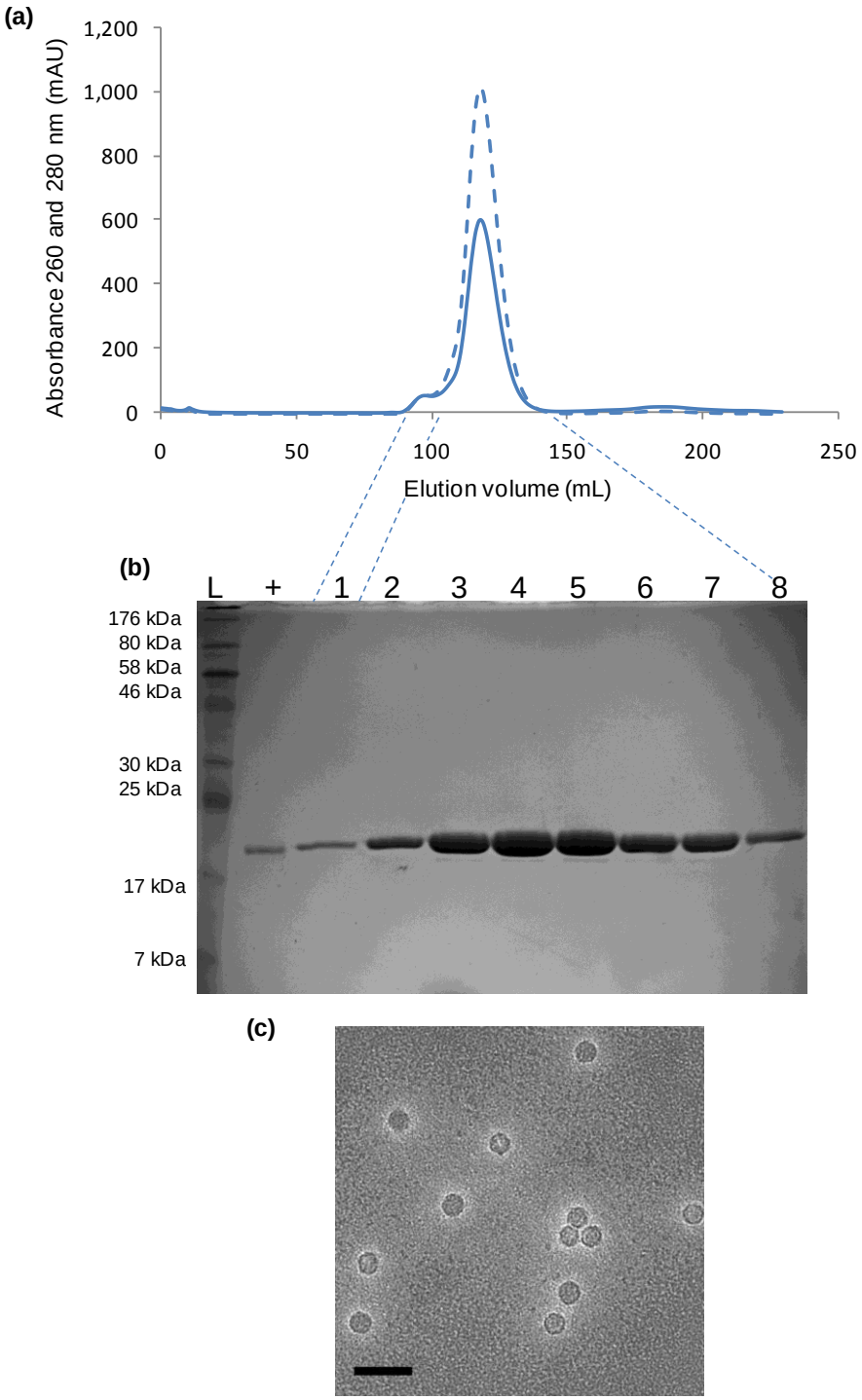


Figure 3.4: SEC purification and characterisation of STNV VLPs
(a) Size exclusion chromatogram of purified STNV VLPs
Chromatogram shown by absorbance at 280 nm (solid line) and 260 nm (dashed line).
(b) SDS-PAGE of fractions from chromatogram.
Fractions were collected across the peak between 95 mL and 135 mL. Lanes 1-8 = fractions eluted from the shoulder (1) and the large peak (2-8) following size exclusion chromatography. Dashed lines connect the fractions from the chromatogram with analysis by SDS-PAGE.
(c) EM image of purified STNV VLPs
Pooled fractions diluted 10-fold and visualised by electron microscopy. Images were taken 52,000 times magnified; scale bar = 50 nm.

3.2.5 *In vitro* disassembly and reassembly of STNV VLPs – proteolysis and effect on capsid assembly

In order to effectively disassemble and reassemble STNV VLPs *in vitro*, two methods were attempted by Dr Steven Lane that were known to disassemble two other ssRNA viruses, namely MS2 and TCV[89]. MS2 disassembles with the addition of two volumes of glacial acetic acid[126], while high salt (1 M) and pH (8.5) buffer in the presence of ethylenediaminetetraacetic acid (EDTA) (3 mM) successfully dissociates other plant viruses, such as Turnip Crinkle Virus (TCV)[34]. Only the latter method resulted in the elimination of *T*=1 capsids following dialysis, as shown by TEM. When the disassembled capsids were subjected to dialysis against reassembly buffer (containing 3 mM $CaCl_2$), *T*=1 capsids re-emerged as shown by TEM[123]. This showed for the first time that STNV VLPs could be disassembled and reassembled *in vitro*.

However, when STNV VLPs are disassembled, the coat protein becomes very proteolytically sensitive, and appears to be at the mercy of trace proteases[28]. When I attempted to perform these disassembly and reassembly experiments using Dr Lane's protocol, the STNV CP became extensively proteolysed. Subsequent attempts to reassemble STNV CP where proteolysis has run to completion results in failure (data not shown).

Several methods were employed in an attempt to prevent this proteolysis and the subsequent aberration of capsid assembly. The first of these was the inclusion of complete protease inhibitor tablets (Roche) in all buffers when dealing with the *in vitro* disassembly and reassembly of STNV VLPs. Although inefficient capsid assembly was observed, there was still considerable proteolysis of STNV CP (data not shown).

The inclusion of complete protease inhibitors in all the buffers was not enough to prevent proteolysis, so more drastic action was required to tackle this problem. In addition to having protease inhibitors in the buffers, each sample was incubated with α_2-Macroglobulin (α_2M) (Roche) for 30 mins prior to being exposed to disassembly conditions. The N-terminal arm of TCV is also incredibly proteolytically sensitive, but the addition of α_2M helps to prevent proteolysis from occurring[34]. With this rationale, disassembly and reassembly experiments were performed in the presence of α_2M.

α_2M is a 800 kDa homotetrameric plasma protein that circulates around the bloodstream of vertebrates[127]. It functions as a protease inhibitor *in vivo*, and is able to inactivate a huge variety of proteases including members from all the known families of proteases[128]. The mechanism of protease inhibition is due to a unique mechanism: the protease binds to α_2M at a stretch of amino acids termed the 'bait' region[127; 128]. This 'bait' region contains a number of peptide bonds recognised for cleavage by the various proteases. Subsequent cleavage of the bait region of α_2M induces a massive conformational change in α_2M, which results in entrapment of the protease, and renders the protease inactive[127; 128].

In order to characterise whether α_2M prevents proteolysis and mediates efficient capsid assembly, 0.5 mL STNV VLPs (2 mg/mL) were dialysed in disassembly buffer containing 1 unit of α_2M in two separate dialysis reactions. In addition, 0.5 mL STNV VLPs (2 mg/mL) containing 1 unit of α_2M was analysed by size exclusion chromatography, where the fractions were collected for SDS-PAGE.

Prior to chromatography, an aliquot was taken for TEM to confirm the presence of VLPs.

After 6 hours of dialysis against disassembly buffer, one of the samples was removed and dialysed against reassembly buffer for 24 h at 4ºC. An aliquot from the disassembled sample was also taken for TEM to confirm whether disassembly was successful, before being analysed by size exclusion chromatography. The reassembled sample was also subjected to analysis by TEM, size exclusion chromatography and SDS-PAGE. Figure 3.5 shows an overlay from each size exclusion chromatogram relating to STNV exposed to native, disassembly and disassembly/reassembly conditions (top). In addition, how the chromatogram relates to fractions on an SDS-PAGE and the corresponding TEM (bottom).

The size exclusion profile from the native STNV sample showed two major peaks which eluted between 7-12mL and a minor peak which eluted between 13-14mL. The SDS-PAGE from these fractions identified α_2M as the major species on the gel from the first peak, whereas STNV was the major species in the second peak. The third minor peak contains BSA, which is also present in the α_2M sample (Roche), presumably to prevent undesired aggregation. The reason why α_2M appears in both these major peaks is not known, but is probably due to aggregation. In contrast, it is known that binding proteases into the bait region of α_2M induces a large conformational change and that α_2M also binds other proteins such as apoE, with complexes detected as being as large as 18.5 nm[129]. This phenomenon may also contribute to this profile.

The elution profile from the disassembled STNV shows that eluted STNV CP shifts from between 9-11mL to approximately 15-17mL, confirmed by 15% SDS-PAGE. This had the same shift in the elution profile observed in prior disassembly and reassembly experiments without α_2M, but without the same proteolysis of the N-terminus (data not shown). The absence of STNV VLPs in the TEM of the disassembled sample provided further evidence that STNV VLPs were successfully disassembled (see Figure 3.5).

The chromatogram obtained from the reassembled VLPs showed that the main peak returns to that which corresponds to native STNV capsid. This is further confirmed by analysis of reassembled STNV by TEM, and by the corresponding SDS-PAGE analysis from the size exclusion fractions. Although reassembly has been successfully initiated, it appears from both the chromatogram and SDS-PAGE analysis that some STNV CP becomes aggregated. This could be due to the α_2M. It is known that α_2M has a very broad specificity for its protein targets and that its protein targets are not always proteases[129]. Under conditions that favour STNV assembly, it may be that some aggregation takes place with α_2M.

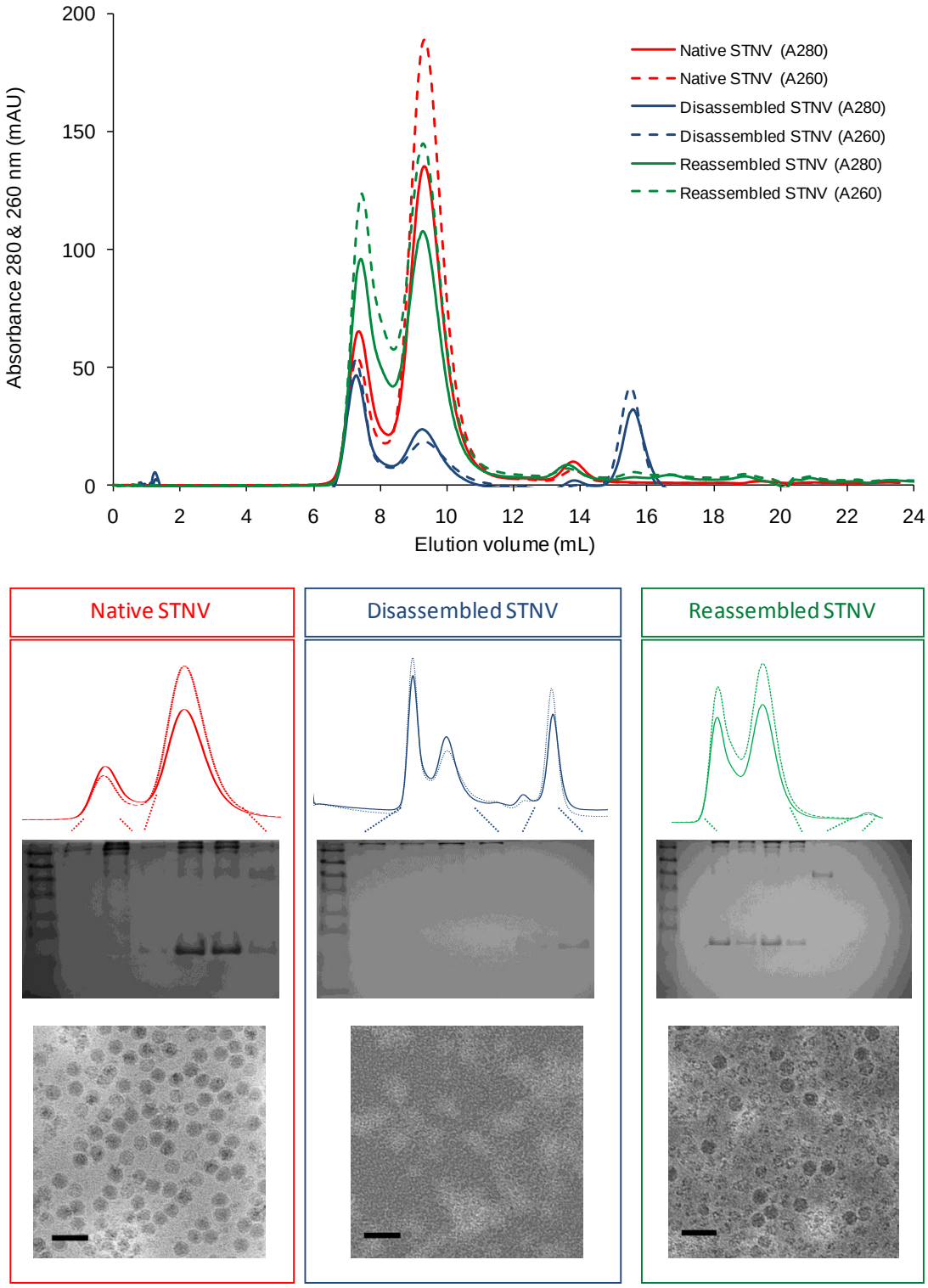


Figure 3.5: STNV disassembly/reassembly in the presence of α_2M.
Native, disassembled and reassembled samples of STNV (0.5 mg) in the presence of α_2M (1 U) were analysed by size exclusion chromatography (SEC), where the fractions that absorbed at 260 and 280 nm were analysed by SDS-PAGE. An aliquot was taken prior to SEC for TEM. Top panel shows an overlay of the SEC chromatogram from each sample. Bottom panel shows each individual chromatogram above the SDS-PAGE connected by dotted lines to the relevant fractions, and the image captured by TEM is shown at the bottom of the panel. Scale bar = 50 nm.

3.2.6 STNV CP purification with α_2M

To investigate how RNA influences STNV capsid assembly, the STNV CP must be purified in a manner which leaves the CP i) free from proteolysis; ii) assembly competent; and iii) free from RNA packaged in *E. coli*. Previous attempts to purify STNV CP by size exclusion chromatography failed on all three counts[28], thus a new approach was adopted using ion exchange chromatography. The premise for this was simple; to disassemble STNV VLPs in a manner that preserves full-length STNV CP and to separate the RNA from the STNV CP based on their net charge. STNV CP has an isoelectric point (pI) of 10, presumably due in part to the highly basic N-terminal region[89], whereas RNA contains a large net negative charge relative to the size of the RNA molecule.

There was one main caveat to using ion exchange chromatography with disassembled STNV. Previous conditions for STNV disassembly involved dialysis into a buffer containing elevated pH (8.5), EDTA (3 mM) and high NaCl concentration (1 M)[28; 123]. Serendipitously, it was observed that STNV VLPs disassembled without increasing the NaCl concentration. It was therefore possible to carry out ion exchange chromatography on disassembled STNV using disassembly buffer minus NaCl.

STNV VLPs were disassembled by dialysis into disassembly buffer without NaCl and then subjected to Q-Sepharose anion exchange chromatography. It was presumed that the Q-Sepharose resin would be sufficient to out-compete the STNV CP and bind all the RNA, and that the STNV CP would elute in the flow through from the column. The results from these initial experiments showed

that although the majority of the RNA was successfully removed, there remained fragments of RNA that was bound to the coat protein (data not shown). This shows that STNV CP binds tightly to RNA, and that Q-Sepharose™ anion exchange chromatography on its own was not sufficient to purify STNV CP.

To overcome this limitation, an SP Sepharose™ cation exchange column was used to bind STNV CP following Q-Sepharose™ anion exchange chromatography. In this way, the Q-Sepharose resin binds RNA, whereas the SP resin binds the STNV CP.

STNV VLPs (5 mg) were disassembled by dialysis in disassembly buffer without NaCl in the presence of 3 U α_2M (Roche). The sample was then loaded onto a Q-Sepharose anion exchange and SP cation exchange columns in series. Post injection, the flow through from these columns eluted after about 2 mL, where the absorbance at 260 and 280 nm were monitored. The 280 nm signal was initially higher in the sample, where three fractions were collected between 2 and 8 mL, followed by a larger 260 nm peak, presumably due to displaced RNA. When the UV absorbance returned to baseline (after 10 mL), the Q-Sepharose column was removed from the SP column. The SP column was washed for 20 column volumes before a NaCl gradient was applied. The NaCl gradient induced an increase in conductivity which eluted whatever was bound to the column. Three fractions from the two peaks were taken, and analysed by SDS-PAGE (see Figure 3.6).

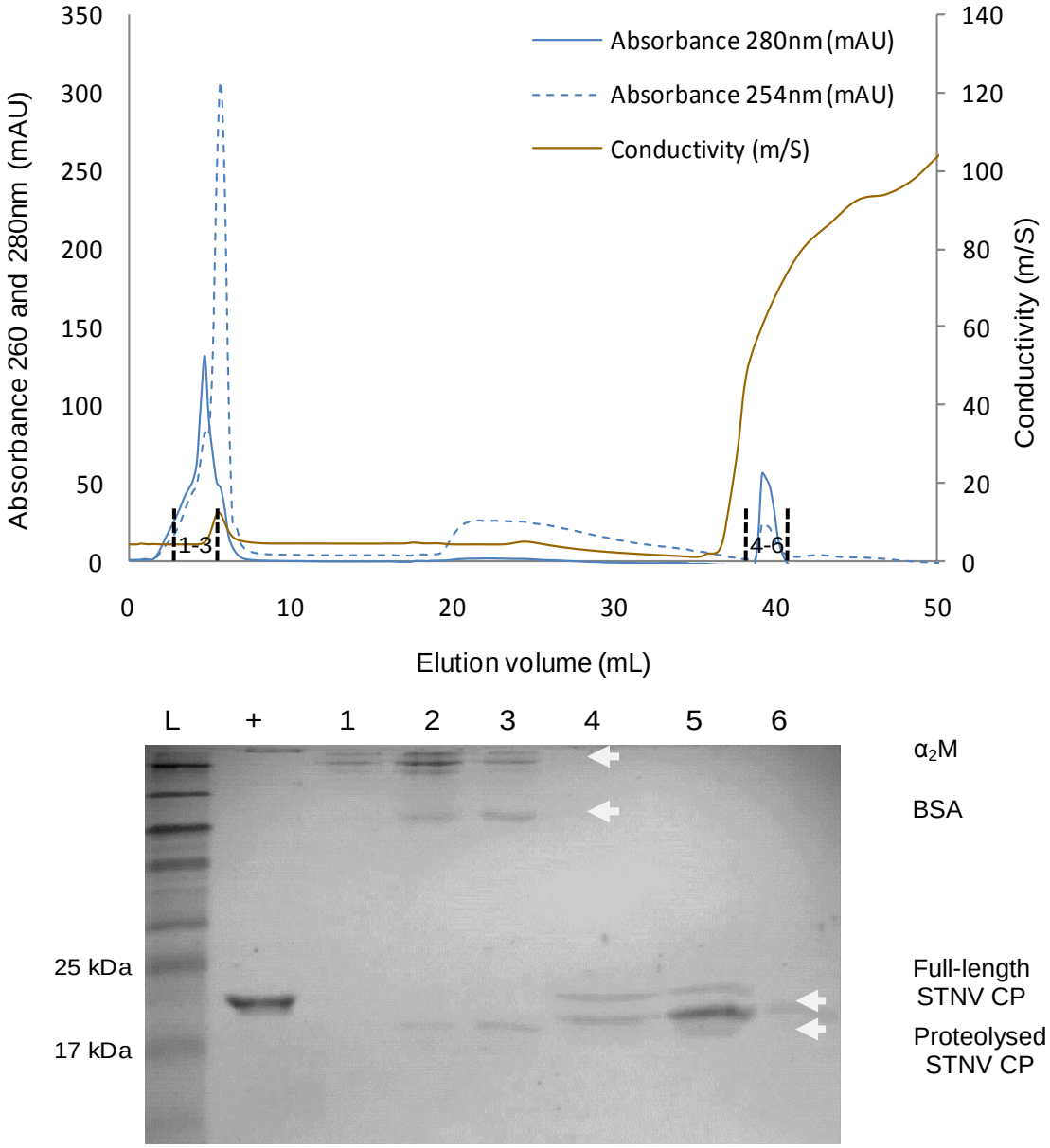


Figure 3.6: STNV CP purification and analysis in the presence of α_2M
(a) Chromatogram of STNV CP purification using Q-Sepharose anion exchange and SP cation exchange chromatography in series. The absorbance was monitored at 254 nm (blue dashed line) and 280 nm (blue unbroken line). Conductivity monitored on second axis (brown).
(b) SDS-PAGE showing proteolysed STNV CP. Three fractions from two peaks were collected and analysed by SDS-PAGE. The proteins that eluted in the first peak were collected in fractions from the flow through (i.e. did not bind to either Q-sepharose or SP resin) were α_2M and BSA, and faint bands corresponding to proteolysed STNV CP. The second peak was eluted with a NaCl gradient from the SP column, and contained a mixture of full length STNV CP and proteolysed STNV CP. Each component is clearly labelled by an arrow on the gel.

Fractions analysed from the first peak showed that α_2M and BSA was present in flow through, as well as proteolysed material from STNV CP. The fractions from the second peak showed that STNV CP was eluted from the SP column. Two species were present on the gel, corresponding to full length STNV CP and proteolysed STNV CP. The $A_{260/280}$ ratio of this material was ~0.6, suggesting that all of the RNA had been successfully removed. However, proteolysis remained a problem, even in the presence of α_2M. This suggests that trace proteases in the sample need to be removed prior to STNV CP purification.

3.2.7 B3-mediated capsid assembly using proteolysed STNV CP

The fractions of STNV CP from Figure 3.6 were pooled and the concentration was determined by UV absorbance, using the Beer-Lambert law (STNV-1 CP ε = 16,960 M^{-1} cm^{-1}). STNV CP (0.5 mL, 0.5 mg/mL) was dialysed in reassembly buffer overnight at 4ºC in the presence (~80 µg/mL) and absence of aptamer B3[81]. Aptamer B3 (30-mer) was synthesized by Amy Barker, using solid phase synthesis as described by Murray and Prokop [130]. The sequence is given in Table 1.2 and Chapter 2.4.4, respectively.

Following dialysis, size exclusion chromatography was performed before fractions were analysed by SDS-PAGE, and fractions between 8-12 mL analysed by TEM. It was hypothesized that only full length STNV CP would support capsid assembly, based on previous Δ15 STNV CP truncation experiments, and that aptamer B3 was selected in the presence of immobilised STNV CP[81]. Previously, *in vitro* capsid assembly experiments with B3 failed to produce VLPs[28], although STNV CP was digested prior to experimentation.

The chromatogram obtained from STNV CP in the absence of aptamer B3 eluted between ~17-20 mL, suggesting that capsid assembly did not take place. Fractions were taken between 10-12 mL and 18-20 mL and concentrated using Amicon Ultra® spin concentrators and the fractions were analysed by SDS-PAGE. Both proteolysed STNV CP was present in the 18-20 mL fractions, but no evidence was obtained for any STNV CP in the 10-12 mL fractions. An aliquot taken prior to size exclusion chromatography was also analysed by TEM, but no capsids were observed.

In contrast, the size exclusion elution profile of STNV CP in the presence of aptamer B3 shows three peaks; a peak that elutes between 10-12 mL, one that elutes between 18-20 mL and one that elutes between 20-22 mL. Fractions were taken from peaks 1 and 2 and concentrated using Amicon Ultra® spin concentrators. Presumably, the peak that elutes between 22-24 mL is aptamer B3, based on the size of the aptamer and the $A_{260:280}$ ratio. The concentrated fractions were analysed by SDS-PAGE, where it was observed that only full length STNV CP was present in the first peak, whereas proteolysed material was present in the second. TEMs prepared from peak one confirm that the shift in this peak is the result of capsid assembly (see Figure 3.7).

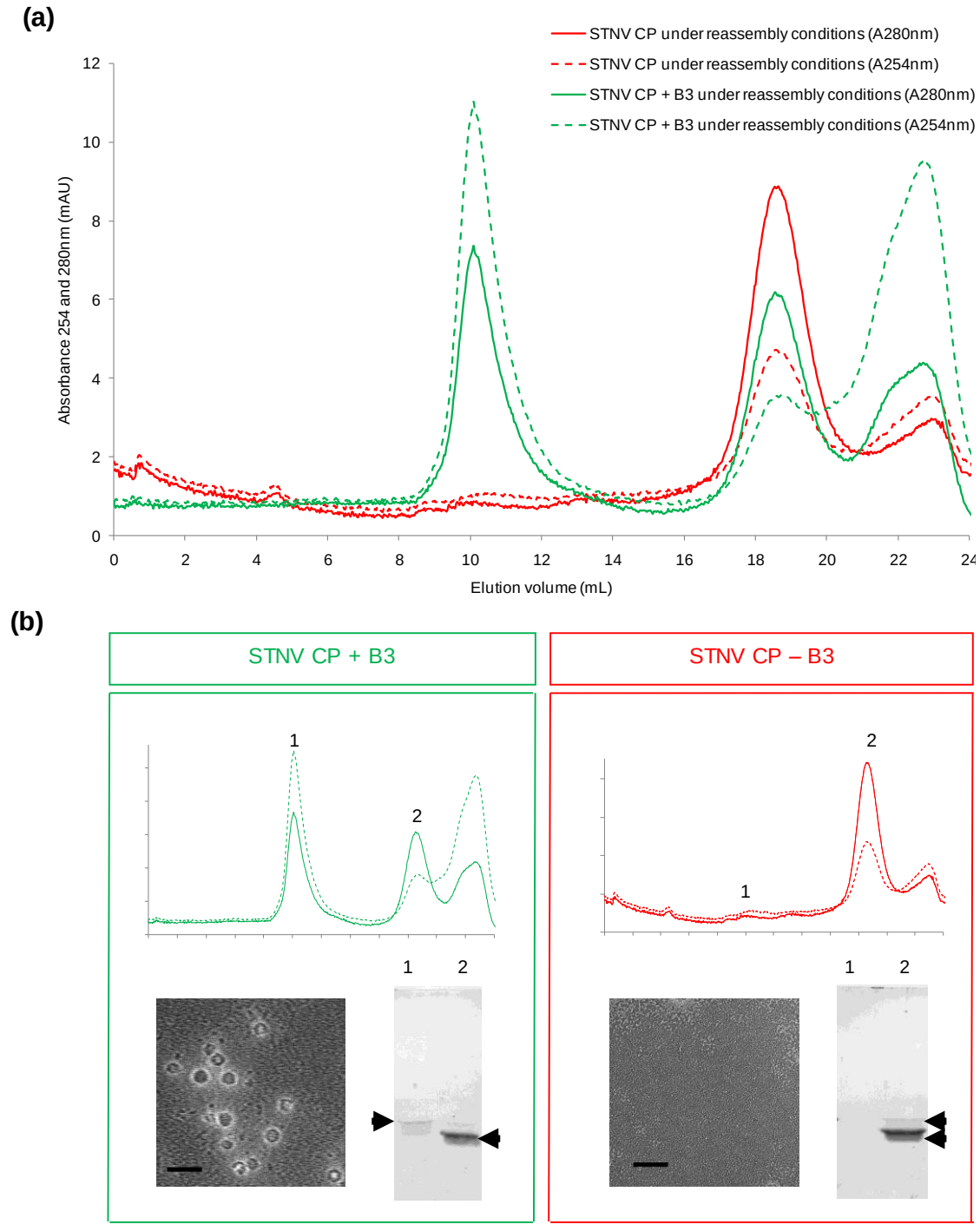


Figure 3.7: Proteolysed STNV CP reassembly in the presence of B3
(a): Overlay of size exclusion chromatograms looking at STNV reassembly with and without B3. The chromatogram in green is STNV CP reassembled with aptamer B3. The chromatogram in red is STNV CP exposed to reassembly conditions without RNA. The solid lines represent absorbance at 280 nm whereas the broken lines represent absorbance at 254 nm.
(b): Characterisation of capsid assembly. Each chromatogram is shown separately, with 2 fractions taken (indicated as numbers). These fractions were analysed by SDS-PAGE and by TEM. Upper arrow shows full length STNV CP and lower arrow shows proteolysed STNV CP. Scale bar = 50 nm

3.2.8 Sucrose density gradient fractionation

In order to remove α_2M from STNV VLPs, it was imperative to find a new solution to prevent proteolysis. Being as trace proteases are still present in the purified VLPs, a new approach was taken to remove the proteases from the sample. Instead of size exclusion chromatography being used as a clean-up step after Q-Sepharose chromatography, sucrose density gradient fractionation was performed for the final purification step.

After Q-Sepharose anion exchange chromatography, fractions from peak 1 were pooled and ammonium sulphate precipitated (see Figure 3.3). The ammonium sulphate precipitated sample was then centrifuged at 26,892 g, and the pellet was resuspended in 12 mL native buffer. The sample was then extensively dialysed against native buffer to remove the ammonium sulphate.

In total, 6 sucrose density gradients (15-45% (w/v)) were prepared in 38 mL tubes (Seton), where 2 mL of sample was loaded on top of each sucrose density gradient. The gradients were centrifuged using a SW32ti rotor at 18,500 rpm for 21 h at 4ºC and allowed to stop without braking to avoid disruption to the gradients. The gradients after centrifugation contained a band near the middle of the tubes with a 'cloudy' appearance (see Figure 3.8d, white arrow shows cloudy band). The gradients were fractionated and the absorbance at 260 and 280 nm was measured (see Figure 3.8a). The fractionation profile showed 2 peaks which absorbed at both 260 and 280 nm. These peaks were then analysed by SDS-PAGE, where the 1st peak contained contaminating proteins and the 2nd peak contained STNV VLPs. The 2nd peaks from all the gradients were pooled and extensively dialysed against native buffer to remove

the sucrose. An aliquot was taken and diluted 10-fold before being analysed by TEM, which confirmed the presence of STNV VLPs (see Figure 3.8c). The yield according to UV absorbance and level of purification estimated by SDS-PAGE analysis suggests that sucrose density gradient fractionation was at least equivalent to size exclusion chromatography (see Figure 3.8).

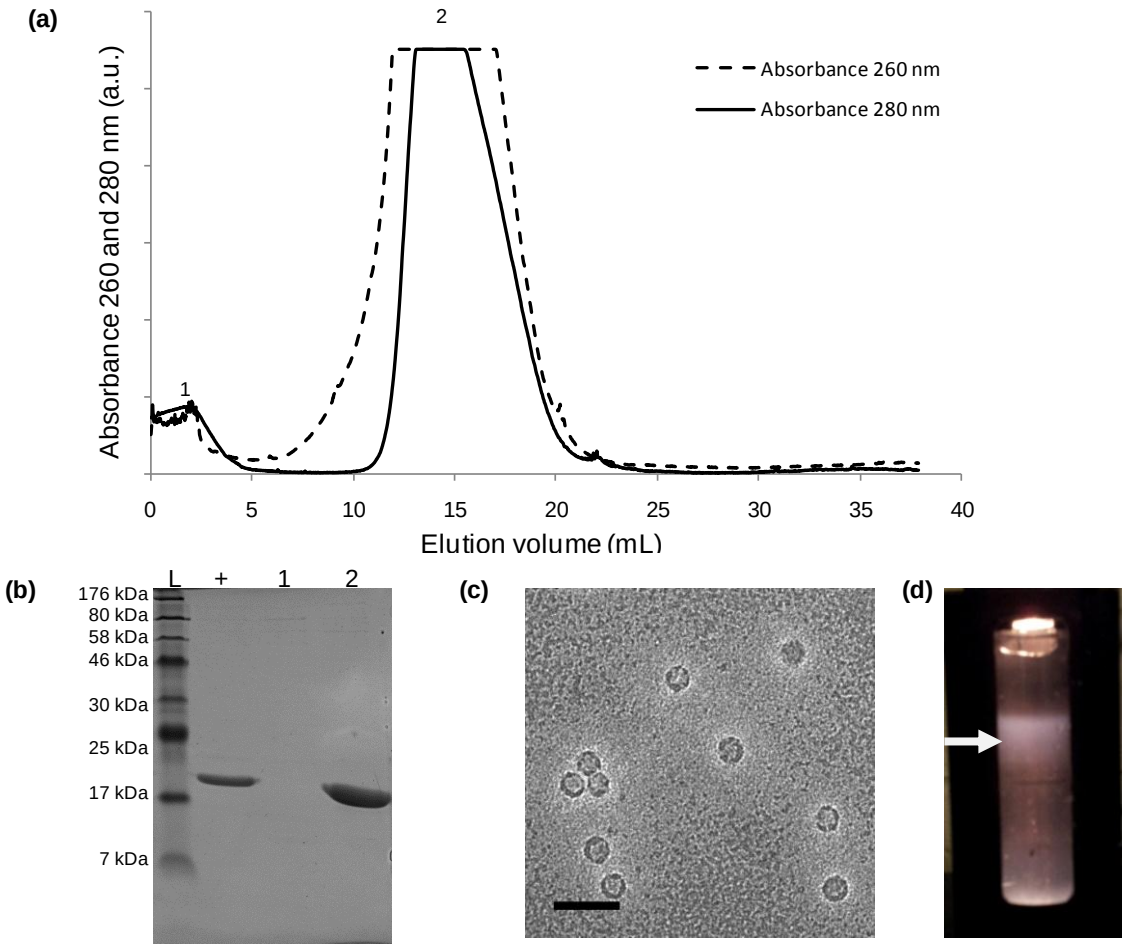


Figure 3.8: Sucrose density gradient fractionation of STNV VLPs
(a) Sucrose density gradient fractionation profile of STNV VLP purification.
The solid line monitors the A_{280} and the broken line monitors A_{260}. The peaks labelled 1 and 2 were analysed by SDS-PAGE and peak 2 by TEM. Flattening of the peak indicates going off scale.
(b) SDS-PAGE analysis of peaks 1 and 2.
Contaminating proteins are eluted in peak 1 whereas STNV VLPs are eluted in peak 2, indicating that STNV VLPs were efficiently purified.
(c) Electron micrograph of STNV VLPs from peak 2.
STNV VLPs from peak 2 were diluted 10-fold and analysed by negative stain TEM.
(d) Photograph of sucrose density gradient.
The gradient holder (black) is cropped to show the sucrose density gradient. The white cloudy band in the middle of the gradient is STNV VLPs.

3.2.9 One-pot disassembly/reassembly of STNV VLPs after sucrose density gradient fractionation

To assess whether sucrose density gradient fractionation could remove trace proteases from STNV VLPs, another disassembly/ reassembly was performed with Amy Barker. STNV VLPs (0.5 mL, 1 mg/mL) using sedimentation velocity experiments in the absence of α_2M. Aliquots were taken for SDS-PAGE analysis and TEM at every stage prior to being analysed by sedimentation velocity.

The STNV VLP has a sedimentation coefficient ($S_{20,w}$) ~42 S, comparable to the native STNV-1 virion (38-45 S)[80]. When the material is disassembled, several heterogeneous peaks emerged, ranging from ~1.5 S to 30 S. Amongst these heterogeneous peaks, a sharp peak was present at 1.5 S, presumably owing to the STNV CP (see Figure 3.9). Several peaks are present that correspond to the RNA. These peaks are likely to be different sizes and/or conformers of the RNA packaged in *E. coli*. Sedimentation velocity of STNV CP mRNA transcript also produces heterogeneous peaks beginning at ~10 S (see Figure 4.5), suggesting that a subset of this material is the mRNA.

The reassembled material contains no heterogeneous peaks and sediments at ~42 S, suggesting that the STNV VLPs have reassembled. The SDS-PAGE from each sample showed that no proteolysis occurred (see Figure 3.9c), therefore the efficiency of reassembly is very high (>95 %) when proteases are not present. TEMs from each sample showed that STNV VLPs are removed when exposed to disassembly conditions, whereas the capsids re-emerge under reassembly conditions (Figure 3.9b).

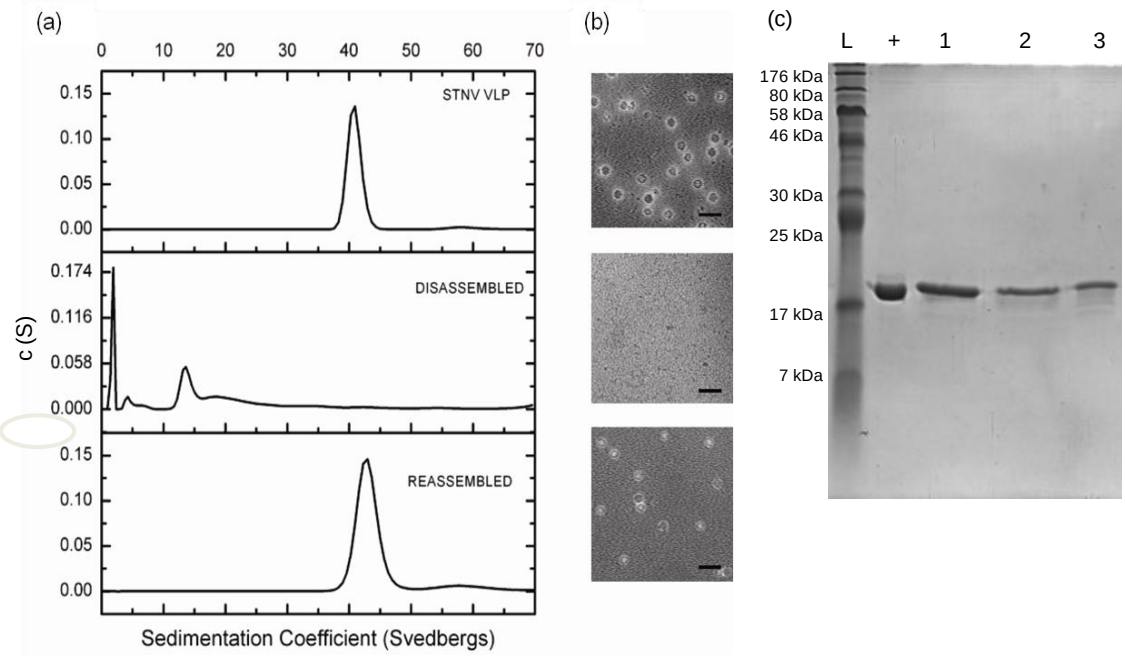


Figure 3.9: *In vitro* disassembly and reassembly of STNV VLPs
(a) Normalised c (S) versus (S) plots of, from top to bottom, STNV VLP, disassembled STNV and reassembled STNV VLP.
(b) TEM micrographs corresponding to the samples shown on the left.
(c) SDS-PAGE from each sample. L = 7-176 kDa prestained ladder (NEB); + = STNV VLP control; 1 = STNV VLP (top); 2 = Disassembled STNV (middle); 3 = Reassembled STNV (bottom)

3.2.10 STNV CP purification after sucrose density gradient fractionation

Previously, it was clear that by including a sucrose density gradient fractionation step into STNV VLP purification, removal of proteases could be achieved (see Figure 3.10). Although dialysis into disassembly buffer without NaCl proved successful at disassembling STNV VLPs, it was not as efficient as with 1 M NaCl. Directly adding 10 X disassembly buffer without NaCl to a 4/5th volume of the STNV VLPs resulted in efficient disassembly. In addition, this method proved to be quicker, thus minimising the amount of time STNV CP was exposed to any potential trace proteases.

STNV VLPs (12 mL, 1 mg/mL) purified by sucrose density gradient fractionation were disassembled by adding 1/5th volume of disassembly buffer. The sample was then immediately loaded onto Q-Sepharose anion exchange and SP cation exchange columns in series. After 30 mL, the Q-Sepharose anion exchange column was removed and the SP column was washed for 30 column volumes. Following washing, the STNV CP was eluted by a NaCl gradient and the fractions were analysed by SDS-PAGE (see Figure 3.10). The SDS-PAGE analysis showed that STNV CP was not proteolysed and the $A_{260:280}$ ratio of ~0.6 indicated that the RNA was successfully removed.

In order to characterise the STNV CP in both disassembly buffer and reassembly buffer, sedimentation velocity experiments were performed under these conditions with Amy Barker. The results from sedimentation velocity under disassembly conditions showed that STNV CP has a major component which has a $S_{20,w}$ =2.037 S (estimated M_r 18,033 Da), with minor species observed both at ~4 S and ~6 S, corresponding to monomer, trimer and pentamer, respectively. Sedimentation velocity experiments performed with STNV CP under reassembly conditions produced identical results to that of STNV CP under disassembly conditions, suggesting that STNV VLPs cannot assemble in the absence of RNA (Appendix, Figure 9.1). Equilibrium sedimentation of these samples under these conditions showed that there was no appreciable self aggregation (data not shown). The mass of STNV CP was determined using mass spectrometry, and is calculated at 21,584.17±0.67 Da, which is consistent with what is expected of STNV-1 CP minus the N-terminal methionine (21,583.52 Da) (Appendix, Figure 9.2).

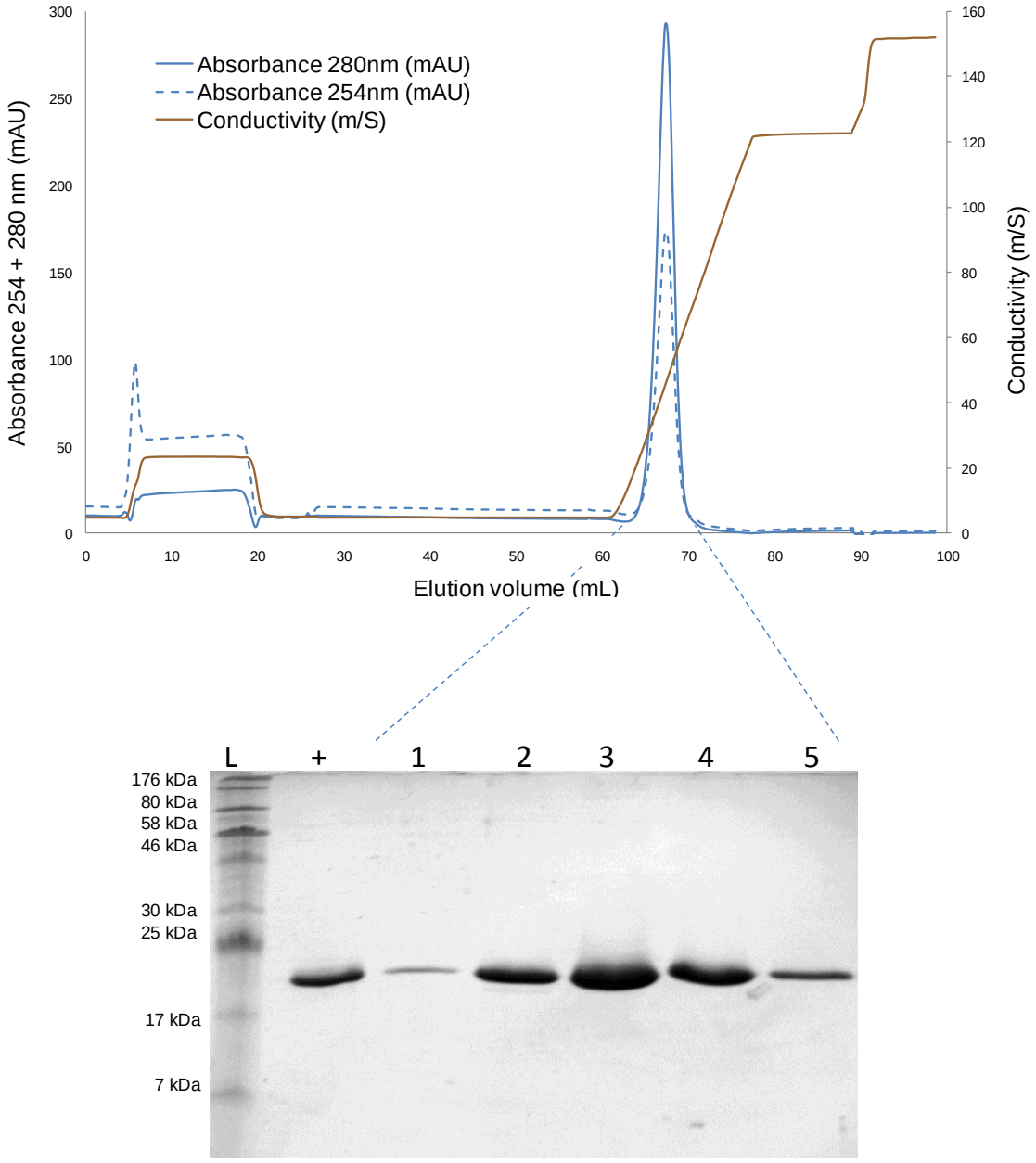


Figure 3.10: Purification of full length STNV coat protein
STNV VLPs were disassembled and purified in parallel by Q-Sepharose/SP ion exchange chromatography. STNV RNA binds to the Q-Sepharose anion exchange resin, whereas the coat protein binds to the SP resin. The anion exchange column is removed and the coat protein is eluted by a gradient of NaCl. Solid blue line represents A_{280} whereas the blue dashed lines represent A_{254}. The conductivity is shown in brown. The fractions from the peak were analysed by 15 % SDS-PAGE, where dashed blue lines shows which fractions from the chromatogram were analysed. L = 7-176 kDa prestained ladder (NEB); + = STNV VLP control; 1 – 5 = Fractions from NaCl eluted STNV CP

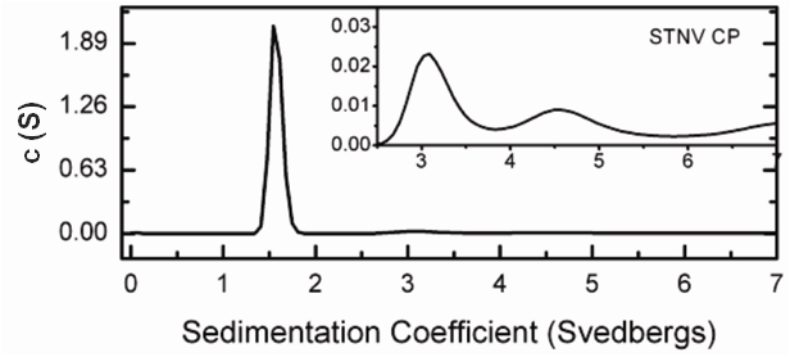


Figure 3.11: Sedimentation velocity of STNV CP
STNV-1 CP was purified by ion exchange chromatography and analysed by AUC. The inset shows a rescaled c (S) axis from 2.5-7 S highlighting the minor species present in the sample.

Figure 3.11 indicates that 2/3 objectives were achieved when the goals of STNV CP purification were set out. In order to test the third criteria of successful STNV CP purification, namely capsid assembly competency, reassembly experiments were carried out using purified STNV CP in the presence and absence of aptamer B3.

3.2.11 B3-mediated capsid assembly with full length STNV CP

In order to determine whether the purified STNV CP remains capsid assembly competent, capsid assembly experiments were performed in the presence and absence of aptamer B3. Although it was observed by AUC that STNV CP does not form capsids in the absence of RNA, this was included as a negative control.

0.5 mL STNV CP (0.5 mg/mL) was subjected to dialysis against reassembly buffer in the presence and absence of aptamer B3 (100 µg/mL). The dialysed samples were then analysed by size exclusion chromatography using a pre-equilibrated 10/300 GL S-200 column. Each sample was loaded onto the column at a flow rate of 0.5 mL/min and the absorbance was monitored at 260

and 280 nm. The fractions with UV absorbance were analysed by SDS-PAGE and TEM.

STNV CP that had been dialysed against reassembly buffer in the presence of aptamer B3 mainly eluted in one peak between ~8-12 mL. Minor peaks are also present between 18-20 mL and 22-24 mL, presumably corresponding to STNV CP monomer and aptamer B3, respectively. TEMs of the peak which eluted between 8-12 mL showed the presence of capsids and species which look like incomplete capsids (see Figure 3.12).

Fractions from peaks 1 and 2 were concentrated with Amicon Ultra® centrifugal concentrators and analysed by SDS-PAGE. STNV CP was present and unproteolysed in peak 1, whereas no STNV CP was detected in peak 2, as is expected at this very low UV absorbance. The peak between 22-24 mL is where aptamer B3 eluted previously (see Figure 3.7).

STNV CP dialysed in the absence of RNA was also loaded on to a pre-equilibrated 10/300 GL S-200 size exclusion column at a flow rate of 0.5 mL/min. As expected, the STNV CP eluted between ~16-20 mL, with no UV absorbance detected where capsids elute. Analysis by SDS-PAGE shows that no STNV CP is detected between 8-12 mL, suggesting that STNV CP doesn't spontaneously assemble into capsids in the absence of aptamer B3, thus confirming what was observed with sedimentation velocity analysis. SDS-PAGE also confirms that no proteolysis was observed in either sample and no capsids were observed by TEM.

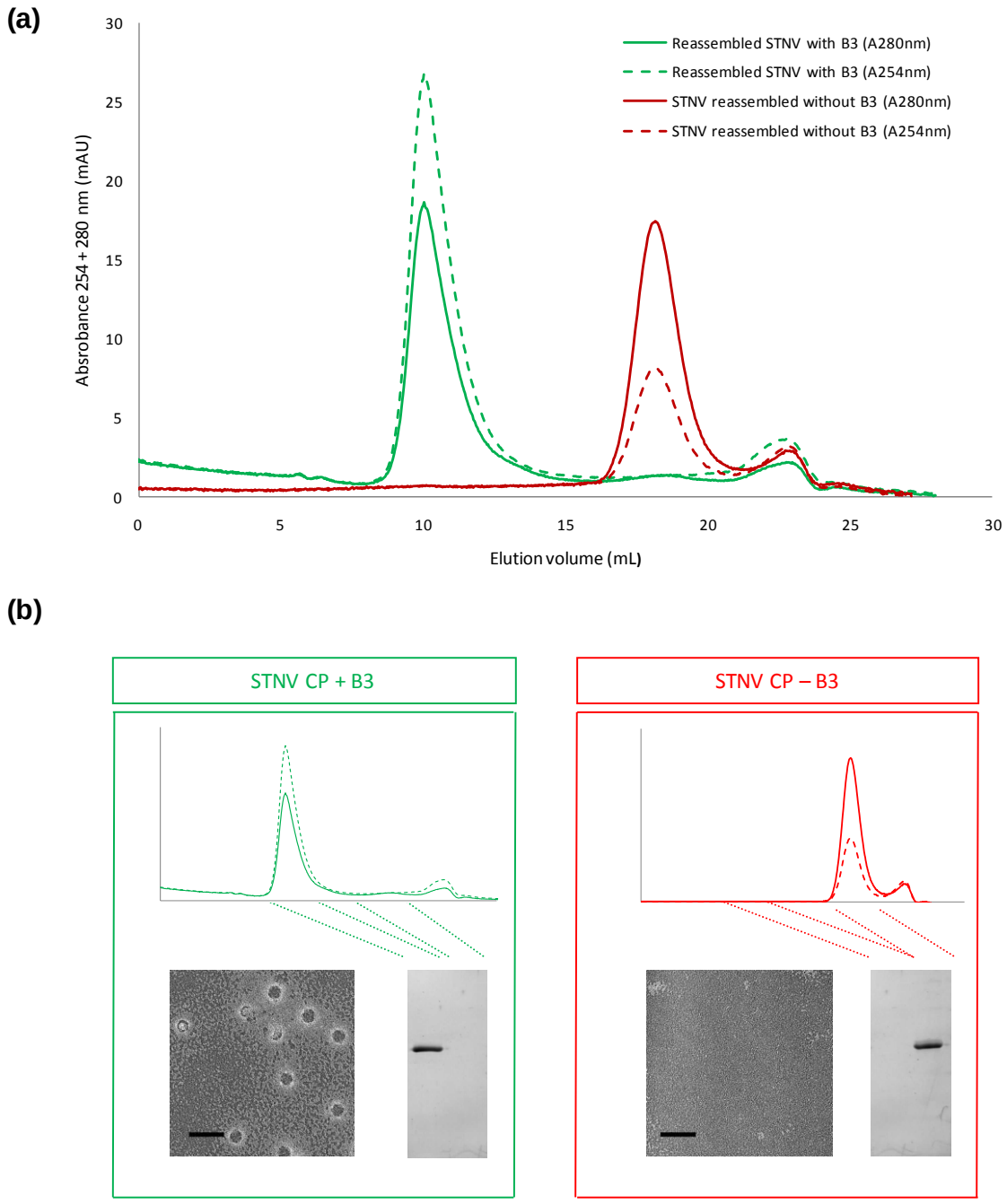


Figure 3.12: B3-mediated capsid assembly
(a) Overlay of size exclusion chromatograms
Capsid assembly studies were carried out using STNV CP with and without aptamer B3. Solid lines represent 280 nm whereas dashed lines represent 254 nm.
(b) Characterisation of capsid assembly
The chromatograms are shown along with fractions that were analysed by SDS-PAGE connected by dotted lines. TEMs of each sample are also shown; scale bar = 50 nm.

3.3 Discussion

The results in this chapter illustrate the importance of the highly basic N-terminal region in regards to STNV capsid assembly. Many other viruses also contain highly basic N-terminal regions, and some of these viruses can assemble into empty particles when their CP N-terminus is proteolysed[34; 88; 131]. However, truncation of the N-terminus does not promote capsid assembly in STNV, and no appreciable self aggregation has been observed in this proteolysed state. Tryptic digests of EDTA-expanded STNV particles results in one cleavage product, where Arg 28 becomes cleaved, just past the N-terminal helix[80].

Previously, STNV CP truncation mutants (STNV Δ12 CP) were expressed in *E. coli* to analyse the effect of the N-terminus on capsid assembly *in vivo*[81]. During purification, it became evident that this protein was assembly deficient. This deficiency in capsid assembly implicates the first 12 residues as being vital for making appropriate contacts between either neighbouring subunits or nucleic acid. From these 12 deleted residues, 5 of them are positively charged at neutral pH, suggesting they perform crucial charge neutralisation functions. In the recombinant STNV X-ray crystal structure, it's clear that at least residues 12-24 are in close proximity to RNA[89]. Residues 1-11 are invisible in both recombinant and wild-type STNV crystal structures, probably due to these residues being disordered[77; 78; 89]. It is conceivable that these residues make contacts with the asymmetric RNA genome, and therefore density is lost by icosahedral averaging.

Despite multiple attempts, no capsid assembly was observed in the absence of RNA. Sedimentation velocity analysis of STNV CP overwhelmingly suggests that STNV CP exists as a monomeric species (>99 %) under both disassembly and reassembly conditions, with minority species existing in trimeric and pentameric forms. It is possible that the reason why STNV CP will not self-assemble into capsids in the absence of RNA is because of repulsive effects from the highly basic N-terminus. If protein-protein interactions were the dominant determinant of capsid assembly, then surely there would be assembly products when proteolysed STNV CP is placed under assembly conditions *in vitro*. In addition, STNV Δ12 CP failed to assemble *in vivo*, whereas the full length CP produced STNV VLPs under the same conditions. This suggests RNA: CP interactions are critical for capsid assembly.

But what is the nature of these RNA: CP interactions? Are they sequence specific, or are they solely dependent on electrostatic interactions between the N-terminus of the CP and the phosphodiester backbone of the RNA? These questions have partially been answered by SELEX experiments in the presence of immobilised STNV CP[81]. The results of this analysis showed that from the 17 aptamers generated, all but 2 had a stem-loop structure displaying the motif (AXXA) in a 4, 5 or 6 loop, respectively. Bioinformatics analysis showed that these aptamers had significant sequence similarity to several positions on the STNV-1 genome. The functional relevance of one of these aptamers, B3, was analysed by capsid assembly experiments *in vitro*. These capsid assembly experiments showed that aptamer B3 could initiate capsid assembly *in vitro* and that minimal fragments (30-mers) were capable of supporting assembly. The $A_{260:280}$ ratios of these STNV-B3 VLPs were ~1.7, suggesting that multiple stem

loops were encapsidated. The next Chapter describes the sequence specificity of the STNV: CP interactions on the efficiency of STNV VLP assembly, and characterises the X-ray crystal structure of STNV-B3 VLP.

4. Sequence-specific interactions overcome an electrostatic barrier preventing capsid assembly of empty STNV VLPs

4.1 Introduction

Previously, it was determined that STNV CP forms sequence-specific interactions with a series of stem loops, each displaying an AXXA motif[81]. This motif can be displayed in the context of a 4, 5 or 6 nucleotide loop. In the previous Chapter, it was shown that aptamer B3 could trigger capsid assembly *in vitro*, and that STNV CP will not assemble in the absence of RNA.

This Chapter addresses whether other RNAs support STNV assembly, what the efficiency of capsid assembly is relative to each of these RNAs, and whether multiple copies of the AXXA motif within genomes are functionally significant. This has been explored using a combination of svAUC and TEM to determine the relative capsid assembly efficiency of STNV CP in the presence of these different RNAs. These include aptamer B3 and a sequence variant, other short RNAs that present the AXXA motif differently and longer RNAs including a cognate STNV genome (see Figure 4.1). These RNAs were subjected to reassembly conditions at a fixed concentration, with increasing amounts of STNV CP titrated into these RNAs. In addition, VLPs were reassembled in the presence of aptamer B3, B3 short and TR on a large scale and purified. These purified VLPs were crystallised. STNV-B3 VLP crystals diffracted to a high resolution and their structure was solved by X-ray crystallography to 2.3 Å.

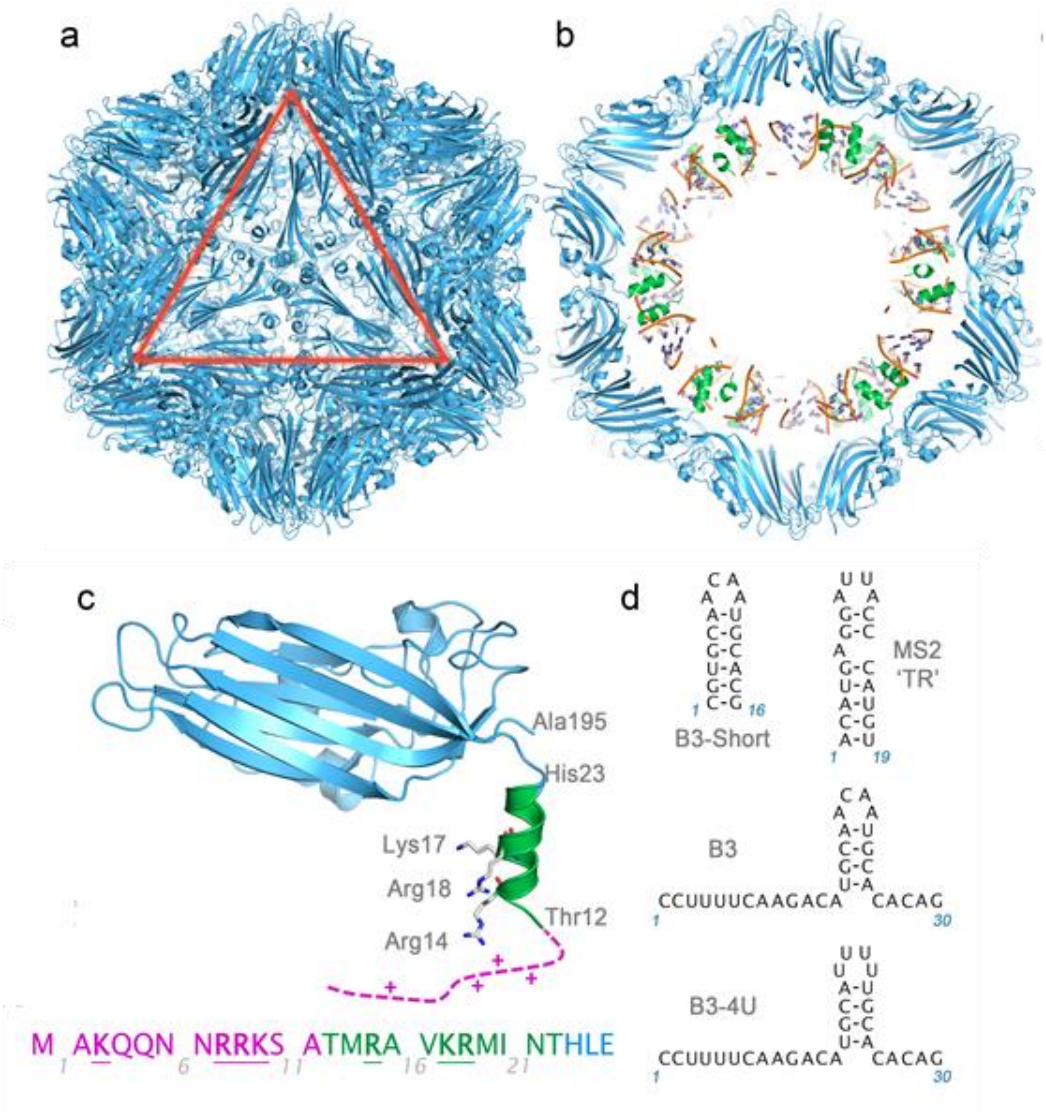


Figure 4.1: The STNV system and stem loops used in capsid assembly
(a) The STNV capsid.
The X-ray crystal structure of recombinant STNV is shown as a blue cartoon. The view is looking down the three-fold axis of symmetry.
(b) Central view through the STNV VLP
A 45 Å slab is cut through the STNV VLP, where the N-terminal helices are coloured green, and the structure of 7 nucleotides per coat protein monomer of ordered RNA is also shown (orange).
(c) Close up view of the STNV coat protein subunit.
STNV CP is a wedge-shaped jelly roll β-sandwich. The N-terminal region contains a net positive charge, where it is visible until Thr12. The unstructured portion (magenta) of the N-terminus contains a further four basic amino acid side chains as indicated by the underlined sequences shown.
(d) The sequences and MFold predicted secondary structures of the RNA stem loops used for probing capsid assembly.
B3 short and MS2 TR are shown next to each other, with B3 and B3 4U shown below. Note B3 and B3 4U are identical in terms of their sequences and structures other than the 4 residues in the loop. Figure taken from Ford *et al* (submitted) courtesy of Dr Neil Ranson.

4.2 Results

4.2.1 RNA production and purification

In order to investigate the sequence specificity of RNA on the efficiency of capsid assembly, both STNV CP and RNAs must be amenable to experimentation. Chapter 3 describes in detail how STNV CP was purified and the measures that were put into place to ensure the production of full length, assembly-competent STNV CP. Similarly, the RNAs were prepared diligently, being careful to avoid RNase contamination. These were prepared either by *in vitro* transcription from an already characterised pre-existing DNA template or by using RNA synthesis, utilising solid-phase phosphoamidite chemistry as described in Chapter 2.4.4. Oligonucleotides ≤30 nt in length were prepared by solid phase synthesis, purified and characterised by mass spectrometry and denaturing PAGE, as described in Chapter 2.2.1.4. RNAs >30 nts were transcribed from a pre-existing template and purified as described in Chapter 2.2.3. The purified products were analysed by 1 % (w/v) denaturing agarose gel electrophoresis as described in Chapter 2.2.1.3.

The capsid assembly efficiency of aptamer B3 and a sequence variant, B3 4U, were compared by STNV CP titrations using a svAUC assay and complemented by TEM images. These RNA oligonucleotides were identical in sequence, other than the ACAA loop sequence of B3 was replaced with UUUU in B3 4U (see Figure 4.1). This sequence variant was designed to test the importance of the AXXA motif for capsid assembly. A truncated version of the B3 stem loop was also prepared with an additional base pair to stabilise the stem and a single-stranded base at either end (B3 short, 16 nt). In addition, the

known capsid assembly initiation signal in MS2, TR (19 nt)[64], was synthesised. This RNA stem loop oligonucleotide is composed of a stem interrupted by a single base bulged adenine, but displays the motif (AUUA) in the loop. Capsid assembly using these oligomeric RNA stem loops were compared, both to analyse whether the sequence of the stem was important for capsid assembly, and whether the 5' and 3' flanking sequences in aptamer B3 affected capsid assembly.

Although STNV VLPs can package RNA oligonucleotides, STNV packages its genome of 1239 nt *in vivo*. It was shown previously that the STNV-1 genome contains up to a total of 30 stem loops displaying the -AXXA- motifs[81]. Longer RNAs were therefore prepared by *in vitro* transcription from DNA templates for capsid assembly. The STNV CP mRNA[89] (590 nt) and a cognate STNV-C[115] (1221 nt) genome were transcribed from digested pET22b-STNV[89] and pUBS-STNV-C[115] plasmids, respectively. These RNAs were compared for *in vitro* capsid assembly studies to determine whether the stem loops in the STNV genomes were functionally significant. STNV CP mRNA contains fewer stem loops than the STNV-C genomes, both because the mRNA is shorter and because it was mutated to ensure it was codon optimised for efficient translation in *E. coli* [89]. It was therefore hypothesised that STNV-C RNA would be more efficient at assembly than STNV mRNA.

Similarly, multiple putative packaging signals (PSs) have recently been identified in the MS2 genome. These PSs were identified using a combination of bioinformatics, structural and biophysical data (Dykeman *et al*, submitted). There are similar PSs present in the genomes of STNV. For these reasons, two

MS2 sub-genomic RNA fragments were prepared. One of these RNAs, MS2 iRNA (927 nt), was transcribed from a PCR template derived from the pSMART plasmid encompassing the internal fragment from the MS2 genome. This RNA contained the TR stem-loop as well as similar PSs, but is ~25% shorter than the STNV genome. Furthermore, these PSs may not be optimally spaced for STNV encapsidation. For these reasons, it was hypothesized that MS2 iRNA should trigger capsid assembly, but may not be as efficiently packaged as the cognate genome. MS2 5'RNA was transcribed directly from the same linearised pSMART plasmid, and is 2469 nt long (i.e. roughly twice as large as the STNV genome). The reason why this RNA was selected is because it contains multiple PSs for MS2, but should not be encapsidated by STNV CP, due to the size constraints imposed on the interior of the *T*=1 capsid.

In vitro transcription was performed using the commercial MEGAScript (Ambion®) transcription kit using manufacturer's instructions, and purified as described in Chapter 2.2.3. The yield routinely obtained was ~10 µg per 25 µL transcription reaction per RNA. Each of the long RNAs were analysed by denaturing 1 % (w/v) agarose gel electrophoresis to determine the size and the purity of the product (see Figure 4.2a).

Oligonucleotides (≤30 nt) were synthesised and purified by an established in-house method by Amy Barker according to Chapter 2.4.4. Each oligonucleotide was subjected to denaturing PAGE and the molecular weight was determined by ESI mass spectrometry (see Appendix, Figures 9.3-9.6).

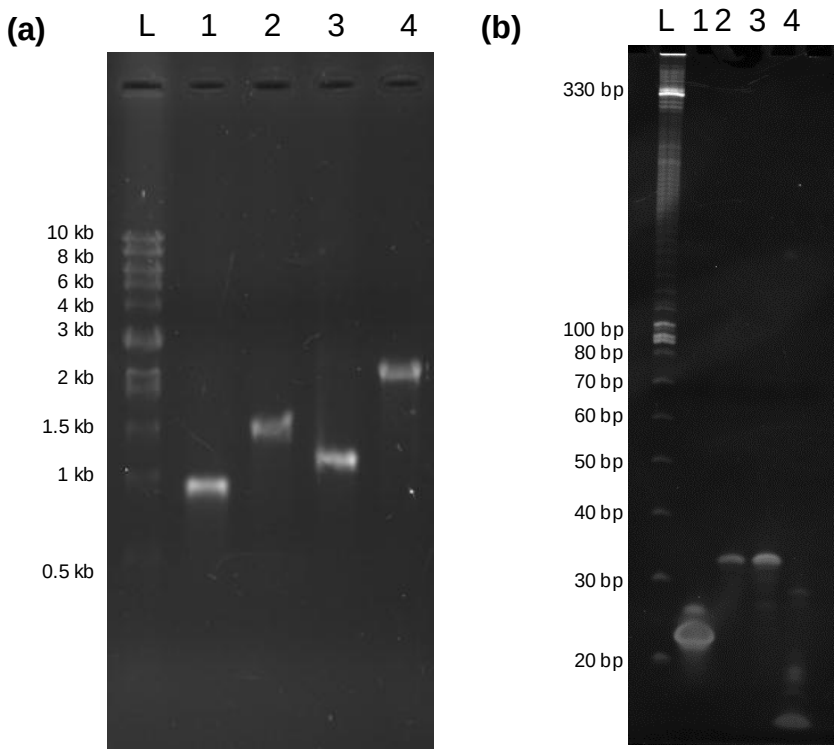


Figure 4.2: Gel electrophoresis of RNAs
(a) Denaturing agarose electrophoresis of RNA transcripts
1 = STNV mRNA; 2 = STNV-C RNA; 3 = MS2 iRNA; 4 = MS2 5' RNA; L = 1 kb DNA ladder (NEB)
(b) Denaturing PAGE of RNA oligonucleotides
1 = MS2 TR; 2 = B3; 3 = B3 4U; 4 = B3 short; L = 10 bp DNA ladder (NEB)

4.2.2 Determining the effect of sequence-specificity by STNV CP titrations

In order to determine whether capsid assembly was sensitive to the preferred loop sequence motif derived by SELEX (AXXA)[81], two 30-mer oligonucleotides encompassing variations on this motif were prepared (see Figure 4.1). These consisted of aptamer B3 and an otherwise identical fragment (B3 4U) in which the 4 base loop sequence of B3 (ACAA) was replaced (UUUU). The efficiency of these fragments at promoting capsid assembly was tested. STNV CP was then titrated into a fixed concentration of RNA (2 μM) in each reaction and dialysed against reassembly buffer. For 0.5:1 reactions, this compared 1 μM STNV CP in 2 μM RNA, i.e. less than stoichiometric with respect to a *T*=1 VLP,

if each CP subunit binds one RNA stem-loop, rising to 5:1 or 10 μM of STNV CP. The end products of assembly were analysed using sedimentation velocity analytical ultracentrifugation (svAUC) and negative stain transmission electron microscopy (TEM) as described in Chapters 2.4.6 and 2.2.2, respectively. The svAUC was analysed using a continuous distribution c(S) vs. S Lamm equation model[118]. The percentage of assembled *T*=1 capsids in each reaction was estimated from the area of the peaks in the c(S) vs. (S) plots (see Table 4.1). The data are displayed as c(S) vs. S plots for CP: RNA molar ratios ranging from 0:1 to 5:1. Under the same buffer conditions, the recombinant STNV VLP has a sedimentation coefficient of 42 S (indicated by grey line in each titration)[81]. There is an obvious difference in the efficiency of capsid assembly between aptamer B3 and the mutant lacking the AXXA motif (Table 4.1, Figure 4.3). Aptamer B3 promotes formation of species sedimenting at 42 S at a molar ratio of 1:1. This is below the protein: RNA stoichiometry expected for a *T*=1 capsid with 30 stem-loops bound. As well as the 42 S species, the reassembly also yields an intermediate species sedimenting at ~28 S. TEMs show that they correspond to intact *T*=1 capsids (42 S) and incomplete shells with the dimensions compatible with a *T*=1 architecture (28 S), respectively. In contrast, the mutant sequence appears to be less efficient at stimulating assembly at a 1:1 ratio, and is significantly hindered at a ratio of 2.5:1. Even at high STNV CP concentrations (5:1) the mutant fragment creates more of the 28 S intermediate than the VLP. These results are fully consistent with the AXXA motif being a preferred RNA binding sequence for STNV CP[81].

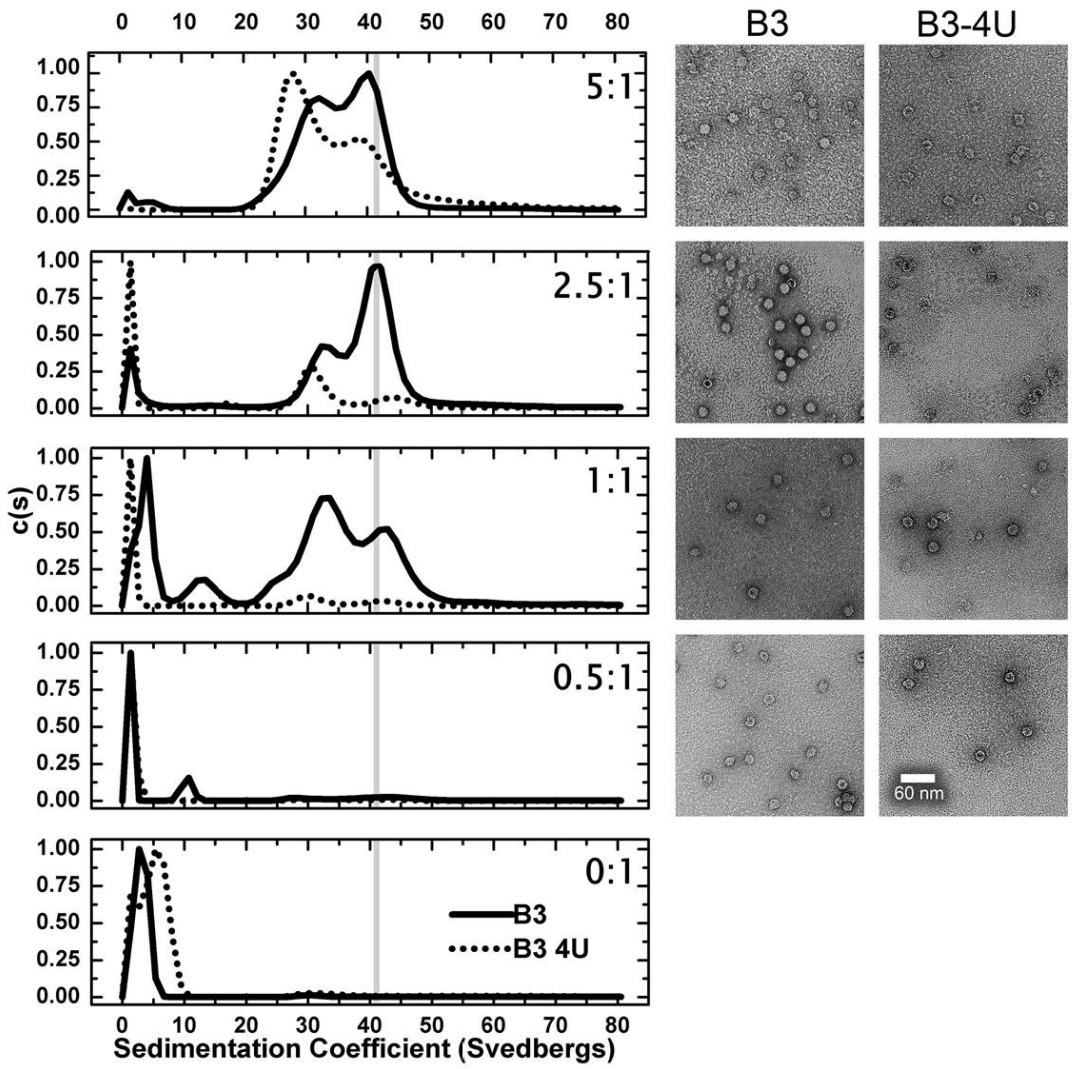


Figure 4.3: A comparison of STNV CP assembly efficiency reassembly using B3 4U and B3 RNAs.
Left panels show a titration of increasing CP: RNA molar ratios in reassembly reactions carried out as described in Chapter 2.4.5 analysed by velocity sedimentation. The corresponding images on the right are TEMs of the sample at the end of capsid assembly reactions. The RNA concentrations were held constant at 2 μM. Grey line represents the $S_{w,20}$ of the recombinant VLP.

4.2.3 Purification of 28 S intermediate from B3-reassembled VLPs

In order to determine whether the 28 S intermediate could be isolated, a large scale capsid assembly reaction at a 1:1 molar ratio was prepared, as described in Chapter 2.4.8. Because sucrose density gradient fractionation and svAUC

are similar techniques, separation was attempted by sucrose density gradient fractionation with a 15-45 % (w/v) gradient. During fractionation, it was evident that 3 species were present, analogous to the sedimentation velocity assay. The first peak at the top of the gradient eluted between 0-1 mL, where the high absorbance at 260 nm suggests this peak is aptamer B3.

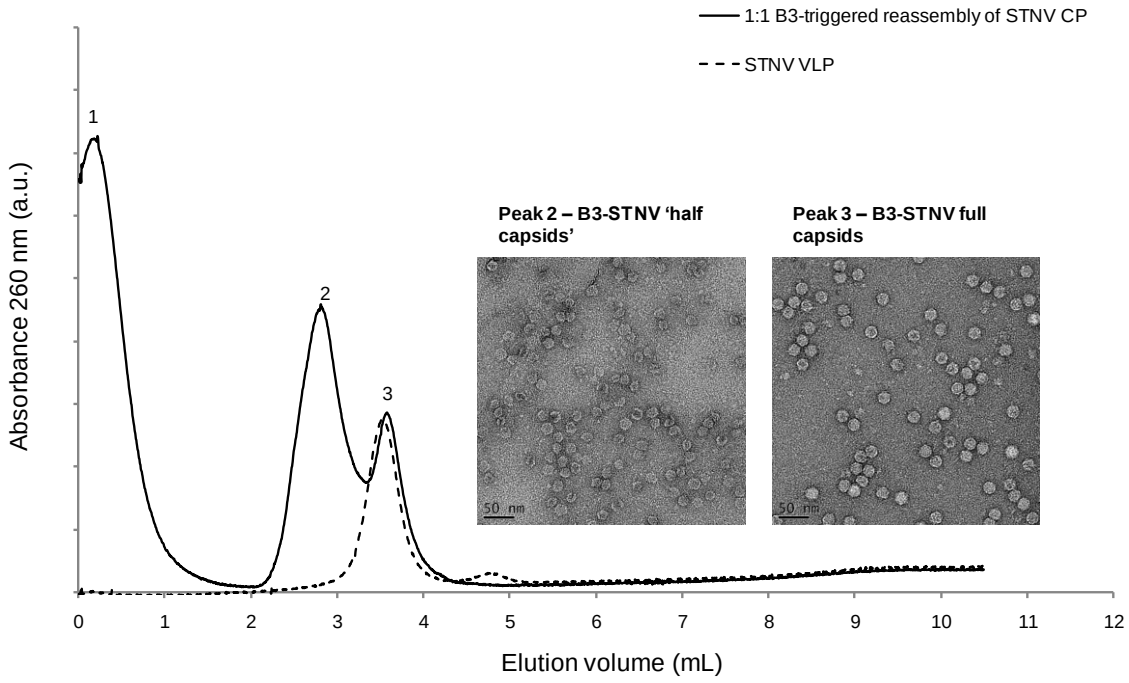


Figure 4.4: Separation of STNV 'half capsids' from complete capsids
STNV capsid assembly was performed with STNV CP and aptamer B3 at a 1:1 molar ratio as described in Chapter 2.4.8, where the products were purified by sucrose density gradient fractionation. Two higher order species were separated, where the corresponding peaks have been labelled 2 and 3. Peak 1 is free aptamer B3. Peaks 2 and 3 were dialysed in native buffer to remove the sucrose and analysed by TEM (inset). Peak 2 contained STNV 'half capsids', where the curvature was similar to the complete *T*=1 capsid. Peak 3 contained purified STNV particles with a *T*=1 morphology.

Peaks 2 and 3 were collected and dialysed against native buffer. These samples were then analysed by UV absorbance spectroscopy and TEM to determine their $A_{260:280}$ ratios and morphology of these species relative to STNV VLPs. The UV absorbance showed that STNV VLPs prepared at this concentration ratio isolated from peak 3 had an $A_{260:280}$ ratio of ~1.2, whereas

the species isolated from peak 2 had an $A_{260:280}$ ratio of ~1.4. This suggests that both species contained less RNA than both wild-type STNV and recombinant STNV. The implication is that STNV VLPs can assemble at this sub-stoichiometric concentration, despite a subset of these STNV CP binding sites not being occupied by RNA. TEMs of these fractionated species clearly show that the 28 S species (peak 2) are partially assembled STNV VLPs with similar curvature to recombinant particles (see Figure 4.4, peak 2 (inset)). Further analysis of this species by sedimentation equilibrium suggests that the molecular weight of these species is ~750 kDa, which roughly corresponds to half the molecular weight of an STNV capsid (Appendix, Figure 9.7). Presumably this species is an intermediate en route to a complete STNV-B3 VLP, as the relative amounts of T=1 capsid increase with STNV CP concentration.

4.2.4 Removal of flanking sequences increases the efficiency of capsid assembly, but efficiency is indifferent to base-paired sequences in the stem

A truncated version of aptamer B3 encompassing the stem loop region was synthesised. In order to stabilise its secondary structure as a stem loop, an additional base pair was added to extend the stem (B3 short, 16 nt). In addition, TR, the known assembly initiation site for RNA bacteriophage MS2 (19 nt), which has a loop sequence of (AUUA) was prepared (see Figure 4.1d). B3 short and TR present their AXXA motif on either a fully base paired stem or one interrupted by a single base bulge (see Figure 4.1d). However, they both lack the 5' and the 3' flanking sequences of aptamer B3. In contrast to the loop sequence sensitivity seen previously, both the shorter stem loops promote capsid assembly with similar efficiencies. Formation of the intermediate species

is only seen at high CP concentrations, and even then the intermediate species are in minority. B3 short is significantly more efficient at promoting capsid formation than aptamer B3, where at the highest CP concentration, approximately 90 % of the material being *T*=1 capsids compared to ~50-60 % for the larger RNA encompassing the same preferred binding motif. These results suggest that flanking sequences either side of the stem loop are partially inhibitory of assembly, and that capsid assembly can be initiated by stem loops as short as 16 nt in length. It also suggests that capsid assembly is not sensitive to sequences in the stem of the oligonucleotide.

Table 4.1: Capsid assembly efficiency of *T*=1 particles with different RNAs

[CP]	**1 µM**		**2 µM**		**5 µM**		**10 µM**	
RNA	***%RNA***	***%T=1***	***%RNA***	***%T=1***	***%RNA***	***%T=1***	***%RNA***	***%T=1***
B3	53.7	13.6	19.9	26.9	7.2	61.9	3.0	52.6
B3 4U	87.7	4.2	66.4	11.6	35.3	15.4	0.1	40.6
B3 Short	83.1	14.8	57.3	31.6	2.7	92.9	4.0	86.7
TR	80.1	15.1	64.7	29.2	15.9	69.9	0.9	73.4
mRNA	11.4	35.0	1.3	23.0	0.5	37.3	0.0	25.3
STNV-C	18.4	19.9	7.4	54.1	0.0	23.1	0.0	27.7
MS2-i	39.3	0.0	26.2	*	0.0	50.5	0.0	41.3
MS2-5'	*	0.0	6.5	*	0.0	14.9	0.0	17.0

* indicates where the peaks of the c(S) vs. S plots overlap making it impossible to distinguish these components from other reassembly intermediates.

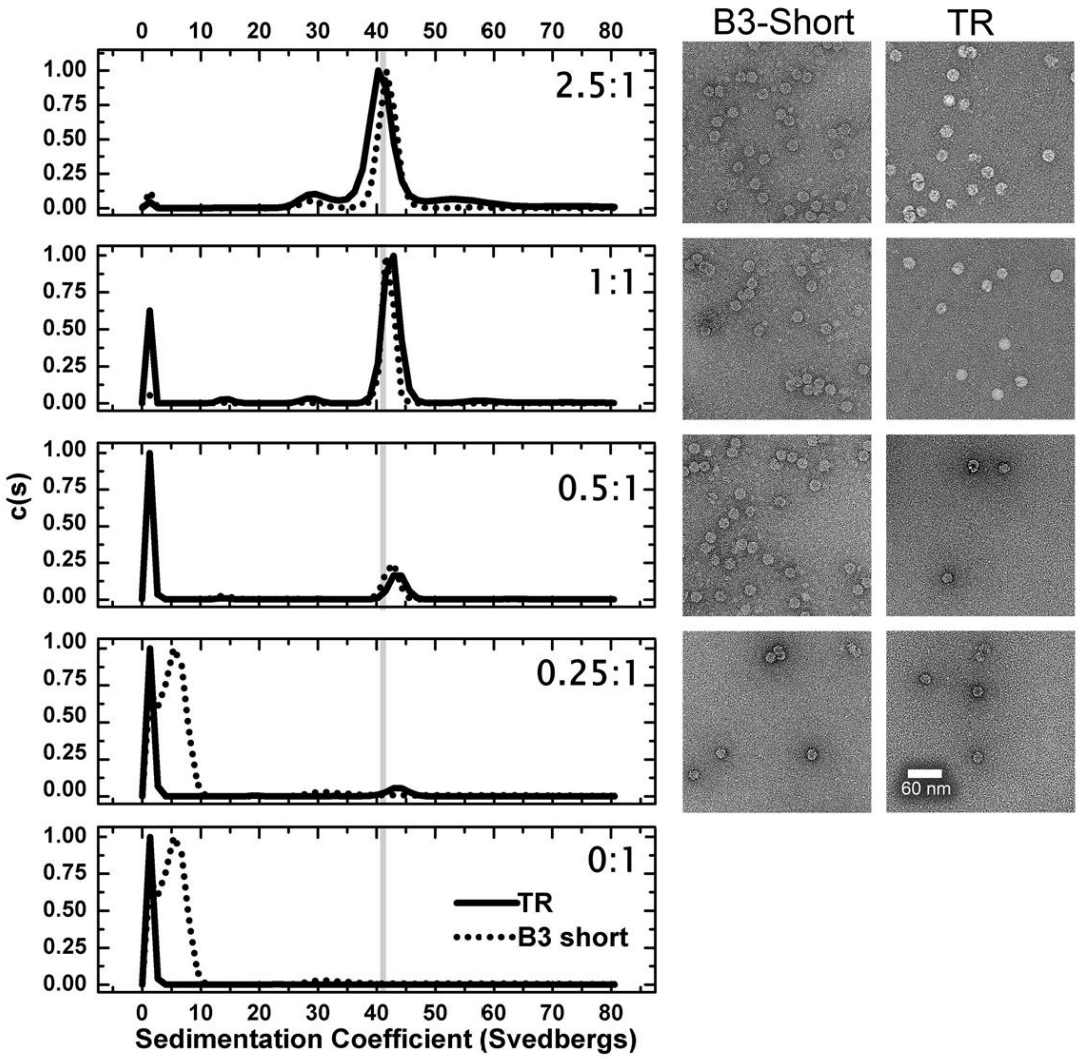


Figure 4.5: A comparison of STNV CP assembly efficiency reassembly using B3 short and MS2 TR stem loops. Left panels show a titration of increasing (from bottom to top) CP: RNA ratios in reassembly reactions as described in Chapter 2.4.5. The corresponding TEM images (right) are the representative of the sample at the end of the reassembly reaction. The RNAs were kept at a fixed concentration of 4 µM.

4.2.5 STNV X-ray structure

4.2.5.1 Large-scale purification of STNV-B3, -B3 short and -TR VLPs for crystallisation

In order to characterise the structure of the VLPs by X-ray crystallography, the capsid assembly reactions in the presence of aptamer B3, B3 short and TR were repeated on a larger scale. Large scale STNV CP purifications were performed as described in Chapters 2.3.9 and 3.2.10, respectively. These large scale reassembly reactions were purified by size exclusion chromatography using a Superdex-200 10/300 GL column. VLPs eluted between 8-12 mL, the fractions were analysed by SDS-PAGE and VLPs from 8-12 mL collected and pooled (see Figure 4.6). The VLPs were then dialysed in ddH_2O and concentrated using Amicon Ultra® spin concentrators. The purified version from all of these VLPs had an $A_{260:280}$ ratio of ~1.7, which indicates they had a comparable amount of encapsidated RNA relative to one another and to the recombinant VLP.

The STNV VLPs described above were then subjected to crystallisation conditions as described by Lane *et al*.[89] (Crystallisation, cryo-protection and data collection of STNV-B3 VLPs were performed by Dr Saskia Bakker). Each sample produced crystals that diffracted well, however full data sets could not be obtained from crystals of STNV-TR and STNV-B3 short VLPs. These crystals appeared smaller, cracked and thinner than their STNV-B3 VLP counterparts. The reasons for this are unclear, but could be due to differences in crystal packing. It may be that further optimisation of crystallisation conditions are required for the acquisition of crystals that produce full data sets. Due to these reasons, data concerning the large-scale production of STNV-TR and

STNV-B3 short VLPs for crystallisation will not be presented here, but chromatograms are available in the Appendix (Figures 9.8 and 9.9).

The data from STNV-B3 were collected by Dr Saskia Bakker at the Diamond Light Source on Beam line IO2, using a wavelength of 0.9795 Å. These data sets were recorded at different detector distances and beam stop distances, generating data over a range of resolutions (see Figure 4.8). It is important that data are collected at high and low resolutions because the RNA component is not seen at high resolution with STNV[77; 78; 89].

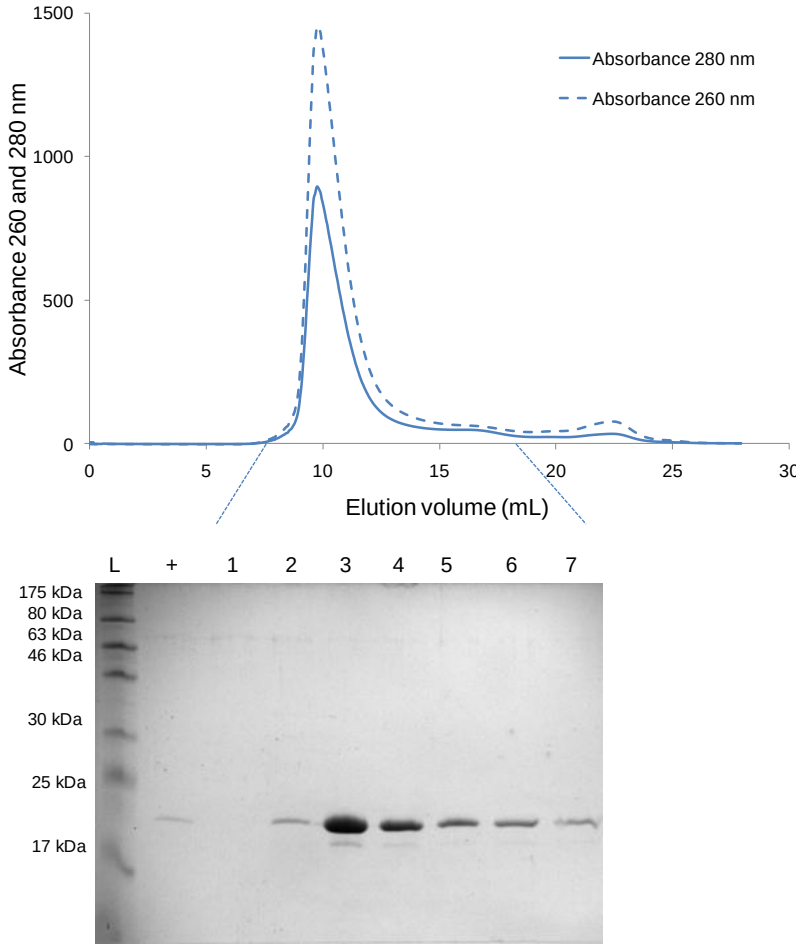


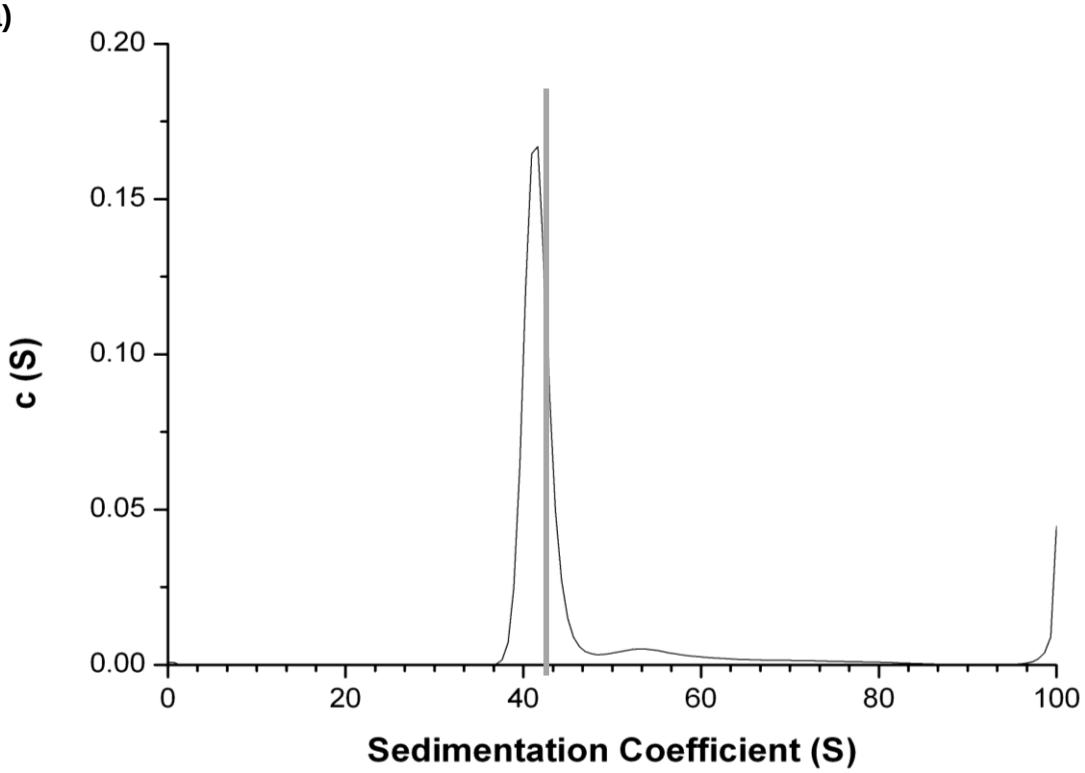
Figure 4.6: Large-scale purification of STNV–B3 VLPs
STNV-B3 VLPs were purified by size exclusion chromatography (top). The VLPs eluted between 8-12 mL. Fractions were analysed by SDS-PAGE (bottom) where the dotted lines signify the fractions that were analysed. + = STNV VLPs; L = 7-176 kDa prestained protein marker (NEB)

4.2.5.2 Characterisation of STNV-B3 VLPs

STNV-B3 VLPs were characterised by svAUC and TEM to assess the quality of the sample prior to crystallisation (svAUC was performed by Amy Barker as described in Chapter 2.4.6.2).

The svAUC analysis shows that the overwhelming majority of the sample sediments at a $S_{20,w}$ = 41.8. This corresponds very well with the value obtained from svAUC of recombinant VLPs ($S_{20,w}$ = 41.7)[81] (see Figure 4.7a). The TEM also confirms the presence of purified STNV-B3 VLPs, with few aggregates or partially-formed VLPs present in the sample (see Figure 4.7b). These results suggest that size exclusion chromatography has sufficiently purified STNV-B3 VLPs in a way to support crystallisation.

(a)



(b)

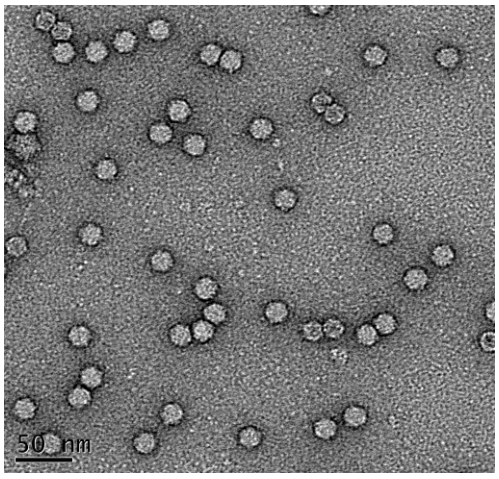


Figure 4.7: Characterisation of STNV-B3 VLPs
(a) svAUC of STNV-B3 VLPs purified by size exclusion
svAUC analysis shows that the VLPs are uniform in size and sediment with similar $S_{20,w}$ relative to recombinant STNV VLPs. STNV VLPs control is shown as a grey line.
(b) TEM of STNV-B3 VLPs
TEMs were captured at 49,000X magnification. Scale bar (bottom left) = 50 nm

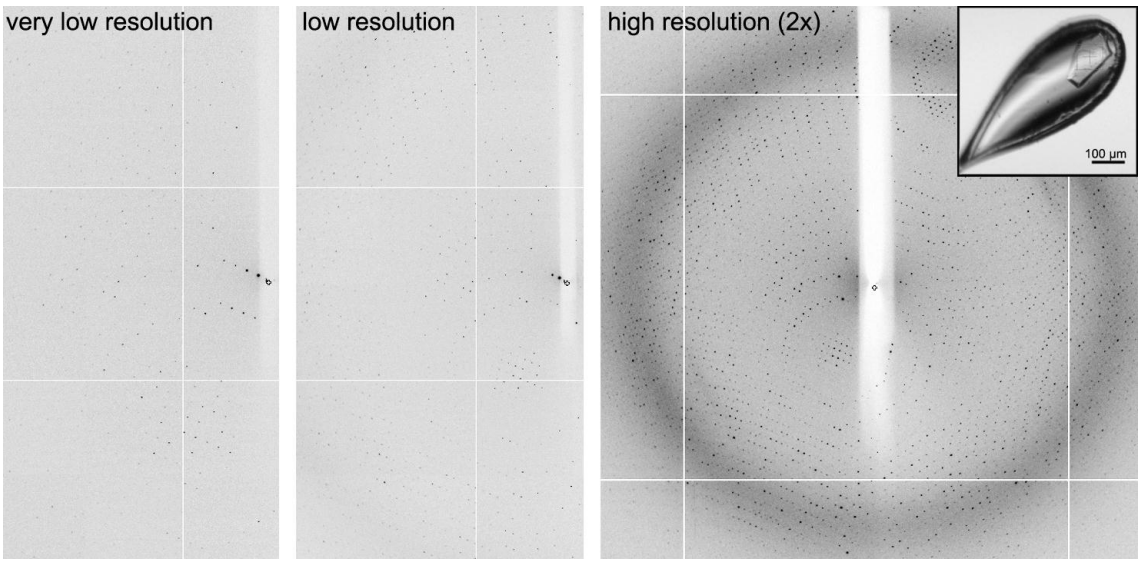


Figure 4.8: STNV-B3 diffraction patterns.
Data collection from STNV-B3 VLPs were recorded at very low resolution (217-30 Å; left), low resolution (116-3.8 Å; middle) and high resolution (30-2.29 Å; right). The crystal that was used for the collection of these data sets is shown top right. Figure taken from Dr Saskia Bakker's thesis.

4.2.5.3 STNV-B3 protein structure

The structure of STNV-B3 VLP coat protein is very similar to those of previously published structures[77; 78; 89]. The CP has a backbone r.m.s.d. of 0.138 Å compared to wild-type STNV-1 (PDB entry 2BUK)[77] and 0.271 Å compared to the VLP containing the mRNA (PDB entry 3S4G), respectively[89]. From these structures, the most variable region identified is the N-terminal helix, which was previously described as the most flexible[89]. The main difference between the STNV-B3 VLP structure and previous structures is that the N-terminus is more ordered in this structure, permitting an additional turn at the N-terminus of the helix to be modelled, which correspond to residues 8-11 (RKSA) (see Figure 4.9). This suggests that sequence-specific binding of aptamer B3 juxtaposes the RNA such that phosphodiesters in the stem help to overcome electrostatic repulsions between CP N-termini. A similar mechanism is expected to occur with longer RNAs displaying multiple AXXA motifs, but due to the intrinsic

asymmetry of those RNA sequences, the conformations at the differing N-termini cannot follow strict icosahedral symmetry. This may account for the apparent disorder of the N-termini in the previous STNV structures[77; 78; 89].

4.2.5.4 STNV-B3 RNA density

In order to identify the STNV CP B3 binding site, 60-fold averaged 2Fo-Fc maps were calculated using data between 217 and 6 Å resolution. Even though the quality of the electron density does not permit the modelling of aptamer B3, there is strong electron density for the RNA that encompasses what is observed in the previously reported recombinant STNV VLP[89]. There is also considerable additional density around the end of the N-terminal helices that lie along the particle two-fold axes (see Figure 4.9b). However, this density cannot be distinguished between the remaining residues of the CP N-terminus or additional RNA.

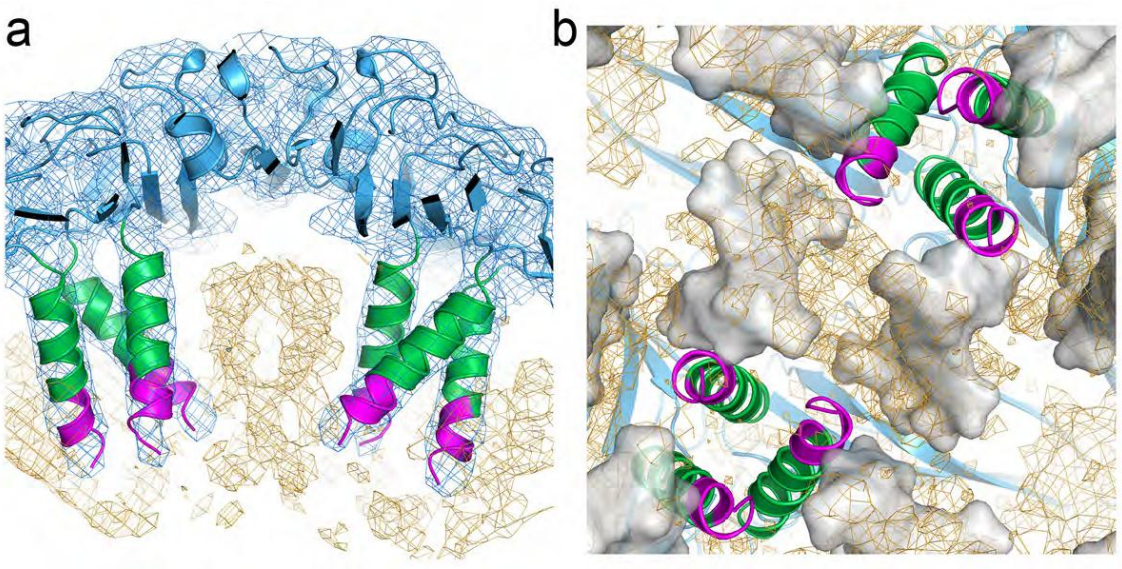


Figure 4.9: The X-ray structure of the STNV-B3 VLP.
(a) Ordered density in the N-terminal helices. The STNV coat protein is shown in cartoon representation and coloured as in Figure 4.1. The orange mesh is the 60-fold averaged 2Fo-Fc electron density of the STNV-B3 structure with the parts of the map corresponding to the STNV capsid masked away, and contoured at ~0.03 $^{e-}$/Å^3. The blue mesh is the density for the proteinaceous shell of the structure contoured at ~0.07 $^{e-}$/Å^3. The extra density at the N-terminal end of the helices is clearly visible, and an atomic model for the extra turn of helix that occupies this density is shown in magenta.
(b) The B3 RNA density. The view is from the inside of the capsid, outward along an icosahedral 2-fold axis. The protein cartoon and orange mesh are as described in (a), whilst the grey surface is the solvent accessible surface of the mRNA modelled in [89]. Figure taken from Ford *et al* (submitted).

4.2.5 The effects of multiple packaging signals on capsid assembly *in vitro*

In this Chapter and Chapter 3, I have shown that STNV CP can self-assemble in the presence of short stem loops under reassembly conditions. However, wild-type STNV-1 self-assembles around its genome of 1239 nt. A series of titration experiments were thus set up using RNAs of a longer length as described in Chapter 4.2 under the same conditions as those of the stem loops. RNA concentrations for each titration were fixed at 20 µg/mL to directly compare capsid assembly efficiency of each RNA relative to the concentration of the phosphodiester backbone. Molar ratios are shown in Table 4.2. The results obtained from each titration were strikingly different.

The STNV-C RNA forms a considerable amount of VLP (~50%) at a CP concentration of 2 μM compared to ~20% for the mRNA at this concentration. Subsequently at higher ratios (≥5:1) the STNV-C appears to form a second higher order species which TEMs suggest are fused double capsids. Subsequent RNase A treatment of these fused capsids result in the double shells falling apart (see Figure 4.12). This result is suggestive of multiple assembly initiation events on the same RNA and that these fused capsids are tethered by RNA. Such double shells also occur at lower ratios on the mRNA, which is less efficiently packaged throughout the titration.

The results from this series of titrations suggest that the STNV-C genome is more efficient at producing *T*=1 capsids than a mutated shorter sub-fragment. This makes sense if the multiple preferred binding sites, suggested to exist in Bunka *et al.*, have sequences at defined spacings on the RNA genome that facilitate the process[81]. Further, it is also consistent with the specificity and fidelity of assembly seen in single molecule Fluorescence Correlation Spectroscopy (smFCS) experiments[60]. Note, the CP subunits in these assembly reactions are from STNV-1 and so the genomic interaction with the STNV-C genome is not specific to its cognate coat protein. However, both RNAs are expected to present multiple copies of the preferred protein binding stem loops at similarly defined spacings[81]. Unfortunately, the STNV-1 genome could not be obtained for these CP titrations monitoring capsid assembly. It would be interesting to see how capsid assembly efficiency varied between the STNV-1 genome and the other sequences of RNA.

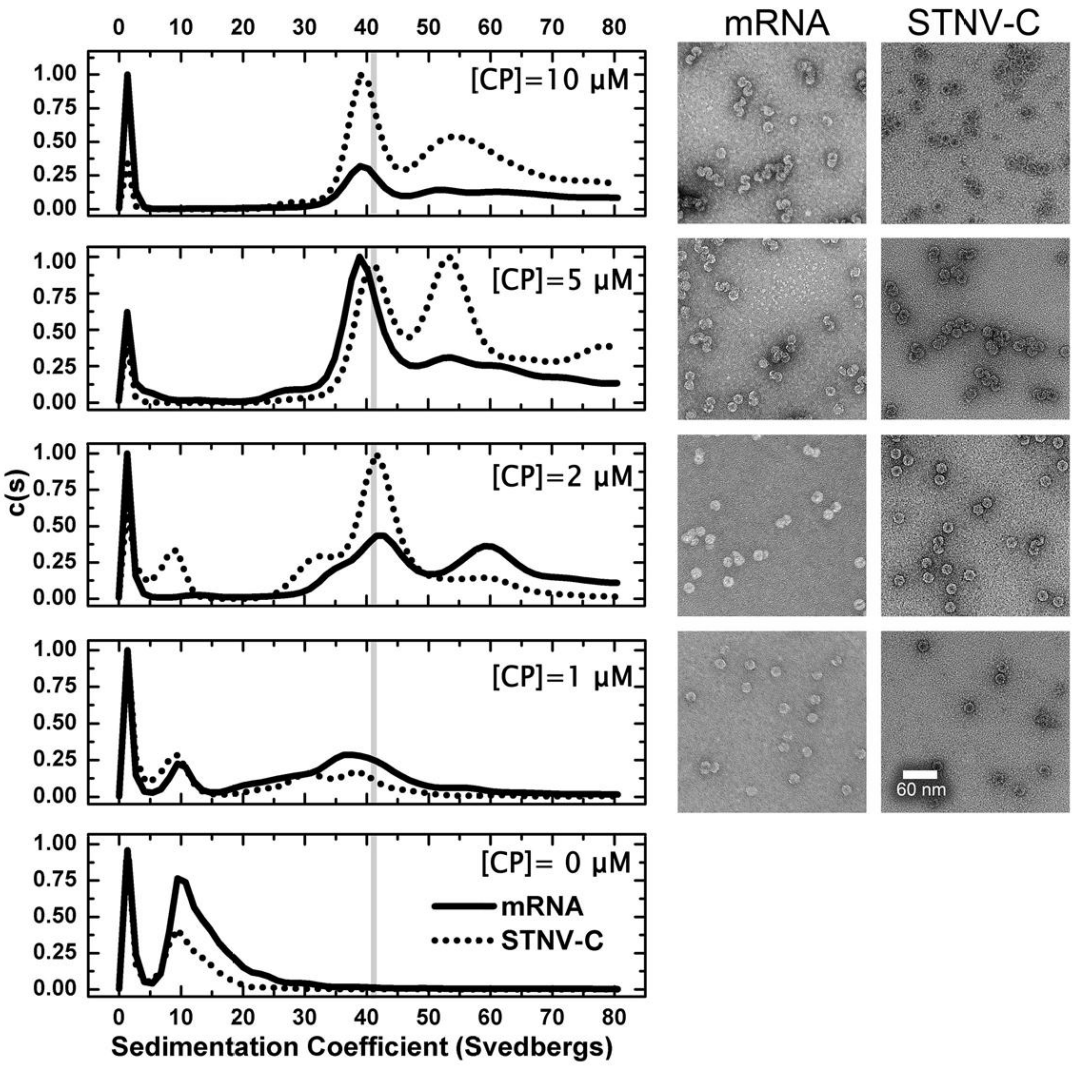


Figure 4.10: A comparison of the STNV CP assembly efficiency reassembly using STNV mRNA and STNV-C RNA. Left panels show a titration of increasing (from bottom to top) CP: RNA ratios in reassembly reactions as described in Chapter 2.4.5. The corresponding TEM images (right) are representative of the sample at the end of each reassembly reaction. The RNAs were at a fixed concentration of 20 µg/mL (100 nM mRNA; 50 nM STNV-C RNA)

Table 4.2: Concentrations of long RNAs

RNA	Length	Concentration
STNV mRNA	590 nt	100 nM
STNV-C RNA	1221 nt	50 nM
MS2 iRNA	927 nt	60 nM
MS2 5'RNA	2468 nt	24 nM

4.2.6 Sub-genomic MS2 RNAs are not efficiently packaged by STNV CP

Similar multiple putative packaging signals (PSs) have recently been identified in the MS2 genome (Dykeman *et al*, submitted). These putative PSs are variations on the TR stem loop used here to trigger STNV assembly. It was hypothesised that STNV RNAs would be more efficient at triggering capsid assembly compared to the MS2 counterparts because the structure and spacing between binding sites are different relative to the two sets of RNAs. Discrimination between genomes in assembly has already been shown by the respective coat proteins in the single molecule FCS assays, which re-emphasises the importance both of such sequences and their locations[60].

Many of the MS2 PSs contain a loop with an -AXXA- motif, and so capsid assembly was examined *in vitro* by packaging sub-genomic MS2 RNAs with the STNV CP in titration assays described previously. Two MS2 sub-genomic RNAs were compared, iRNA (927 nt) and MS2 5'RNA (2469 nt). Astonishingly, both RNAs will support capsid assembly as shown by the production of particles which sediment at 42 S (see Figure 4.11). However, the majority of the products from MS2 5'RNA sediments with a much higher S-value than STNV VLPs. TEM images show that they consist of clumps of multiple capsids, each roughly the same diameter as STNV capsids. Such species could be created if multiple assembly initiation events occurred on a single RNA and proceed independently. Since these cannot form completed capsids the RNA cannot be fully protected, and brief treatment with RNase A released single particles from the larger aggregates consistent with this view (see Figure 4.12). Similar formation of these multi-capsids have been observed *in vitro*, when packaging “excess” RNA has been reported for cowpea chlorotic mottle virus[132]. In

addition, capsid assembly from MS2 RNAs is of a far lower efficiency compared to the STNV RNAs. At 1 µM, two defined species are present in assembly with MS2 5' RNA. One of these species sediments at ~35 S, suggesting partially-assembled capsids are present. TEM images of this sample shows that these species are indeed incomplete capsids, with what appears to be RNA protruding out at one distinct site (see Figure 4.11, first TEM panel). The implication is that STNV CP may struggle to completely package the large size of MS2 5'RNA. In contrast, MS2 iRNA does not initiate capsid assembly with high efficiency at low CP concentrations. Only at a CP concentration of 10 µM are capsids observed with any high degree, with a subset of these species existing as large aggregates. Together, these results suggest that the packaging signals in STNV RNAs are functional. Similar signals that are present in the sub-genomic RNAs of MS2 do not efficiently trigger assembly, whereas the STNV-C genome does.

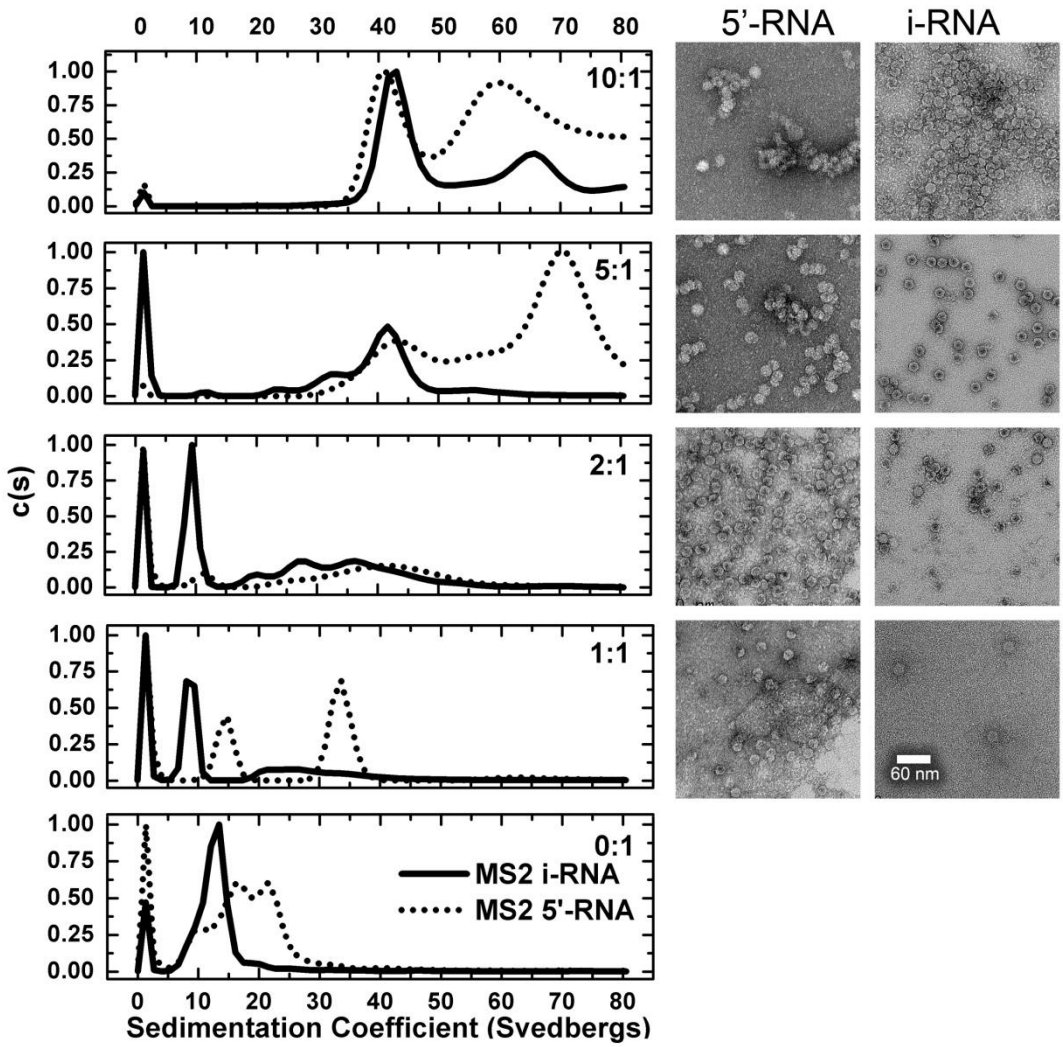


Figure 4.11: A comparison of the STNV CP assembly efficiency reassembly using STNV mRNA and STNV-C RNA. Left panels show a titration of increasing (from bottom to top) CP: RNA ratios in reassembly reactions as described in Chapter 2.4.5. The corresponding TEM images (right) are representative of the sample at the end of each reassembly reaction. The RNAs were at a fixed concentration of 20 µg/mL (MS2 5'RNA ~0.024 µM; MS2 iRNA ~0.06 µM)

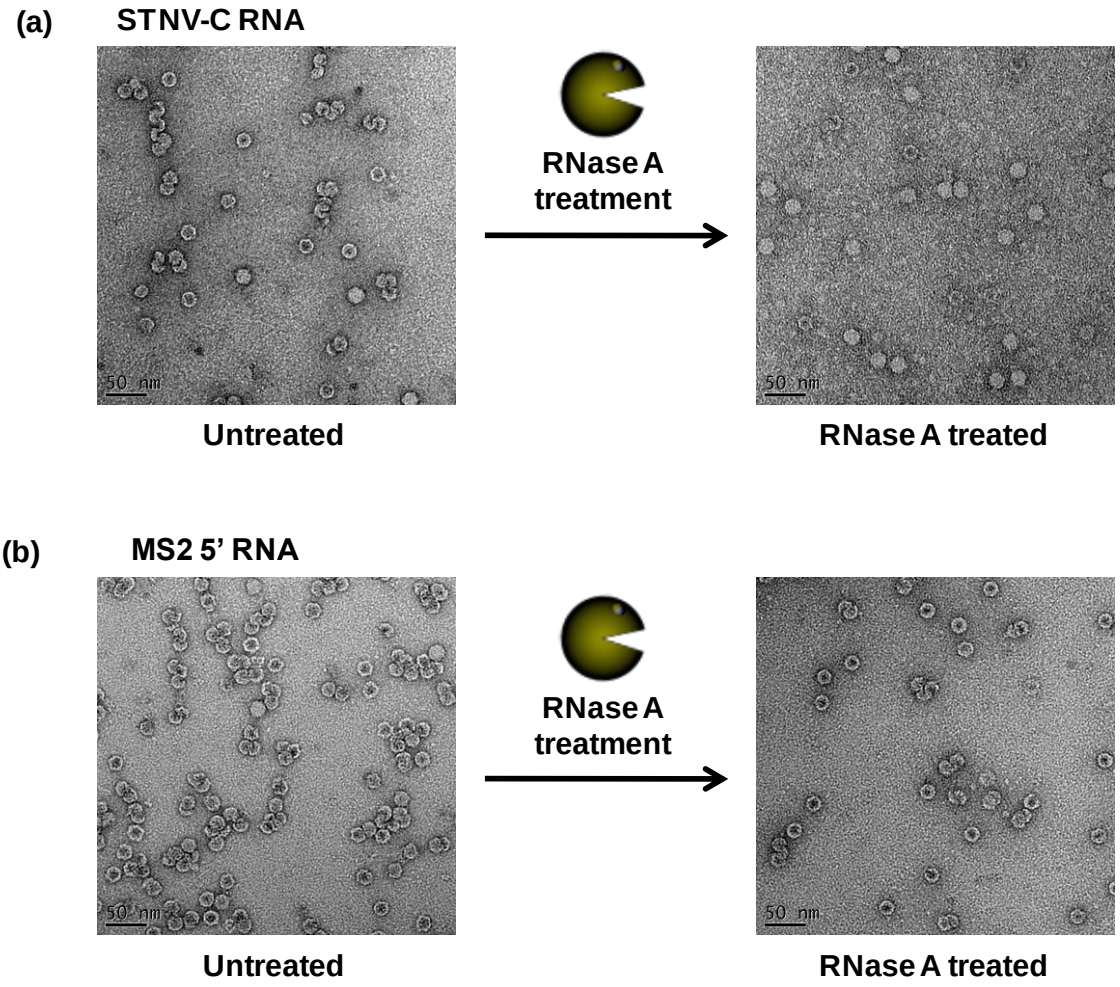


Figure 4.12: RNase A treated samples
(a): RNase A treated STNV-C reassembly
TEMs were taken of reassembled STNV CP with STNV-C RNA ([CP] = 5 µM) prior and post RNase A treatment as described in Chapter 2.4.7.
(b): RNase A treated MS2 5'RNA reassembly
TEMs were taken of reassembled STNV CP with MS2 5'RNA ([CP] = 5 µM) prior and post RNase A treatment as described in Chapter 2.4.7.

4.3 Discussion

In Chapter 3, it was shown that the N-terminus of STNV CP is essential for capsid reassembly *in vitro*. Further evidence of this was presented in Bunka *et al*., where N-terminal truncation mutants were prepared that failed to assemble *in vivo*[81]. STNV capsid assembly is also dependent on the presence of RNA. Many viral coat proteins have extended N-terminal regions encompassing sequences rich in basic residues. It was always assumed that these N-terminal regions made non-specific electrostatic interactions with the phosphodiester backbones of their genomes. However, no details have been obtained from these interactions to date. The results in Chapters 3 and 4 have substantially increased our understanding of these interactions, and have enabled a detailed model for STNV capsid assembly to be proposed (see Figure 4.13).

In the presence of aptamer B3, STNV CP assembles into a *T*=1 capsid, although an intermediate species is also produced. RNA-induced assembly appears to be sensitive to the loop regions, since the sequence variant of B3 4U is much less efficient at triggering assembly. Although details of the sequence specific interactions are not present in the electron density map, the interaction must be an important determinant for capsid assembly, as mutation of the loop region severely impairs assembly. The loop sequences are possibly important for juxtaposing the stem loops in a way that enables the stems to lie adjacent to the N-termini helices. Presumably this orientation allows the section of polypeptide from residue 12 to become ordered by neutralising the positive charges of the N-termini at these regions. This would add to the positive charges (R14, K17 and R18) that have already been recognised in the VLP. Similar groups are present at the N-termini of STNV-2 and STNV-C CP,

implying that these positive charges are positioned in structurally relevant positions to make contacts with RNA, and that these contacts are important for capsid assembly. The fact that capsid assembly is less efficient with aptamer B3 than it is with B3 short and TR suggests that the 5' and 3' extended regions in B3 are inhibitory to assembly. This implies that assembly is sensitive to sequences outside the preferred stem loop.

The data presented in Chapters 3 and 4 show that the N-terminal region plays a specific role in capsid assembly, and that assembly efficiency is governed by RNA sequence specificity. The RNA seems to aid the assembly of protein: protein interactions by overcoming an electrostatic barrier present in the absence of RNA. In Chapter 3, it has been shown that purified STNV CP does not assemble in the absence of RNA. The state of STNV CP aggregation was analysed by svAUC under both reassembly and disassembly buffer conditions. The results from these were identical, where the overwhelming majority of this species is monomeric. Two minority species are also present in these samples, possibly corresponding to trimer and pentamer, respectively. The STNV-B3 X-ray crystal structure implies that RNA binding shifts the equilibrium towards that of CP trimeric formation, suggesting that the trimer is the basic capsomere of STNV assembly.

If RNA binding initiates formation of the trimeric capsomere, what is it that brings together each capsomere into a capsid? The answer may lie in the X-ray crystal structure of the VLP[89]. Positively charged residues from neighbouring trimers (R66, R91 and K123) are positioned in a way so their side chains can make contacts with the phosphodiester backbone of the RNA of a neighbouring

trimer. These contacts are made to both the stem region (R66) or may interact with the loop region (R91 and K123) around the three-fold axes of the VLP.

In single molecule capsid assembly studies, STNV CP compacts and assembles STNV-C RNA into *T*=1 capsids, but fails to compact MS2 RNA under the same conditions[60]. This observation suggests that STNV-C RNA has PSs at defined spacings relative to one another that are not as ideally spaced in the MS2 RNA. Other RNAs are going to contain STNV PSs by chance, but these RNAs are unlikely to contain the frequency of these PSs at specific locations with ideal spacings relative to one another. By initially contracting the genome, the virus pays an initial ‘entropic deposit’, such that this energetic cost can be recouped later by a reduction in the subsequent reduced energy barriers involved in the formation of assembling a compacted genome. The CPs bound to the compacted genome are likely to be more stable due to enhanced protein: protein interactions as well as the protein: RNA contacts compared to transient species that form on cellular RNAs.

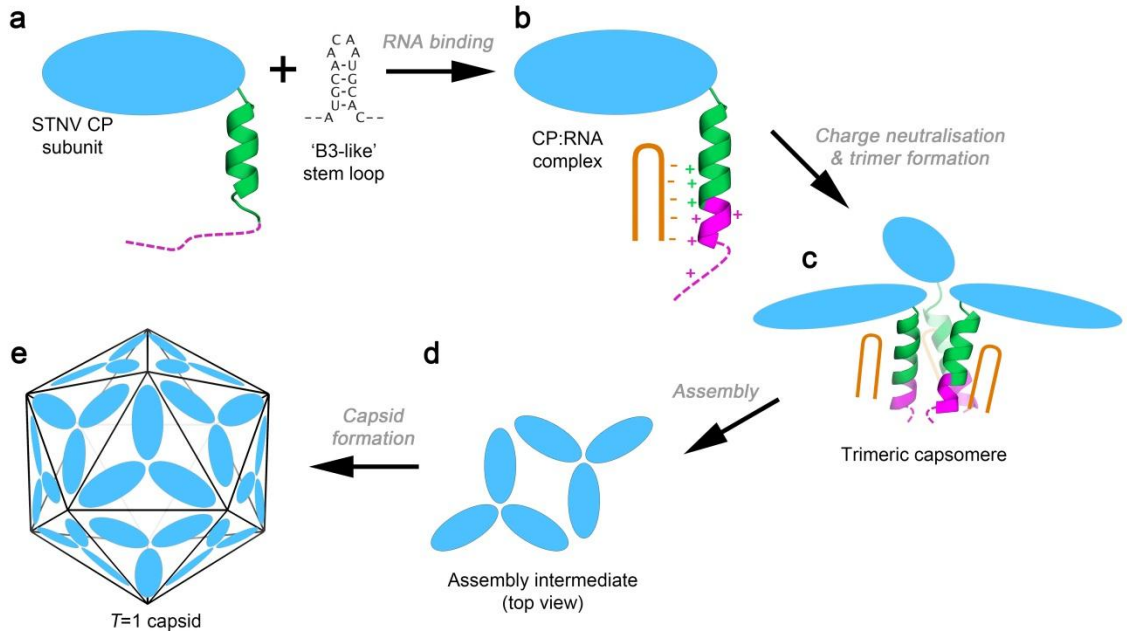


Figure 4.13: Schematic model of the STNV assembly process. (a) STNV coat protein (CP) subunits (blue oval) have an N-terminal extension that is partly helical (green) and partly unstructured (magenta). These extensions are highly positively charged with 7 Arg/Lys residues in the 24 residues at the N-terminus. (b) Binding of RNA stem loops, exemplified by the B3 stem loop, at least partly neutralises the positive charge on the N-terminus allowing the N-terminal helix to become longer and more ordered. (c) Once the N-terminal region has had its positive charge neutralised by RNA binding, the CPs can trimerise, forming an assembly-competent trimeric capsomere. (d) These trimeric capsomeres can then assemble to form higher order structures and ultimately the *T*=1 capsid (e).

5. Mechanisms for uncoating in (+) ssRNA icosahedral plant viruses

5.1 Introduction

Plant viruses can cause significant losses in terms of crop production in the horticultural and agricultural industries. For example, viral diseases of rice in Southeast Asia can cause devastating losses, and have been estimated at costing around $1,000,000,000 per annum[133]. Host defence mechanisms have recently been recognised that control the spread of infection *in planta*, including roles in RNA silencing[134]. These host defence mechanisms are of consequence to the viral life cycle. In TCV, the coat protein, p38, has been shown to have a dual functionality; both protecting the genome from environmental factors and providing protection against some of these host defences during uncoating[109].

The work described in this Chapter aimed to probe the roles of the capsid in facilitating controlled disassembly in two model (+) ssRNA icosahedral plant virus systems, TCV and recombinant STNV. The main focus of this work is on TCV, where the project was initially started by Dr Saskia Bakker, who had completed cryo-EM reconstructions of the native virion and an expanded form. I came onto the project because of my knowledge and experience with plant viruses and size exclusion chromatography. The following subsections give some background to this work prior to my involvement with the project. My contribution to this project commences from Chapter 5.2 onwards.

5.1.1 Cryo-EM structure of TCV reveals internal density which is absent in the X-ray crystal structure

A 3D image reconstruction of native TCV was performed by Dr Saskia Bakker and determined by cryo-EM using single-particle processing and icosahedral averaging at a resolution of ~11 Å. The subsequent EM density map contained the spiked-like appearance that is characteristic of the *Tombusviridae.* The electron density corresponding to the S and P domains in the crystal structure of TCV[110] were fitted to the EM map to increase the understanding of the EM structure. A cross-section of the EM map reveals that the virion is composed of three layers of density each connected at defined points on the particle. The atomic coordinates from the S and P domains in the crystal structure fit the EM density from the outermost shell extremely well (between the atomic radii 180-110 Å, see Figure 5.1), suggesting that this density relates to the S and P domains in the native particle. The two inner shells of density present at lower contour levels are therefore likely to be a mixture of N-terminal R domains from the CPs and internal RNA. The middle shell of density (between 110-82 Å) seems to be connected to the S and P domains at 5-fold, 3-fold and 2-fold symmetry axes (see Figure 5.1a). The innermost shell of density (between 72-50 Å) appears to be connected to the middle shell along the 2-fold symmetry axes. At higher contour levels, the innermost shell appears to be almost completely absent. At this contour level, the middle shell has a distinct cage-like appearance, with an apparent hexagonal and pentagonal lattice centred on the three-fold axis. The absence of the inner density at this higher contour level suggests that this region may have a lower occupancy, or is intrinsically disordered relative to the outer shells, or both.

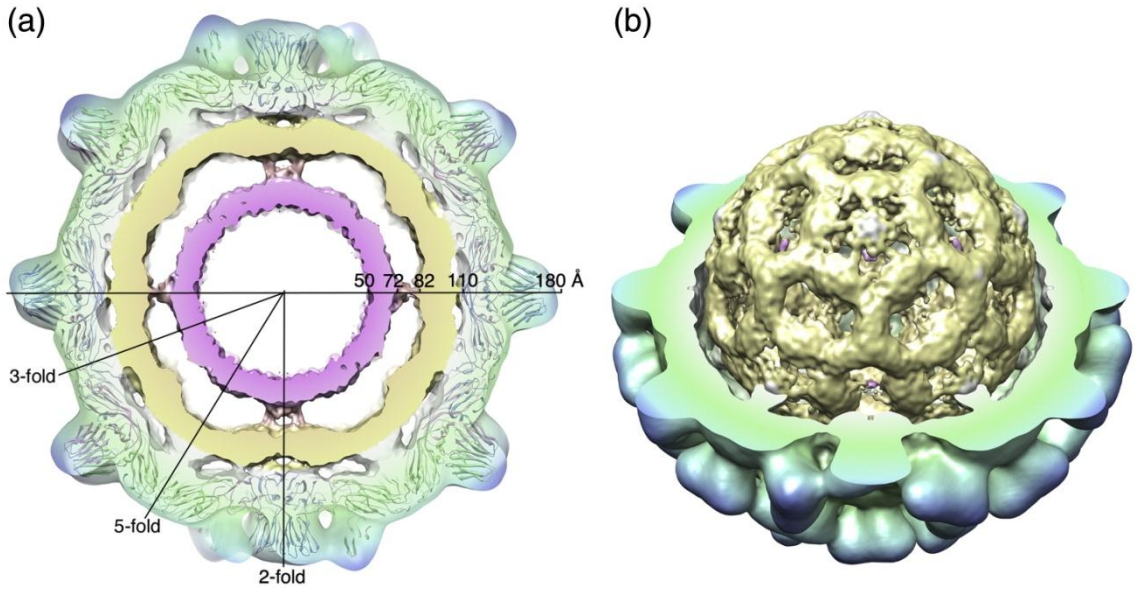


Figure 5.1: Internal density from the cryo-EM structure of native TCV
(a) Cross sectional cut-away view of native TCV
A central 25 Å cut-away cross-section from TCV EM electron density. The outermost layer is transparent to allow the modelled coordinates from the X-ray crystal structure to be fitted. The inner shells are coloured using a radial scheme from pink (50 Å) to blue (180 Å) and labelled across the virion. Icosahedral symmetry axes are also shown by lines.
(b) Cut-away version of the virion
Removal of the outermost layer of density reveals density of a cage-like structure which probably corresponds to N-terminal R domains and RNA. Same colour scheme is used as in A. Figure taken from [113].

5.1.2 Cryo-EM structure of expanded TCV reveals pores at the particles three-fold symmetry axes

TCV, like that of many plant viruses, such as TBSV[21] and SBMV[135], contains calcium ions which function to stabilise the particle. The calcium ions in the homologous TBSV are coordinated by amino acid residues Asp155 and Asp157 from one CP subunit and by Glu127 and Asp199 from the adjacent CP (see Figure 5.2b). When TCV enters the cytoplasm of a newly-infected cell, the low concentration of calcium ions are thought to initiate the removal of these calcium ions. The resulting electrostatic repulsion from side chains that coordinated these calcium ions, as well as the electrostatic repulsions from RNA, results in expansion of the TCV particle. In the crystal structure of

expanded TBSV, this expansion event is stabilised by increased ordering of the N-terminal R domains in the A/B dimers[136]. Treatment with high ionic strength results in the disassembly of expanded TCV, but not expanded TBSV[34].

In order to examine the effects of expansion in TCV by cryo-EM, Dr Saskia Bakker treated TCV with EDTA in a low ionic strength buffer at pH 8.5 to mimic the cytoplasm of the plant cell. This resulted in particles that were expanded: these particles were used to produce a 3D reconstruction of icosahedrally averaged expanded TCV at ~17 Å. The expanded particles were ~5 % larger than the native virions, measuring ~190 Å radii compared to ~180 Å in the native particles[113]. Similarly to the native virions, the expanded particles also contained three layers of density, with each residing at slightly higher radii relative to the native virions[113]. The structure from the outermost layer of the expanded particle was relatively similar in appearance to the native. In contrast, significant differences exist between the inner shells of density. In the expanded particle, the density from the middle shell becomes progressively weaker, and the cage-like structure that appears in the native structure becomes less apparent[113]. However, at the lower contour levels, visible columns of density are present that connect the innermost shell to the middle shell at the two-fold axes. There are two main reasons which may contribute to the weaker density in the expanded state. Firstly, the expanded state could affect the internal volume of the virion (a 5 % expansion results in 17 % larger intrinsic volume). Secondly, RNA and/or N-terminal R domains may be more dynamic in the expanded state relative to the native state. This would make sense if the genome of TCV is preparing for exiting the particle[113].

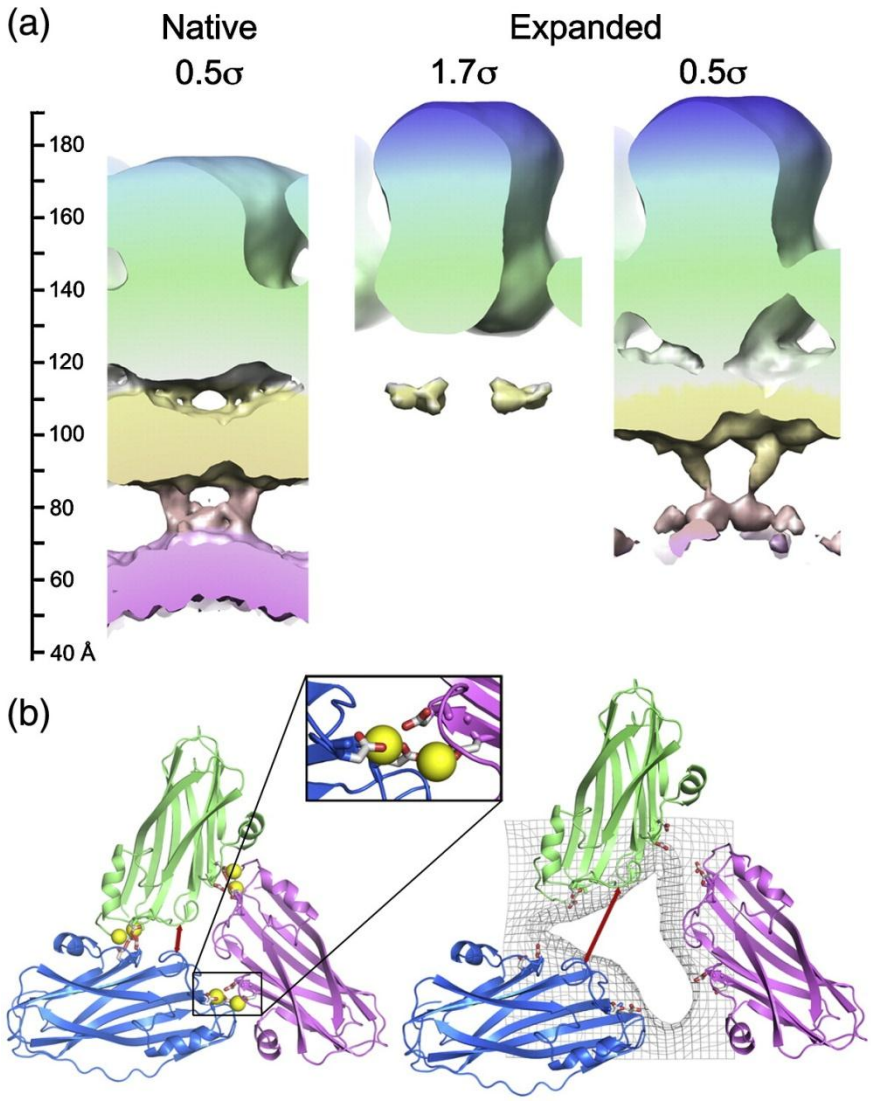


Figure 5.2: Structural comparison between native and expanded states
(a) Radial distribution of states
On the left, the native virion is contoured so that each shell of density is shown. In the middle panel, the expanded particle is shown to show the same internal volume to native virions. The right panel shows density from the expanded particle to at a higher contour level to show density for the weaker inner shells of density. The ruler on the left corresponds to the radial distance from the centre of the particles.
(b) Cartoon representation of the results of expansion.
Only the asymmetric trimers of the particles are shown for clarity. On the left, the native trimer contains calcium ions (shown as yellow spheres) at the interface between CPs in the locations they are shown in the homologous X-ray crystal structure of TBSV (PDB code 2TBV). The calcium ions are coordinated by acidic side chains and shown inset (D155 and D157 from one subunit; E127 and D199 from the other). The pores shown in the cryo-EM in the expanded virus are shown being surrounded by a grey mesh. The effect of expansion of two amino acids which remain at the same radial level (D203 and S202) are separated by a red arrow. The distance between them in the native particle is 6.5 Å compared to 17 Å in the expanded particle. The figure was taken from Bakker *et al.*[113]

There is also a big difference in capsid integrity between the expanded particle and the native virus (see Figure 5.3). The expanded particles appear to have developed pores at distinct sites on the protein capsid, whereas the native capsid does not. To better understand the nature of expansion, the coordinates from the S and P domains from the crystal structure were fitted in to the expanded density. This showed that there was no discernible change in the hinge angles between the S and P domains**,** and that pentamers and hexamers that build the capsid remained intact[113]. The pores occur due to the pentamers and hexamers moving apart by about 10 Å. In the expanded crystal structure of TBSV, the β-annuli are intact[136]. However, the resolution from these cryo-EM maps do not allow direct visualisation of β-annuli, therefore it cannot be confirmed if the β-annuli are intact in the expanded state of TCV[113].

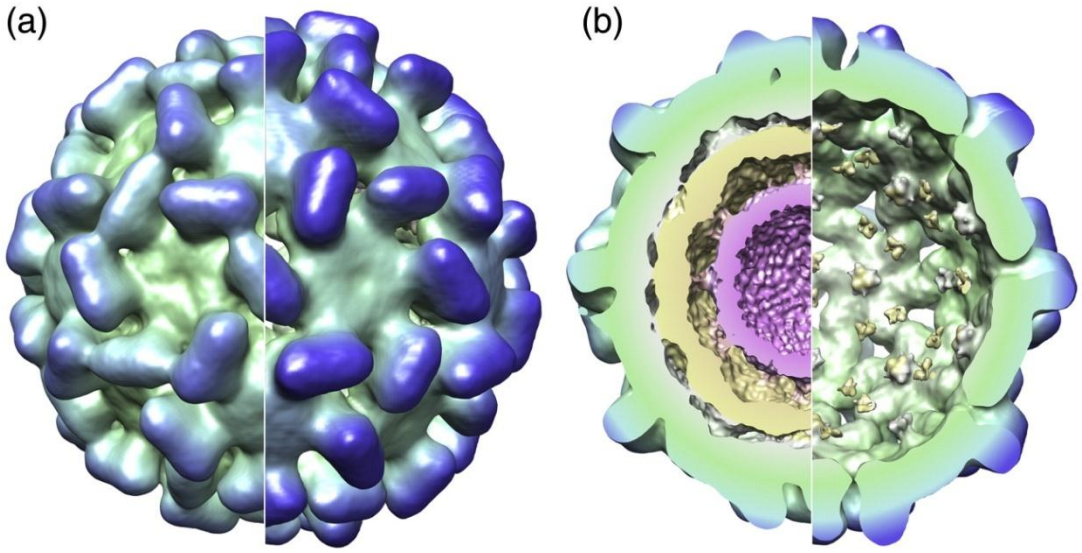


Figure 5.3: Comparison of native and expanded TCV.
(a) Surface representation of native and expanded TCV.
Native virion is on the left and expanded TCV is on the right. The structures are coloured using identical radial colour schemes, so that the expanded particles are shown as a deeper blue compared to their native counterparts. Both models have been contoured to show approximately the same density for their protein capsids.
(b) Inside view of the back half of native and expanded TCV
The native is on the left and the expanded particle is on the right. Native particles have clearly more density on the interior of the capsid compared to the expanded form. Figure taken from Bakker *et al*. [113]

5.1.3 Biochemical analysis of the different states of TCV reveals differences in stability and protease susceptibility, but not RNase susceptibility

Electron density from the cryo-EM structure of the expanded particle described above contains pores. However, the precise size of these pores cannot be determined due to the limits of the resolution on this map. One obvious function of these pores could to mediate the extrusion of molecules from inside the particle to outside, such as their N-terminal R-domains and/or their RNA genome. To assay this possibility, Dr Saskia Bakker performed protease and nuclease digestion on each state of the TCV virion to determine the sensitivity of each of these states to digestion.

Native, disassembled, expanded and recontracted forms of TCV were challenged against chymotrypsin over time and the digestion products were analysed by SDS-PAGE. Native chymotrypsin has a characteristic banding pattern of two major protein species; one band of ~80 kDa, which corresponds to the p80 protein and one product of ~38 kDa, which corresponds to the full length TCV CP. Also, there are 2 minor species of ~30 kDa (see Figure 5.4a). Challenge of native TCV against chymotrypsin reveals no change in the digestion pattern, even at the longest time point, suggesting that native TCV is resistant to proteases. This re-emphasises the great stability of TCV.

In contrast, when disassembled TCV was subjected to chymotrypsin challenge over time, the SDS-PAGE digestion products showed that full-length TCV CP got rapidly degraded into products ~30 kDa (see Figure 5.4a; black arrow). These products were analysed by mass spectrometry and N-terminal

sequencing. This analysis revealed that these digestion products are consistent with the removal of the R domains, confirming what was suggested previously.

Chymotrypsin treatment of expanded TCV particles leads to the same 30 kDa digestion products, albeit with slower kinetics. However, a significant proportion of this material was undigested, suggesting that a subset of the TCV CP remained protease resistant. The approximate ratios of these products on the gel suggest that ~1/3 TCV CP were undigested, suggesting that only the A/B conformers may be protease sensitive. Since the pores in expanded TCV are too small to allow proteases to enter the particle directly, the protease digestion sites for the C/C dimers are likely to be inaccessible.

Expanded particles were recontracted by the addition of excess calcium ions and also challenged against chymotrypsin. From these particles, minimal proteolysis was observed, suggesting that the expansion of TCV was largely reversible. Cryo-EM maps generated against recontracted TCV particles produced identical maps to that of the native particles[113].

As well as protease challenge against the different states of TCV, similar analysis was also carried out by RNase A. As expected, the RNA from native virus was protected against RNase A treatment. Interestingly, considering that expanded TCV is sensitive to protease treatment, the expanded particles showed resistance to RNase A. This was not simply a consequence of the buffering conditions used to expand the virus, as the RNA was rapidly degraded by RNase A in the disassembled virion (see Figure 5.4b).

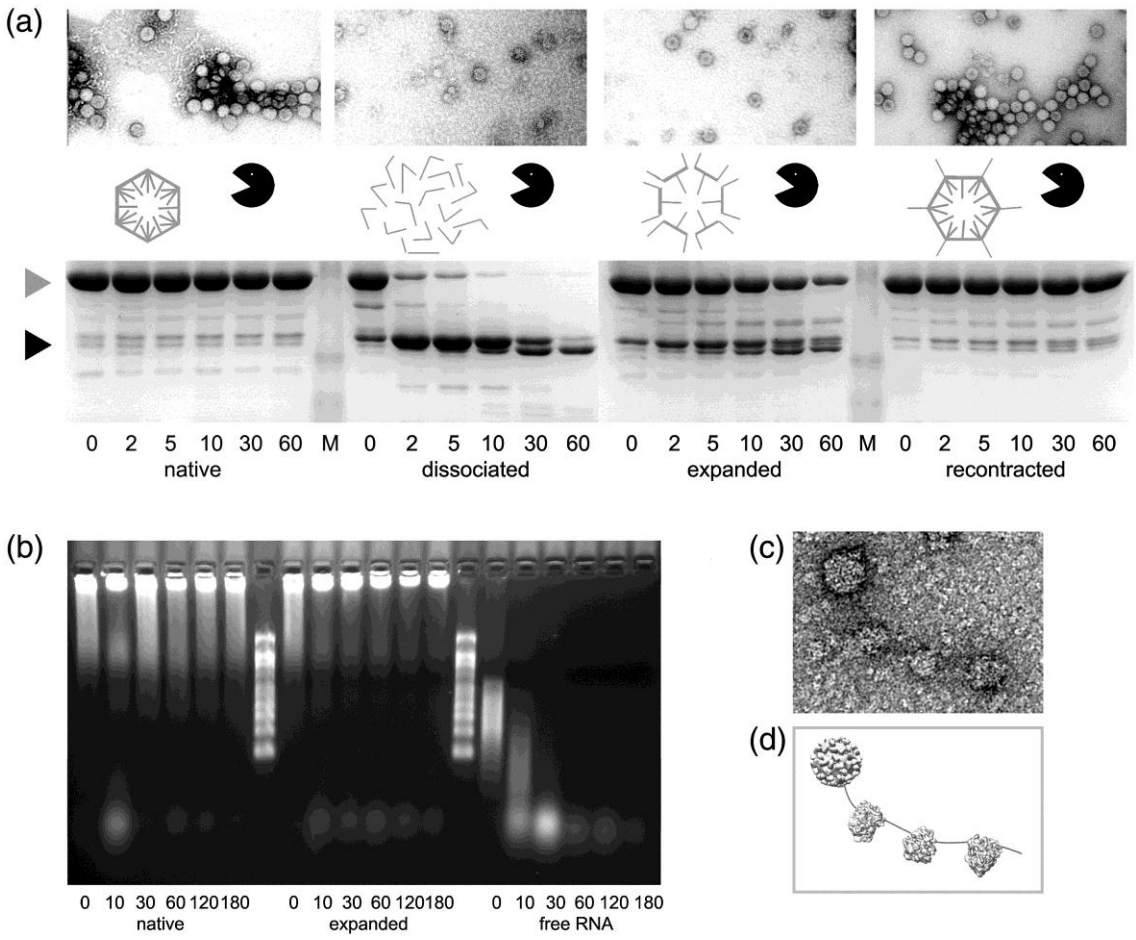


Figure 5.4: Biochemical analysis of the differing states of TCV
(a) Chymotrypsin digestion assay
TCV was challenged against chymotrypsin in each state and its digestion products analysed by SDS-PAGE. The native particles appeared to be unaffected by protease treatment, whereas the disassembled particles were rapidly digested to the ~30 kDa digestion products, which roughly correspond to the CP S and P domains. After 60 mins, the expanded particles have roughly 1/3 of CP as the full-length product. In recontracted virions, chymotrypsin does not appear to cleave the CP, implying that recontraction is largely reversible. Negative stain TEM images are shown above each prior to addition of chymotrypsin. Grey arrow points to full-length CP and black arrow to the 30 kDa digestion products.
(b) RNase A assay of native and expanded TCV
Denaturing agarose gels of RNase A treated samples from native (left), expanded (centre) and free RNA (right). Time points are shown at the bottom of the gels in mins.
(c) TEM of TCV incubated in wheat germ translational extract
TEM shows a TCV particle attached by a thin tether of what looks like RNA to 3 ribosomes.
(d) Schematic of TCV striposome complex
Cartoon based on known 3D structures of what is shown in (c). Figure taken from Bakker *et al.* [113]

Despite the insensitivity of expanded particles to RNase A digestion, when expanded TCV was incubated in wheat germ translational extract and examined by negative stain TEM, particles were seen extruding a thin tether, which is presumed to be RNA, connected by particles which appear to be ribosomes (see Figure 5.4c). This is fully consistent with a co-translational disassembly mechanism mediated by ribosomes. Such striposome formation has also been observed in other plant viruses[99], most notably with TMV[101].

5.1.4 svAUC on the different states of TCV reveal differences in sedimentation coefficients

To determine the stability of the expanded state of TCV, and to characterise the differences which occur during expansion, a series of samples were prepared for svAUC, which was performed by Amy Barker (see Figure 5.5). The $S_{20,w}$ of native TCV was calculated at ~120 S. When expanded TCV was spun at high velocity, the expanded virion seemed to dissociate into smaller constituents, although here was no overlap with fully disassembled TCV. This shows that expanded particles are not as stable as native TCV.

TCV particles were expanded by EDTA chelation and recontracted by addition of excess calcium ions. In contrast to the expanded particles, recontracted particles sediment as though they are VLPs, but having lost some lower molecular weight components. This result contradicts the previous observation from the low resolution cryo-EM reconstruction. This structure suggests that recontracted virus resembles the native particle, and that the expansion state is fully reversible (data not shown). The cryo-EM reconstruction was determined at a relatively low resolution and also employed icosahedral averaging. The

implication is that the cryo-EM reconstruction may not distinguish between any non-asymmetric features seen in the recontracted state. Expanded, chymotrypsin treated and recontracted virus sediments similarly to recontracted virus, suggesting that the recontracted virus has lost some lower molecular weight material.

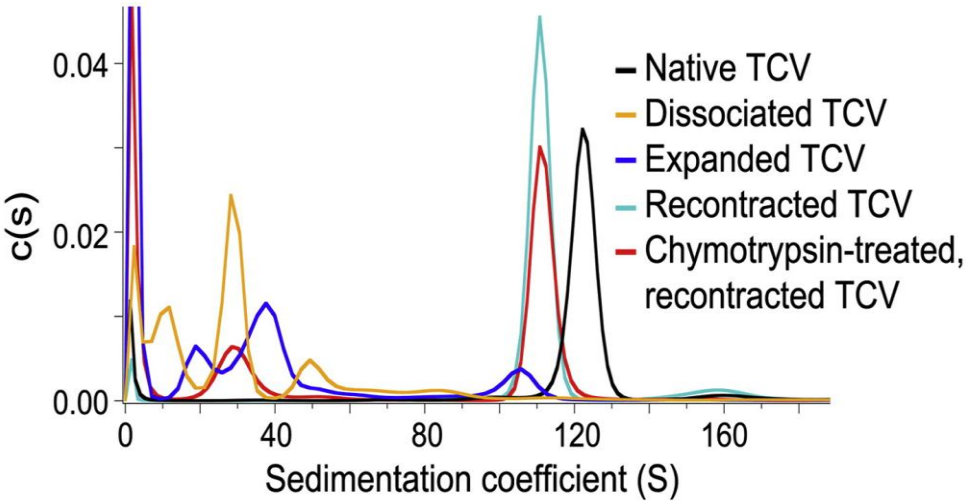


Figure 5.5: svAUC analysis of the different states of TCV
The key (inset) shows the samples analysed by svAUC. Native TCV sediments ~125 S. Both expanded and disassembled TCV show heterogeneous peaks corresponding to fractionated particles, although these peaks do not overlap. This suggests expanded TCV is metastable, and falls apart under the stress of centrifugation. Recontracted and chymotrypsin treated and recontracted TCV both sediment at similar S values, suggesting that recontraction of TCV is not fully reversible. Figure taken from Bakker *et al.* [113]

The results in the next subsection describe the point in time where I came onto the project and discuss my contribution to this work.

5.2 Results

5.2.1 Expanded, chymotrypsin treated TCV leads to partial extrusion of viral RNA

The results from the proteolysis experiments described in Chapter 5.1.3 might arise in two different ways. When particles are expanded and treated with chymotrypsin, perhaps a subset of these particles are completely proteolysed, or it may represent the average proteolysis from all the particles in the sample. To analyse which of these scenarios is most likely, expanded chymotrypsin treated TCV was prepared so that roughly 50 % of the subunits were cleaved. The chymotrypsin was then quenched with excess protease inhibitors as described in Chapter 2.5.2. This expanded proteolysed sample was then fractionated by size exclusion chromatography. The fractions with UV absorbance were collected, and 500 µL from these fractions were concentrated by Amicon Ultra® centrifugal concentrators and analysed by SDS-PAGE. Note, the concentrated particles were possibly disrupted by centrifugation, but the protein content was analysed rather than the particles themselves.

The fractionation profile shows that a VLP peak is eluted between 8-11 mL which also absorbs strongly at 260 nm, suggesting the RNA is retained within VLPs. This peak has significantly less proteolysed CP compared to the material that was loaded (see Figure 5.6). Fractions that eluted between 13-17 mL contained most of the proteolysed material, and had the characteristic banding patterns of TCV CP with the R domains cleaved on SDS-PAGE. Another minor peak was eluted between 17-19 mL, which seems to show smaller fragments of protein, probably the N-terminal R domains and/or separated S and P domains.

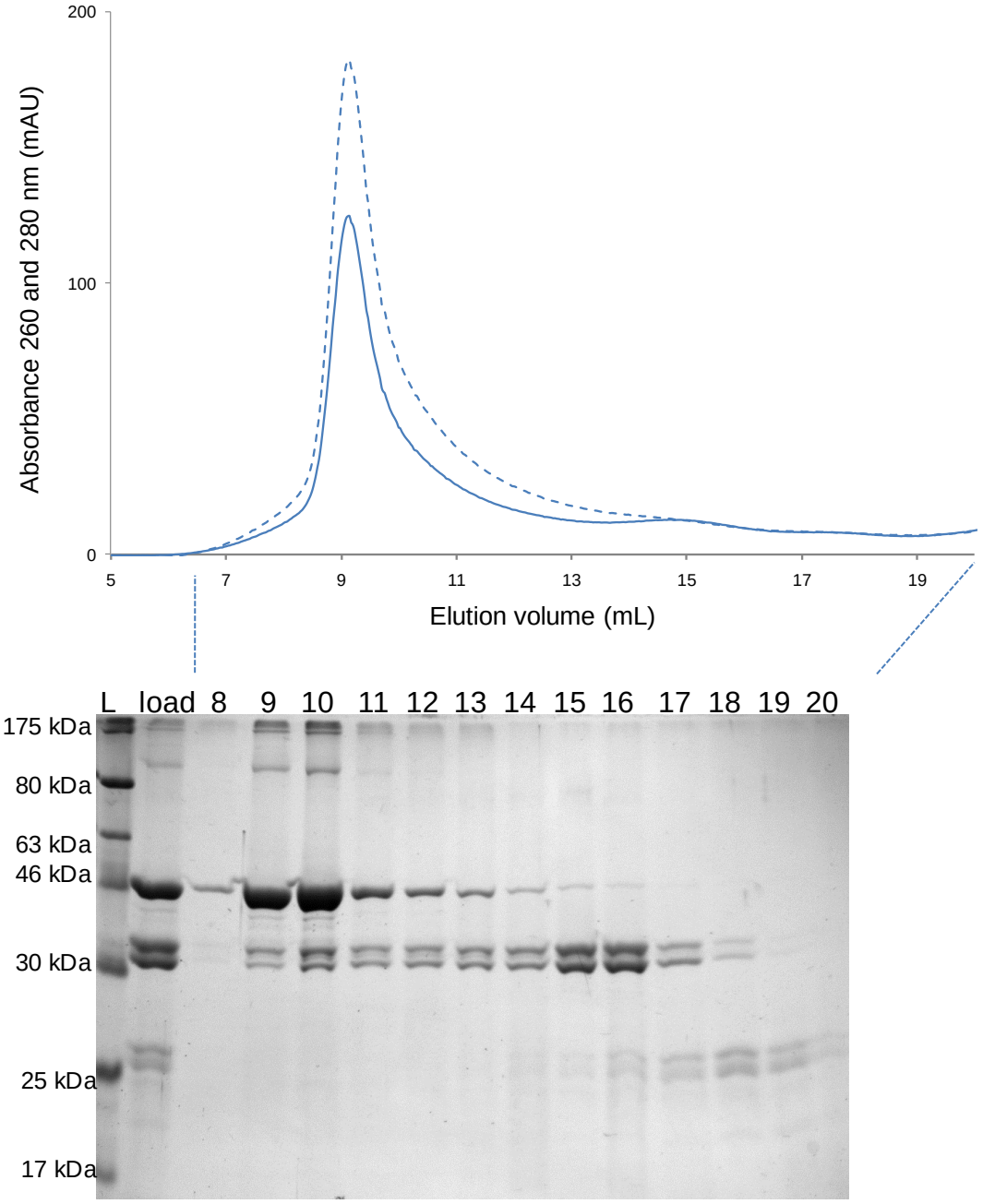


Figure 5.6: Chromatogram of expanded, chymotrypsin treated TCV and SDS-PAGE analysis of fractions

The top of the figure shows a cropped version of a size exclusion chromatogram looking from 5 mL to 20 mL to show the UV absorbance from expanded proteolysed TCV. The dashed lines on the chromatogram represent A_{260}, whereas the solid lines show the A_{280}. The dotted blue lines connecting the chromatogram and the SDS-PAGE indicate the fractions analysed by SDS-PAGE. The exact fraction numbers are shown above the lanes of the SDS-PAGE. L = 7-176 kDa prestained protein ladder (NEB); load = sample prior to size exclusion analysis.

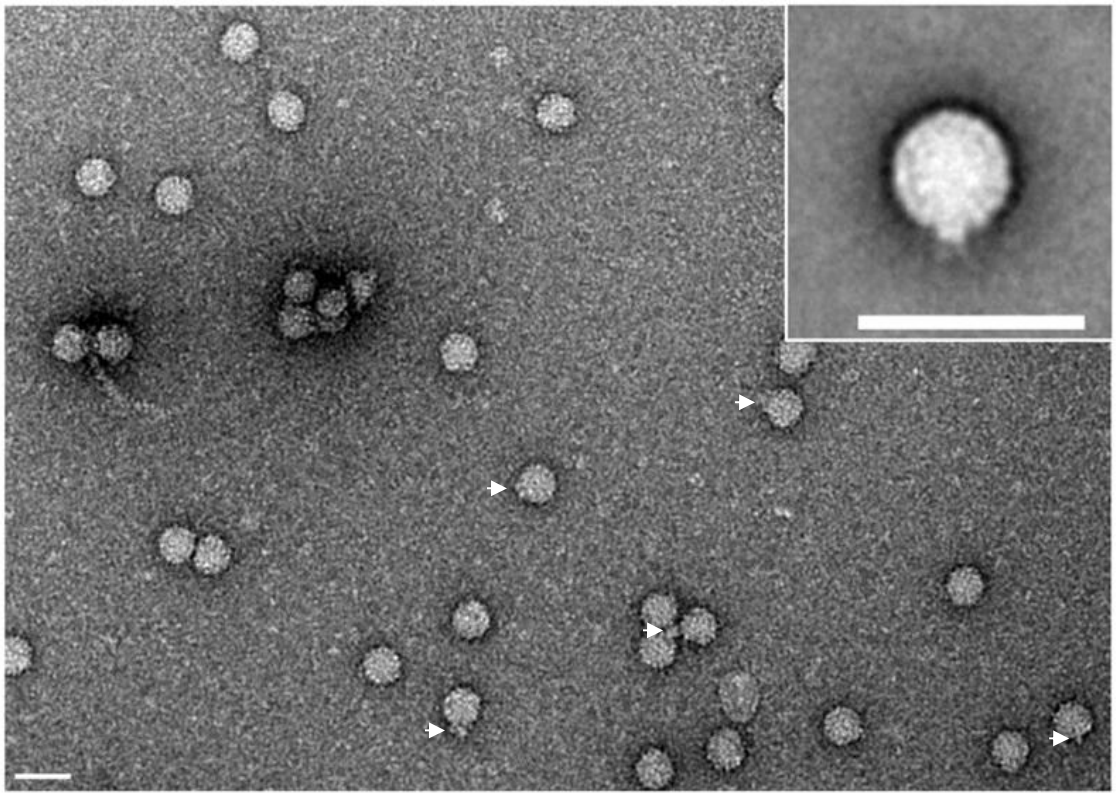


Figure 5.7: Negative stain EM image of expanded and proteolysed TCV
This image shows the heterogeneity and disruption to the particles. Some of these particles have what seem to be their genome and/or N-termini poking out at a defined location (white arrow heads). Inset is an aligned average of this type of particle, calculated by multivariate statistical classification. The scale bars, both in the main panel and the inset, represent 50 nm. Figure adapted from Bakker *et al*[113].

These results are consistent with the dissociation of cleaved subunits from intact particles. Presumably, the A/B dimers would be released by proteolysis because the β-annuli are required for particle integrity. It could be imagined that this dissociation of subunits may arise in stages, where N-termini may become externalised and accessible to protease. Subsequent cleavage results in the dissociation of the cleaved CP dimer, thus making the VLP even more porous than the expanded particle.

Negatively stained expanded and proteolysed particles were observed by TEM (see Figure 5.7). These images are enlightening. Although expanded TCV

particles are icosahedrally symmetric nature, expanded and proteolysed VLPs contained a mixed population of particles that appear to have lost their symmetry. Some of these images are consistent with extrusion of the viral genomes at defined sites from the VLP, although the level of extrusion appears to vary between these particles. However, it seems that most of these particles are significantly disrupted, or at least disrupted to some degree (see Figure 5.7 and inset).

5.2.2 Incubation of expanded, chymotrypsin treated TCV incubated with wheat germ extract provides evidence of striposome formation

In order to test whether expanded proteolysed particles are potential uncoating intermediates, these particles were incubated in the presence of wheat germ extract according to the manufacturer's guidelines. Emma Wroblewski, an undergraduate project student, and I prepared size exclusion chromatography-purified expanded, chymotrypsin treated TCV using the conditions described in Chapter 2.5.2. These particles were then incubated in the presence of wheat germ extract for 30 mins at room temperature following the manufacturer's guidelines (Promega). The sample was then diluted 50-fold and prepared for TEM as described in Chapter 2.2.2. The TEMs showed that striposomes could be formed under these conditions in high yield (see Figure 5.8).

Previously, Dr Saskia Bakker incubated expanded TCV in the presence of wheat germ extract *in vitro* translation kit (Promega). However, no product was detected in these experiments that related to a possible translation product. In order to further characterise TCV disassembly, expanded proteolysed TCV was incubated in the presence of wheat germ extract *in vitro* translation kit

(Promega) using manufacturer's guidelines. However, again, no translation products were detected.

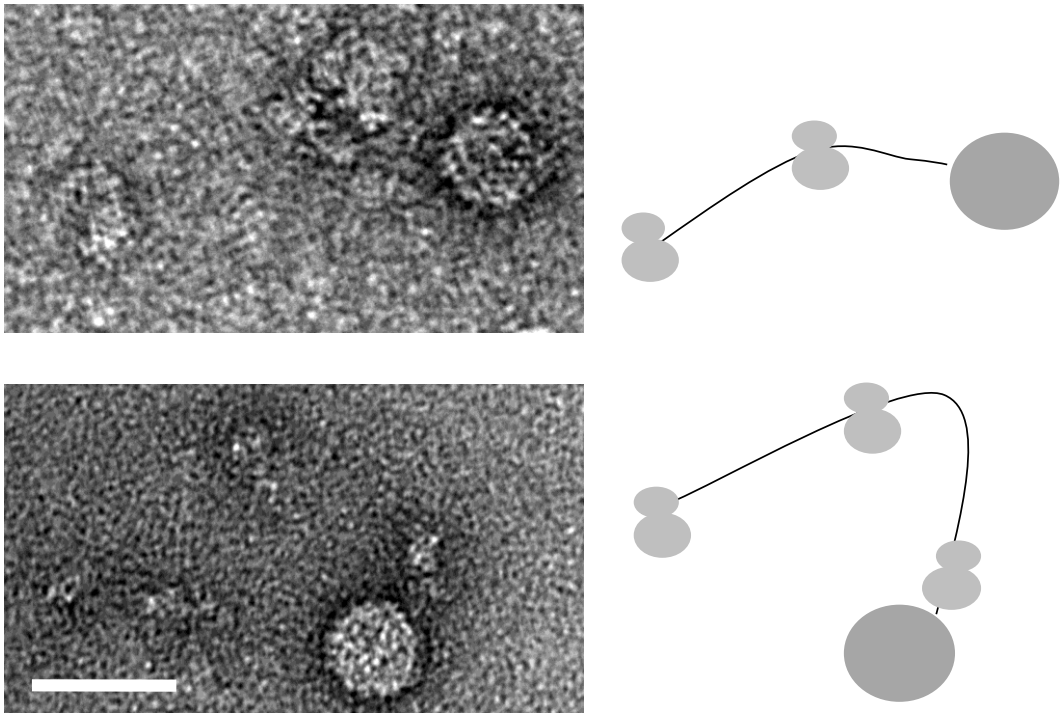


Figure 5.8: Gallery of Striposomes
Striposomes were isolated from expanded TCV treated with chymotrypsin. These striposomes have either 2 (top panel) or 3 (bottom panel) ribosomes bound via a thin tether of RNA. Scale is identical in each micrograph. (Scale bar = 50 nm)

5.2.3 Expanded and proteolysed TCV particles are resistant to RNase A degradation

It was shown in Chapter 5.1.3 that expanded particles are resistant to RNase A treatment, despite the RNA genome being capable of exiting from pores in the particles and forming striposomes. Because the RNA genome appears to extrude from expanded, chymotrypsin treated particles in the EM images in Figure 5.8, RNase A treatment was performed by Emma Wroblewski under my supervision. These RNase A-treated particles were incubated for 1 h at room temperature before the RNase A was quenched through addition of 1 µL RNaseOUT^{TM}, and the RNA purified as specified in Chapter 2.5.4. Remarkably, expanded chymotrypsin treated particles seemed to be resistant to RNase A

treatment. These results also agreed with Dr Saskia Bakker's results (Figure 5.5), which suggest that expanded TCV is resistant to RNase A, but free RNA is readily degraded.

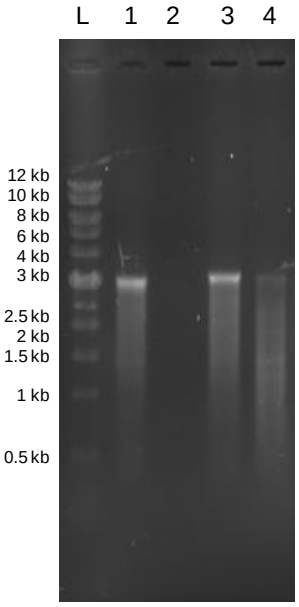


Figure 5.9: 1 % (w/v) denaturing agarose gel electrophoresis of RNase A treated samples
TCV samples were prepared and treated with RNase A. The RNA was then extracted as described in Chapter 2.5.4. Expanded TCV, RNase A treated TCV RNA and untreated TCV RNA were prepared as controls. 1 = Expanded TCV and RNase A treated; 2 = RNase A treated TCV RNA; 3 = RNase A treated expanded, proteolysed particles; 4 = untreated TCV RNA control; L = 1kb DNA ladder (NEB).

5.2.4 Chymotrypsin treatment of RP-complex suggests p80 is protected from cleavage by the RNA genome

Despite being extensively studied, many features of the p80 protein have so far remained a mystery. The elusive nature of p80 is partly due to its scarcity; the stoichiometry being calculated as 1 molecule of p80 protein: 89 CP_2 per TCV particle[34]. There is no open reading frame which codes for a product of 80 kDa in the TCV genome, suggesting that it is made by an as yet uncharacterised method. Stockley *et al* extensively characterised the p80 protein and found that it contained an identical amino acid composition within experimental error to

TCV CP. Tryptic digest and mass spectrometry analysis confirmed that p80 was a covalently linked CP dimer[112]. The exact nature of this cross-link is unknown, but it has been speculated that its formation may be catalysed by the TCV genome, based on the fact that RNA has been shown to catalyse a vast array of chemical reactions. The exact function of p80 is unknown, but putative role(s) in uncoating have been mooted as possible function(s)[34].

Despite this paucity of information, the p80 protein is enriched in the RP-complex[111]. Because of the unique nature of p80, and the p80 analogues found in other plant viruses such as SBMV and TBSV, the RP-complex was chymotrypsin treated. Its susceptibility to proteolysis and its migration by size exclusion chromatography was analysed. The eluted fractions with UV absorbance were then analysed by SDS-PAGE. It was hoped that this experiment would provide useful information regarding the nature of the interaction between p80 and the RNA genome within the RP-complex and/or the virion.

In order to carry out this experiment, copious amounts of RP-complex were needed to be able to detect the p80 protein. In light of this, 200 µL of TCV (75 mg/mL) was disassembled in the presence of protease inhibitors and the RP-complex was purified by size exclusion chromatography in an analogous manner to Sorger *et al* [34] (data not shown). The purified RP-complex was then concentrated by Amicon Ultra® centrifugal concentrators and chymotrypsin treated as described in Chapter 2.5.5. The digestion was quenched after 30 mins, and the digestion products were fractionated by size exclusion chromatography.

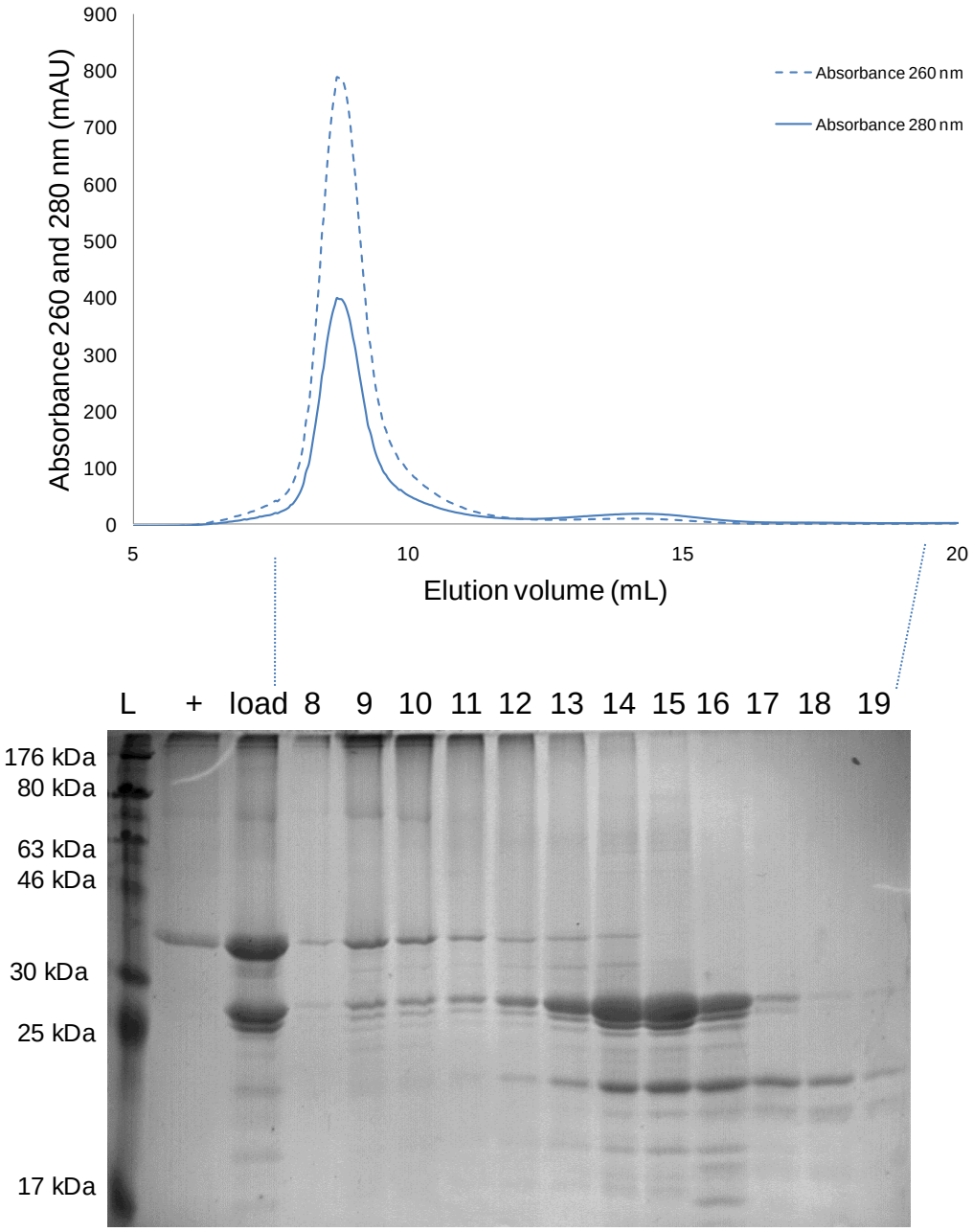


Figure 5.10: Size exclusion chromatography fractionation and SDS-PAGE analysis of chymotrypsin digested RP-complex
The top of the figure shows a chromatogram of proteolysed RP-complex monitoring the absorbance at 260 nm (dashed line) and 280 nm (solid line). The dotted lines connecting the gel and the chromatogram show how the gel corresponds to the gel. Fraction numbers are also indicated above the lanes of the gel. Load = sample prior to fractionation; + = RP-complex prior to concentration and chymotrypsin treatment; L = 7-176 kDa prestained protein marker (NEB).

The chromatogram produced two peaks; a major peak eluted between 8-11 mL, and absorbed strongly at 260 nm, suggesting this species was predominantly RNA (see Figure 5.10). The second peak seemed broader, and eluted between ~12-17 mL. This peak had a higher absorption at 280 nm than 260 nm, suggesting that digested protein was separated from the rest of the RP-complex.

The fractions collected from size exclusion with UV absorbance were then concentrated using Amicon Ultra® centrifugal concentrators. The concentrated fractions were then analysed by SDS-PAGE as described in Chapter 2.2.1.2. SDS-PAGE analysis of these fractions showed that the overwhelming majority of TCV CP_2 was proteolysed, forming digestion products with similar mobility to those seen in the expanded virion, suggesting similar cleavage sites are accessible. These results also suggest that proteolysis mediates dissociation of TCV CP_2 from the RP-complex. In contrast, the majority of the p80 protein remained uncleaved, and eluted with the RNA.

In all, these results show that p80 is much more resistant to proteolysis than TCV CP_2 in the RP-complex, and that once cleaved, TCV CP_2 dissociates from the RNA genome. Furthermore, p80 protein remains undigested, and is eluted with the RNA genome, implying that it is somewhat protected against protease digestion by the RNA genome. But what is the nature of this protection in the context of how it interacts with the TCV genome? It is likely that p80 binds to a unique site on the TCV genome. It is also possible that p80 has two RNA binding sites and that these are located at opposite ends of the genome. The p80 complex would therefore essentially circularise the RNA. This is analogous

to what happens in MS2, where the A protein (which is essential for MS2 infection) binds to sites close to each end of the genome, circularising it[137].

5.2.5 STNV VLPs may have a similar disassembly mechanism to TCV

It has long been hypothesised that calcium release and subsequent expansion mediates the first stage of uncoating in the (+)ssRNA icosahedral plant viruses[82]. However, the results described here suggest that proteolysis of the N-terminal R domain is a putative second stage in the uncoating mechanism in TCV. Other viruses also contain highly basic N-terminal regions[21; 78; 79; 135; 138; 139], and so it is possible that they share this strategy for uncoating. Our recombinant version of one of these viruses, STNV, was therefore subjected to the same EDTA and chymotrypsin treatment as described with TCV. It was hypothesised that removal of the N-termini by proteolysis following expansion may mediate the formation of an uncoating intermediate in an analogous way to TCV uncoating.

STNV virions are very stable, and have been shown to be resistant to protease cleavage[80]. In order to test whether recombinant STNV VLPs were also protease resistant, STNV VLPs were incubated in the presence of chymotrypsin (Roche) as described in Chapter 2.5.6. SDS-PAGE analysis showed that STNV VLPs were resistant to chymotrypsin treatment, analogous to the wild-type virion (see Figure 5.11).

In contrast, STNV VLPs treated with EDTA and chymotrypsin as described in Chapter 2.5.7 resulted in proteolysis of STNV CP. The digestion products seen by SDS-PAGE were very similar to what was observed in Unge *et al*,

suggesting that the N-terminal domain is proteolysed under these conditions (see Appendix, Figure 9.10). This, taken together with the already known observation that the N-terminus of STNV CP is extremely proteolytically sensitive, makes sense if a similar uncoating mechanism to TCV exists in STNV.

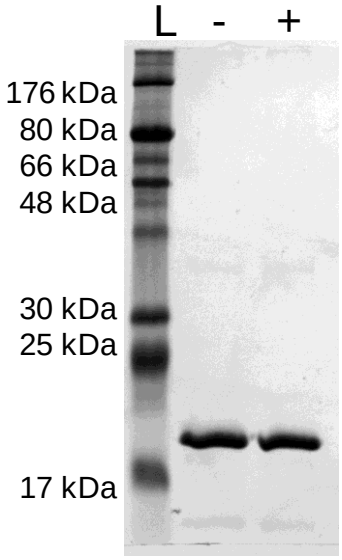


Figure 5.11: Chymotrypsin treatment of native STNV
SDS-PAGE above shows that native STNV is resistant to chymotrypsin treatment. - = STNV without chymotrypsin; + = STNV VLPs treated with chymotrypsin; L = 7-176 kDa prestained protein marker (NEB).

In order to test the hypothesis of proteolysis-mediated uncoating, EDTA and chymotrypsin treated STNV VLPs were subjected to size exclusion chromatography and SDS-PAGE, akin to expanded proteolysed TCV. The results were striking. Expanded, proteolysed STNV eluted as VLPs between ~8-12 mL, similarly to expanded, proteolysed TCV. The proteolysed subunits were also dissociated from these particles (see Figure 5.12). When these purified proteolysed particles were analysed by TEM, the particles appeared asymmetric with extrusions emanating from the STNV particles (see Figure 5.13). These results with a plant satellite virus are also fully consistent with a

proteolysis-mediated mode of uncoating, and with the TCV data suggest it may be a common feature of these viruses.

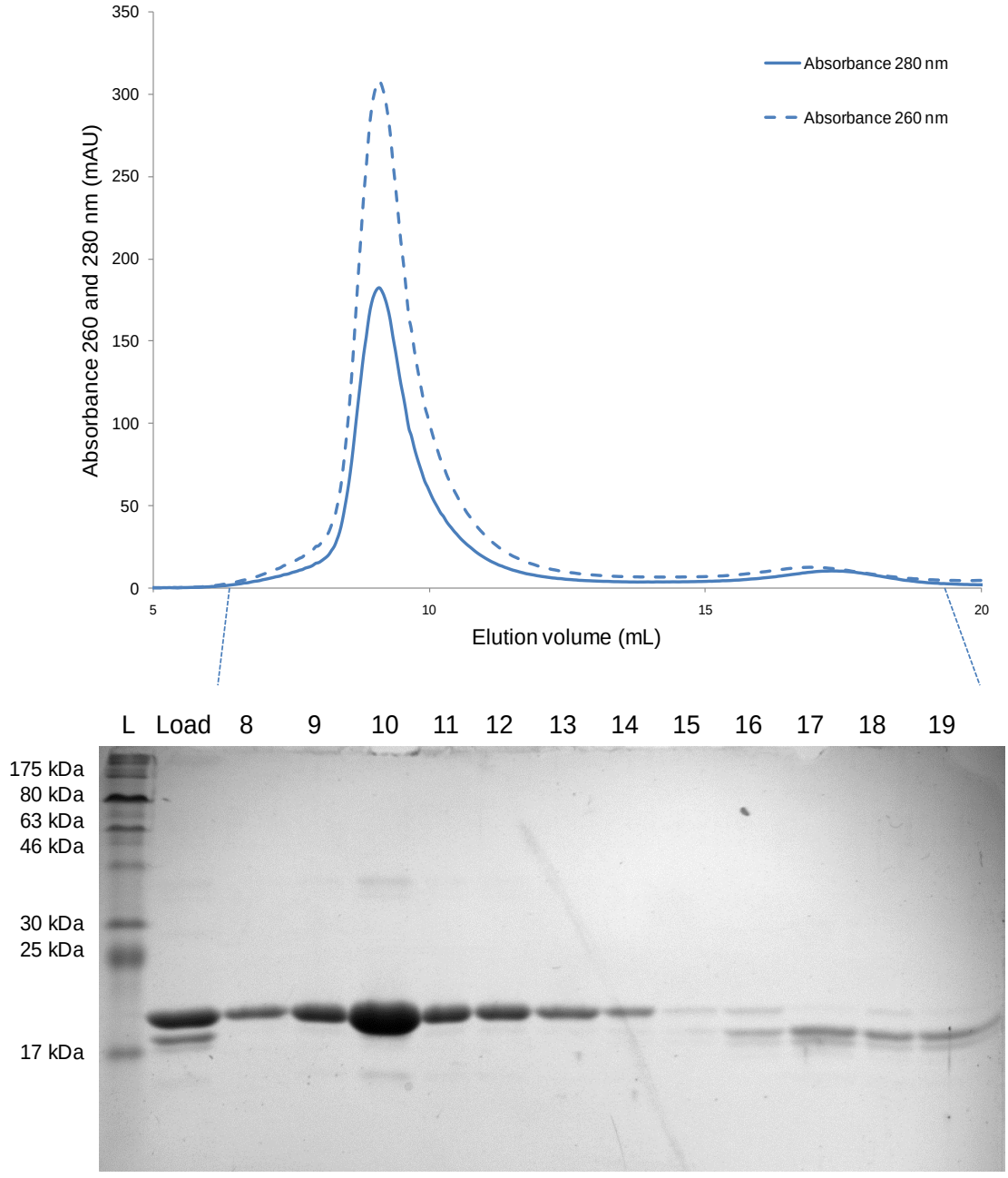


Figure 5.12: Size exclusion chromatography profile and SDS-PAGE analysis of EDTA chymotrypsin treated STNV VLPs elution
The above chromatogram between 5-20 mL shows the UV absorbance (A_{280} = solid line; A_{260} dashed line) profile from EDTA and chymotrypsin treated STNV VLPs and the fractions analysed by SDS-PAGE. Dotted lines connect the fractions to the gel, as well as the fraction numbers being explicitly labelled above the lanes. Load = sample prior to being loaded onto the size exclusion column; L = prestained protein marker (NEB).

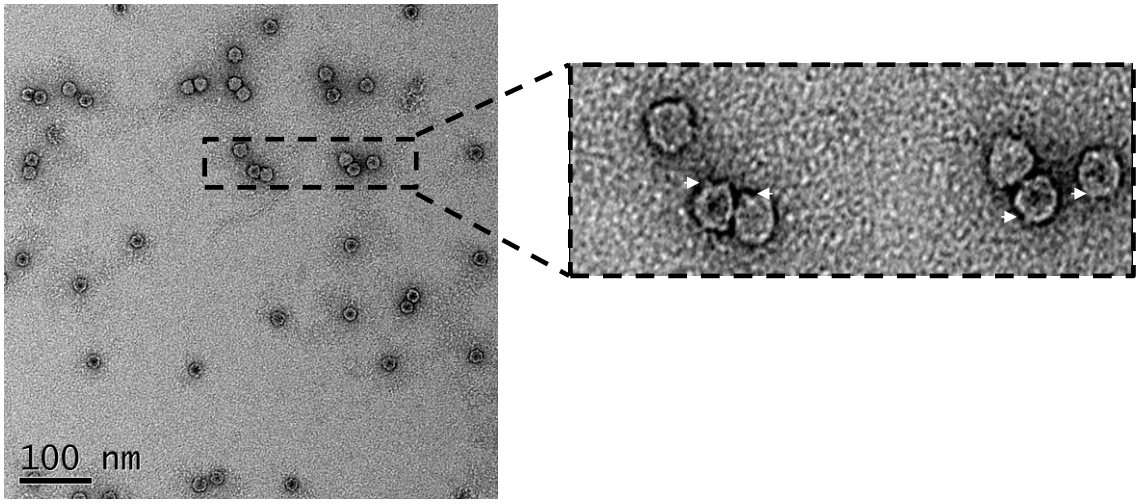


Figure 5.13: TEM images of expanded proteolysed STNV VLPs show similar extrusions to TCV
Purified EDTA and chymotrypsin treated STNV VLPs were analysed by TEM at 30,000 X magnification. A portion of the micrograph is blown up and shown alongside it, showing similar extrusions that were also present in the TCV VLPs.

5.3 Discussion

The tailed DNA bacteriophage are the most abundant entities in the biosphere[140]. Upon binding of the phage to a host cell membrane, the free energy stored in the electrostatic repulsions between the DNA phosphate groups results in the DNA exiting the capsid via the protective tail.

Although some (+) ssRNA icosahedral viruses have similar packaging densities to the tailed bacteriophages, the repulsive nature is reduced due to the existence of positive side chains at the N-or C-termini of these viruses, thus providing a neutralising effect. Also ssRNA genomes are extensively base paired and thus form branched polymers in contrast to the linear polymer of dsDNA. This avoids the problems DNA phages have in overcoming the persistence length of their genomes during the packaging event. Unlike tailed DNA bacteriophage, *in vitro* capsid assembly and packaging of their RNA genomes do not require the input of free energy, probably due to the neutralisation functions of these positively charged regions and the increased flexibility of RNA relative to DNA. If this is the case, disassembly and uncoating must destabilise the native capsids in such a way to mediate the release of their genome, i.e. work needs to be done to release these genomes whereas the pressure inside dsDNA phages provides the necessary impetus for release in those particles. In TCV and STNV, and also other plant viruses, the required capsid instability is driven by chelation of calcium ions from the capsids, which mediates expansion events. The expanded conformation of these virus particles creates pores on the surface of the previously sealed container, theoretically large enough to permit the exit of the viral RNA. In TCV, this expanded conformation is not very stable, dissociating by the addition of high salt

concentration[34] or in response to high speed centrifugation[113]. Uncontrolled dissociation of particles *in planta* would expose the RNAs to the host defences such as the plant RNAi silencing pathway.

It is from this expanded state that host cell factors must interact with the TCV genome. The reduced ordering seen in the expanded particle relative to native TCV suggests that viral RNA and/or the N-termini become more dynamic. This is made evident by the increased susceptibility of expanded TCV CP to chymotrypsin, although no change in RNase susceptibility was found. Since the viral RNA remains inaccessible to nuclease, the implication is that some of the N-termini become dissociated from the RNA, and become extruded from the particle. In the expanded state of TBSV, this appears to be the case, where the N-termini of the B subunits appear to become extruded to the exterior of the particle[136]. Presumably, such a process would likely be mediated by both the expansion event and by the entry of cellular counterions.

In contrast to nuclease protection of the genome in the expanded state, TCV forms striposomes in this state when in the presence of wheat germ extract. This suggests that RNA can exit from pores, and that one end of the RNA can bind to ribosomes. Although a high affinity ribosome binding site has been identified at the 3' end of its genome[141], the likelihood is that the 5' end exits the capsid first. This way the ribosome can bind functionally, and the free energy of translation can be used for a Brownian ratchet mechanism, extruding the RNA and allowing additional ribosomes to bind the 5' end of the genome. This would assist in pulling the rest of the genome through the capsid, coating the RNA in ribosomes, thus masking the genome from the host cell’s RNA silencing

mechanism[108]. The protein capsid would therefore retain its protective functionality right up until the end of genome uncoating and RNA translation.

Although striposomes are seen from expanded particles in the presence of wheat germ extract, these striposomes were relatively rare (Saskia Bakker, pers. comm.), suggesting that additional steps in uncoating may be necessary. The asymmetric particles seen in Figure 5.7 are an obvious candidate for such an uncoating intermediate. The partial extrusion of viral RNA from these particles suggest that removal of N-terminal R domains mediates charge repulsion events allowing the initial extrusion of the genome. The loss of positively charged R domains via proteolysis would increase the electrostatic potential of the RNA, thus allowing for this extrusion event to take place. Figure 5.14 proposes a model for TCV disassembly.

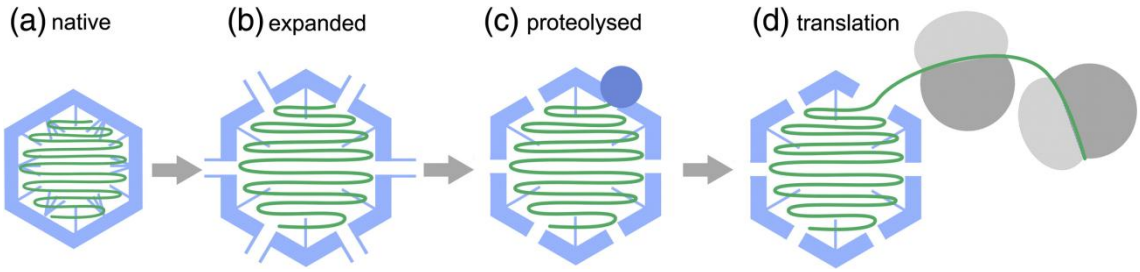


Figure 5.14: A model for TCV uncoating
(a) Native TCV
TCV in its native state is sealed with all the R domains inside the particle contacting the genome.
(b) Expanded TCV
Chelation of calcium ions mediates an expansion event, which makes the genome and N-termini more dynamic. Some of these R domains become accessible to proteases.
(c) Proteolysed TCV
Extruded N-termini become proteolysed, which leads to the extrusion of the genome and generation of asymmetric particles.
(d) Translation
The replication machinery of the plant cell is then allowed access to the extruded genome, and is able to bind to ribosomes, resulting in the formation of striposomes.

This model is consistent with all the data reported in this Chapter, together with the fact that the N-terminal R domains of TCV CP are exquisitely sensitive to proteolysis. These observations strongly suggest that expanded proteolysed particles are indeed an intermediate towards uncoating.

Parallels can be drawn between TCV and many other plant viruses, both in terms of its structural and biochemical similarities[142]. Conditions for *in vitro* expansion and disassembly of one of these viruses, STNV, draw comparisons to TCV. Like TCV, STNV virions will expand upon chelation of stabilising calcium ions, and recombinant particles will dissociate with EDTA treatment in high ionic strength at pH 8.5[34; 80; 81; 113]. However, no observed increase in diameter has been determined for EDTA-treated recombinant STNV, as characterised by TEM, svAUC or cryo-EM (data not shown). It is possible that STNV virions expand more than the recombinant capsid due to the differences of their packaged RNA. It may be that STNV virions have increased electrostatic repulsion resulting from a more condensed RNA genome compared to the recombinant[89]. However, EDTA-treated recombinant particles appear to be more sensitive to proteases than untreated recombinant capsids, suggesting that the N-terminal regions, including the short helix become more dynamic upon chelation of calcium, and that the N-termini can be extruded from pores on the surface of the capsid. It has been shown that not all plant viruses containing calcium expand upon chelation of calcium ions. For example, in red clover necrotic mottle virus (RCNMV), calcium chelation elicits the introduction of pores in the capsid with no discernible increase in capsid diameter[143].

EDTA and chymotrypsin treatment of recombinant STNV particles results in cleavage of the N-terminal region from a subset of CP subunits. The precise location of the cleavage site has not been determined, but the proteolysed subunits appear to have similar banding patterns by SDS-PAGE to the STNV virion treated with trypsin[80]. The trypsinated cleavage site at Lys 28 results in the removal of the N-terminal helix[80]. It is therefore likely that Phe 30 is a potential chymotrypsin cleavage site in the EDTA-treated recombinant particle.

Size exclusion chromatography of EDTA and chymotrypsin-treated STNV produced asymmetric particles with similar characteristics to those seen with TCV. However, when incubated with wheat germ extract, no striposomes were found by TEM. It is possible that the mRNA extruded from the recombinant capsid is not adequately presented to ribosomes, since it is not a *bona fide* STNV genome and lacks suitable eukaryotic ribosome binding sites. It would be interesting to see whether STNV VLPs assembled around an STNV-C genome would expand similarly to STNV virions analysed by Unge *et al,* and if these particles are capable of producing striposomes.

6. General Discussion

The (+) ssRNA icosahedral viruses are pathogens which infect host cells from every kingdom of life. In this class of virus, the processes of capsid assembly and disassembly are extremely poorly understood. These critical stages in the virus life cycle are carefully orchestrated, and maintain a high degree of molecular precision. In this thesis, two model (+) ssRNA icosahedral viruses have been interrogated, and specific mechanisms for capsid assembly and uncoating are proposed. The results described in this thesis suggest that common strategies may be utilised by this class of virus.

In the case of STNV, the CP N-terminus is essential for capsid assembly[81]. N-terminal deletion mutants of as little as 12 amino acids result in an assembly-deficient protein[81]. In addition, the N-terminus of STNV CP is extremely proteolytically sensitive in disassembled VLPs, although STNV capsids are protease resistant. Cleavage occurs at residues close to the N-terminal helix, such that the N-terminal helix is removed. A known tryptic cleavage site has been identified at Arg28, which results in removal of the N-terminal helix[80]. Proteolysed STNV CP in this way is rendered assembly deficient.

Many other viruses also contain basic N- or C-terminal extensions. In most cases, deletion or proteolysis of these extensions prevents capsid assembly of the correct architecture, although in some cases smaller sub-particles are assembled[34; 144; 145; 146]. Deletion of the first 31 amino acids from the N-terminus of FHV CP however, results in the formation of a variety of differently sized particles[147]. In contrast, deletion mutants of up to 20 amino acids in the N-

termini of recombinant norovirus VLPs do not affect capsid assembly, although longer N-terminal deletion mutants abrogates assembly[148].

For these viral CPs with basic N- or C- termini, why assembly is perturbed by their removal could be due to multiple factors. In the case of viruses with a triangulation number greater than 1, the coat protein subunits must be able to exist in different conformations. To compose an icosahedral capsid, these conformations must be present at specific locations within the surface lattice, so as to designate curvature at the right locations. High resolution structural studies of these viruses show that differences between these conformers can be very subtle[149]. Most of these conformers involve the switching between an ordering and disordering of flexible regions in the coat protein at the N- or C-terminus[147]. Truncation of N-terminal regions in TCV for example, prevents the formation of flat contacts between dimers, thus perturbing the formation of *T*=3 particles, and initiating the production of *T*=1 particles.

In the case of STNV, capsid assembly is dependent on the presence of RNA and full length STNV CP with an intact N-terminal helix. The positively charged N-terminal helix is therefore an obvious candidate in mediating important protein: RNA contacts that are essential for the formation of a *T*=1 capsid. The question was therefore asked, what is the nature of these protein: RNA interactions?[81]. In order to determine whether STNV packages RNA in a sequence-specific manner, SELEX experiments were performed using immobilised STNV CP, where RNA aptamers were directed against the RNA binding sites of the CP. Several RNA ligands were identified, which contained sequence similarity to portions of the STNV-1 genome. One of these aptamers,

B3, contained 16/24 matches, including a 10/10 contiguous stretch. Secondary structure predictions of these aptamers by MFold revealed that each aptamer could fold into stem-loop structures, each displaying the motif AXXA in the context of a 4, 5 or 6 nt loop. When the genomes of STNV-1, -2, and -C were analysed for potential stem loop structures based on the structures of these aptamers, around 30 stem-loops were identified for each of the genomes. Since there appears to be no single high affinity packaging signal present in the STNV genome, a series of stem-loops could be used to facilitate capsid assembly. Many RNAs will contain STNV binding sites by chance, thus accounting for non-specific RNA packaging. However, genomic RNAs have an advantage over random RNAs since they contain multiple copies of this preferred binding motif, and at ideal spacings. Binding of STNV CP to multiple binding sites on a cognate STNV-C genome has been shown to trigger collapse of the RNA[60]. This collapsed state does not occur upon STNV CP binding to MS2 sub-genomic fragments under the same conditions. Similar results were observed using the MS2 system in this lab, suggesting that viral CPs chaperone folding of their genomic RNAs through local CP interactions to multiple putative PSs[60].

Recently, SELEX experiments against TCV CP_2 were performed by Gareth Platt under my joint supervision with Dr David Bunka, using material that I purified. Two libraries were produced, one n30 library and an n50 library. Preliminary analysis of these libraries has identified at least two significantly populated aptamer sequence families. Several of these align to multiple regions of the TCV genome, suggesting a similar scenario of packaging signals to STNV and MS2 (see Figure 6.1).

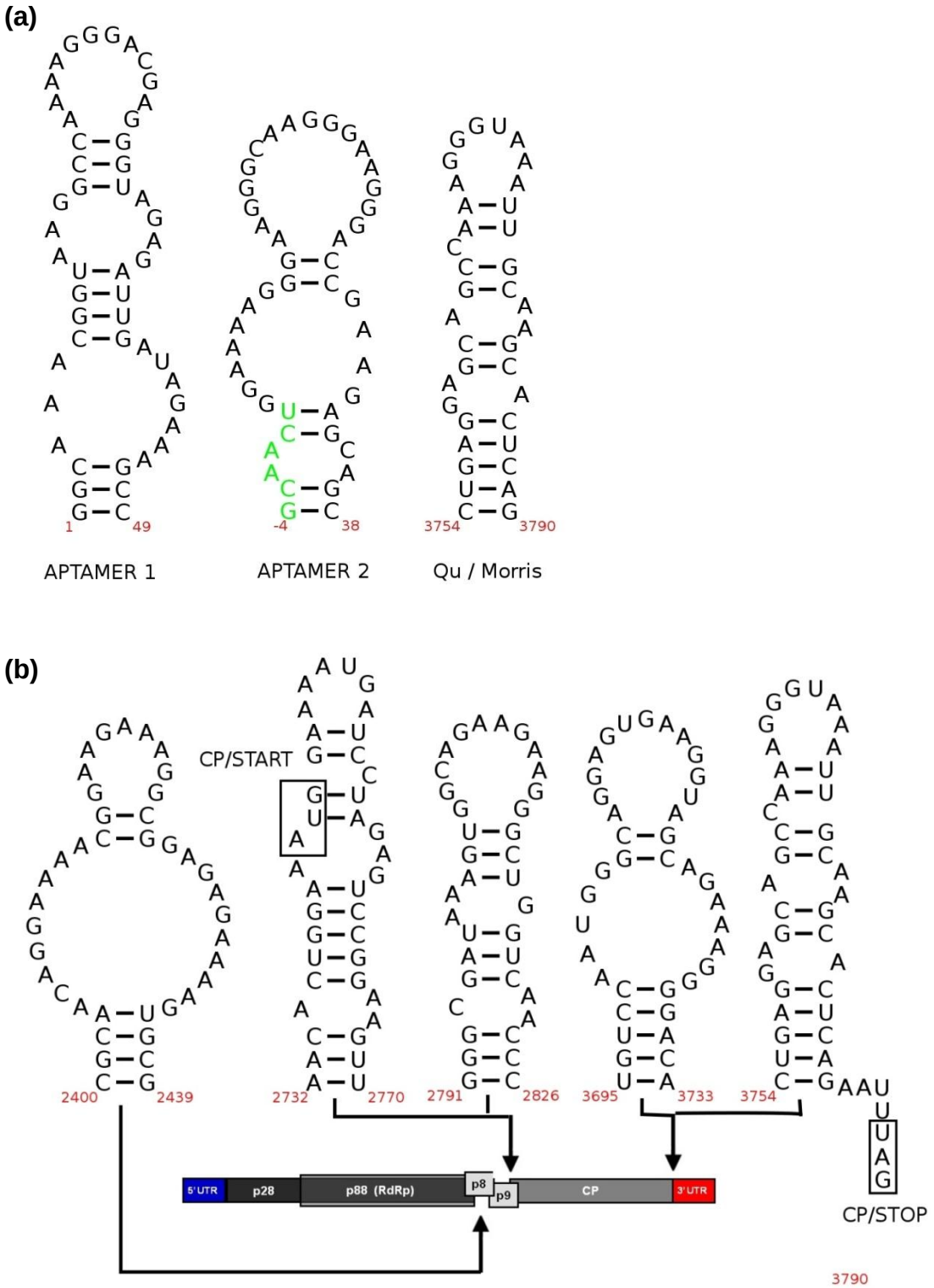


Figure 6.1: SELEX against TCV CP_2 produced several hits from portions on the TCV genome.

(a) Secondary structures of aptamers

Preliminary analysis of SELEX has revealed two aptamer sequences with similar secondary structures to the known PS identified in Qu and Morris.

(b) Matches to the TCV genome

Sequence alignments of aptamers reveal regions of similarities in important regulatory regions, including CP start and stop codons. Figure courtesy of Dr Eric Dykeman.

The aptamers appear to be folded into secondary structures resembling the minimal assembly sequence previously identified *in vivo* by Qu and Morris[33]. The aptamers both contain regions of base-paired stems interrupted by two bulges, whilst displaying a purine-rich loop. The matches of the aptamers to the genome appear to be present in regulatory important regions. Such sites encompass the start and stop codons in the TCV genome, implying that TCV CP may bind to these sites to play important roles in translational repression, in an analogous manner to MS2. These results, although preliminary, further exemplify the exquisite control of capsid assembly, and demonstrate how different viruses may utilise similar mechanisms in their life cycles.

Another example of how viruses may utilise similar mechanisms is in the process of uncoating. Many plant viruses share similar characteristics, both biochemically and structurally. In Chapter 5, a putative disassembly mechanism has been proposed for TCV. It is likely to be shared by other plant viruses, such as STNV[80; 83; 84], SBMV[135] and CCMV[150], where divalent cations are used to stabilise their capsids. These calcium ions can be removed from these capsids *in vitro* by EDTA treatment, which mimics the conditions of the plant cell cytoplasm *in vivo*. In the plant cell, regulation of calcium ions is tightly controlled. Calcium ions in the plant cell are used as an intracellular messenger for signalling. In plant cells, the intracellular calcium ion concentration is thought to be in the region of 100 nM, compared to 1 mM for magnesium ions[84; 151]. Removal of calcium ions from the capsids of these viruses results in charge repulsion between negatively charged side chains, which usually results in expansion of the capsid shell and subsequent pore formation.

The expanded state has long been thought of as the first step in disassembly[82]. In the two model systems tested in this thesis, expansion produces particles that are sensitive to proteases. Proteolysis mediates the release of a subset of coat protein subunits from these particles, such that the remaining protein shells lose their icosahedral symmetry. These particles also appear to be extruding RNA at specific sites. In the case of TCV, when these particles are incubated in the presence of wheat germ extract, the formation of “striposomes”, which are thought to be polysomal arrays of ribosomes bound via thin tethers of extruding RNA, can be visualised by TEM. The formation of these striposomes suggests that ribosomes may be instrumental in the uncoating mechanisms of these viruses.

The protrusions seen in EDTA and chymotrypsin-treated STNV VLPs suggest that this is a common uncoating strategy. Other plant viruses such as SBMV and CCMV have proteolytically sensitive N-terminal domains. It would be interesting to see whether these expanded virus particles contained similar protrusions when protease treated. Protease treatment of the coat proteins of SBMV and CCMV results in the production of a single product, indicating that the protein remains folded in a way that prevents other residues becoming accessible to protease digestion[152; 153]. These results are fully consistent with the uncoating mechanism described in Chapter 5 being ubiquitous for this family of plant viruses.

7. Future Work

- Chapter 4 describes crystallisation of STNV VLPs packaged with TR and B3 short. These were crystallised in identical conditions to that of the wild-type STNV-1, recombinant STNV and STNV-B3 VLPs. However, a full data set could not be acquired from these crystals. If I had more time, I would purify these VLPs on a large scale and try different crystallisation conditions to generate nicely diffracting crystals so that full data sets (high and low resolution) could be obtained. From these data sets, it is hoped that an electron density map could be generated to visualise CP: RNA interactions in high resolution.
- Chapter 4 also describes the formation of 'half capsid' intermediates. If I had more time, I would purify these species on a large scale and perform cryo-EM to generate a three-dimensional structure. I would also attempt to crystallise these species to see if they would diffract. To my knowledge, this would be the first capsid intermediate structure ever to be solved.
- Chapter 5 describes how STNV VLPs undergo an extrusion event upon EDTA and chymotrypsin treatment. It would be interesting to see if these VLPs bind to ribosomes in an analogous manner to TCV if they contained a *bona fide* STNV genome.
- In the general discussion, preliminary data from SELEX against RNA binding sites of TCV CP_2 is discussed. Analysis identified several aptamers mapping the TCV genome at important regulatory regions. It would be interesting to see if these aptamers could initiate capsid assembly, and how efficiently each aptamer was able to initiate assembly relative to other RNAs.

8. References

1. Schneemann, A. (2006). The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* **60**, 51-67.
2. Roberts, R. J. (2005). How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences* **102**, 5905-5908.
3. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annual Review of Immunology* **12**, 433-455.
4. Kost, T. A., Condreay, J. P. & Jarvis, D. L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotech* **23**, 567-575.
5. Manchester, M. & Singh, P. (2006). Virus-based nanoparticles (VNPs): Platform technologies for diagnostic imaging. *Advanced Drug Delivery Reviews* **58**, 1505-1522.
6. Suzich, J. A., Ghim, S. J., Palmer-Hill, F. J., White, W. I., Tamura, J. K., Bell, J. A., Newsome, J. A., Jenson, A. B. & Schlegel, R. (1995). Systemic immunization with papillomavirus L1 protein completely prevents the development of viral mucosal papillomas. *Proceedings of the National Academy of Sciences* **92**, 11553-11557.
7. Lomonossoff, G.P., & Evans, D.J., (2011). Applications of plant viruses in bionanotechnology. *Current Topics in Microbiology and Immunology*, Springer-Verlag Berlin Heidelberg 2011.
8. Fauquet, C., Mayo, M.A., Maniloff, J., Dusselburger, U., and Ball, LA., (2005). Virus taxonomy: International Comittee on the Taxonomy of Viruses.

9. Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J.-M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences* **108**, 17486-17491.

10. Larson, S. B., Day, J., Greenwood, A. & McPherson, A. (1998). Refined structure of satellite tobacco mosaic virus at 1.8Å resolution. *Journal of Molecular Biology* **277**, 37-59.

11. Crowther, R. A. & Klug, A. (1975). Structural Analysis of Macromolecular Assemblies by Image Reconstruction from Electron Micrographs. *Annual Review of Biochemistry* **44**, 161-182.

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

13. Knipe, D., and Howley, P.M., (2007). Field's Virology.

14. Bovin, J. O., Wallenberg, R. & Smith, D. J. (1985). Imaging of atomic clouds outside the surfaces of gold crystals by electron microscopy. *Nature* **317**, 47-49.

15. Marks, L. D. (1994). Experimental studies of small particle structures. *Reports on Progress in Physics* **57**, 603.

16. Bernal, J. D. & Fankuchen, I. (1941). X-ray and crystallographic studies of plant virus preparations. *The Journal of General Physiology* **25**, 111-146.

17. Crick, F. H. C. & Watson, J. D. (1956). Structure of Small Viruses. *Nature* **177**, 473-475.

18. Fraenkel-Conrat, H. & Williams, R. C. (1955). Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid

Components. *Proceedings of the National Academy of Sciences of the United States of America* **41**, 690-698.

19. Caspar, D. L. D. & Klug, A. (1962). Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symposia on Quantitative Biology* **27**, 1-24.
20. Johnson, J. E. & Speir, J. A. (1997). Quasi-equivalent viruses: a paradigm for protein assemblies. *Journal of Molecular Biology* **269**, 665-675.
21. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368-373.
22. Brown, C. S., Van Lent, J. W., Vlak, J. M. & Spaan, W. J. (1991). Assembly of empty capsids by using baculovirus recombinants expressing human parvovirus B19 structural proteins. *Journal of Virology* **65**, 2702-2706.
23. Bancroft, J. B., Hiebert, E. & Bracker, C. E. (1969). The effects of various polyanions on shell formation of some spherical viruses. *Virology* **39**, 924-930.
24. Bancroft, J. B., Hills, G. J. & 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**, 354-379.
25. Zlotnick, A. (1994). To Build a Virus Capsid: An Equilibrium Model of the Self Assembly of Polyhedral Protein Complexes. *Journal of Molecular Biology* **241**, 59-67.

26. Zandi, R., van der Schoot, P., Reguera, D., Kegel, W. & Reiss, H. (2006). Classical Nucleation Theory of Virus Capsids. *Biophysical Journal* **90**, 1939-1948.

27. Chen, Z. G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G. & Johnson, J. E. (1989). Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science* **245**, 154-159.

28. Lane, C. (2007). *Structural studies of satellite tobacco necrosis virus (STNV),* PhD thesis (University of Leeds)

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

30. Larson, S. B., Lucas, R. W., Greenwood, A. & McPherson, A. (2005). The RNA of turnip yellow mosaic virus exhibits icosahedral order. *Virology* **334**, 245.

31. Larson, S. B. & McPherson, A. (2001). Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Current Opinion in Structural Biology* **11**, 59.

32. Johnson, J. M., Willits, D. A., Young, M. J. & Zlotnick, A. (2004). Interaction with Capsid Protein Alters RNA Structure and the Pathway for In Vitro Assembly of Cowpea Chlorotic Mottle Virus. *Journal of Molecular Biology* **335**, 455.

33. Qu, F. & Morris, T. J. (1997). Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J. Virol.* **71**, 1428-1435.

34. Sorger, P. K., Stockley, P. G. & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J Mol Biol* **191**, 639-58.

35. Stockley, P. G., Stonehouse, N. J., Murray, J. B., Goodman, S. T., Talbot, S. J., Adams, C. J., Liljas, L. & Valegard, K. (1995). Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res* **23**, 2512-8.

36. Witherell, G. W., Gott, J. M. & Uhlenbeck, O. C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol* **40**, 185-220.

37. Schmidt, T., Johnson, J. E. & Phillips, W. E. (1983). The spherically averaged structures of cowpea mosaic virus components by X-ray solution scattering. *Virology* **127**, 65-73.

38. Timmins, P. A., Wild, D. & Witz, J. (1994). The three-dimensional distribution of RNA and protein in the interior of tomato bushy stunt virus: a neutron low-resolution single-crystal diffraction study. *Structure* **2**, 1191-1201.

39. Li, T., Chen, Z., Johnson, J. E. & Thomas, G. J. (1992). Conformations, interactions, and thermostabilities of RNA and proteins in bean pod mottle virus: investigation of solution and crystal structures by laser Raman spectroscopy. *Biochemistry* **31**, 6673-6682.

40. Koning, R., van den Worm, S., Plaisier, J. R., van Duin, J., Pieter Abrahams, J. & Koerten, H. (2003). Visualization by Cryo-electron Microscopy of Genomic RNA that Binds to the Protein Capsid Inside Bacteriophage MS2. *Journal of Molecular Biology* **332**, 415.

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

42. Felden, B., Florentz, C., McPherson, A. & Giege, R. (1994). A histidine accepting tRNA-like fold at the 3'-end of satellite tobacco mosaic virus RNA. *Nucl. Acids Res.* **22**, 2882-2886.

43. Schroeder, S.J., Stone, J.W., Bleckley, S., Gibbons, T. & Mathews, Deborah, M. (2011). Ensemble of Secondary Structures for Encapsidated Satellite Tobacco Mosaic Virus RNA Consistent with Chemical Probing and Crystallography Constraints. *Biophysical Journal* **101**, 167-175.

44. Zeng, Y., Larson, S. B., Heitsch, C. E., McPherson, A. & Harvey, S. C. (2012). A model for the structure of satellite tobacco mosaic virus. *Journal of Structural Biology*.

45. Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M. & Johnson, J. E. (2001). The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat Struct Mol Biol* **8**, 77.

46. Johnson, K. N., Tang, L., Johnson, J. E. & Ball, L. A. (2004). Heterologous RNA encapsidated in Pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. *J Virol* **78**, 11371-8.

47. Wery, J.-P., Reddy, V. S., Hosur, M. V. & Johnson, J. E. (1994). The Refined Three-Dimensional Structure of an Insect Virus at 2·8 Å Resolution. *Journal of Molecular Biology* **235**, 565-586.

48. Fisher, A. J. & Johnson, J. E. (1993). Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361**, 176.

49. Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. (1993). The Refined Structure of Bacteriophage MS2 at 2·8 Å Resolution. *Journal of Molecular Biology* **234**, 620-639.

50. Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36-41.

51. Toropova, K., Basnak, G., Twarock, R., Stockley, P. G. & Ranson, N. A. (2008). The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J Mol Biol* **375**, 824-36.

52. Dykeman, E. C., Grayson, N. E., Toropova, K., Ranson, N. A., Stockley, P. G. & Twarock, R. (2011). Simple Rules for Efficient Assembly Predict the Layout of a Packaged Viral RNA. *Journal of Molecular Biology* **408**, 399-407.

53. Krahn, P. M., O'Callaghan, R. J. & Paranchych, W. (1972). Stages in phage R17 infection: VI. Injection of a protein and RNA into the host cell. *Virology* **47**, 628-637.

54. Shiba, T. & Suzuki, Y. (1981). Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **654**, 249-255.

55. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & 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**, 541-552.

56. Morton, V. L., Dykeman, E. C., Stonehouse, N. J., Ashcroft, A. E., Twarock, R. & Stockley, P. G. (2010). The Impact of Viral RNA on Assembly Pathway Selection. *Journal of Molecular Biology* **401**, 298-308.

57. Woodson, S. A. (2000). Compact but disordered states of RNA. *Nat Struct Mol Biol* **7**, 349-352.

58. Tihova, M., Dryden, K. A., Le, T.-v. L., Harvey, S. C., Johnson, J. E., Yeager, M. & Schneemann, A. (2004). Nodavirus Coat Protein Imposes Dodecahedral RNA Structure Independent of Nucleotide Sequence and Length. *J. Virol.* **78**, 2897-2905.

59. Bothner, B., Schneemann, A., Marshall, D., Reddy, V., Johnson, J. E. & Siuzdak, G. (1999). Crystallographically identical virus capsids display different properties in solution. *Nat Struct Mol Biol* **6**, 114-116.

60. Borodavka, A., Tuma, R. & 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*.

61. Butler, P. J. (1999). Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philos Trans R Soc Lond B Biol Sci* **354**, 537-50.

62. Stonehouse, N. J., Scott, D. J., Fonseca, S., Murray, J., Adams, C., Clarke, A. R., Valegard, K., Golmohammadi, R., van den Worm, S., Liljas, L. & Stockley, P. G. (1996). Molecular interactions in the RNA bacteriophage MS2. *Biochem Soc Trans* **24**, 412S.

63. Stonehouse, N. J., Valegard, K., Golmohammadi, R., van den Worm, S., Walton, C., Stockley, P. G. & Liljas, L. (1996). Crystal structures of MS2 capsids with mutations in the subunit FG loop. *J Mol Biol* **256**, 330-9.

64. Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature* **371**, 623-626.

65. Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. & Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol* **270**, 724-38.

66. Nugent, C. I., Johnson, K. L., Sarnow, P. & Kirkegaard, K. (1999). Functional Coupling between Replication and Packaging of Poliovirus Replicon RNA. *Journal of Virology* **73**, 427-435.

67. Frolova, E., Frolov, I. & Schlesinger, S. (1997). Packaging signals in alphaviruses. *Journal of Virology* **71**, 248-58.

68. Aldovini, A. & Young, R. A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *Journal of Virology* **64**, 1920-1926.

69. Zhong, W., Dasgupta, R. & Rueckert, R. (1992). Evidence that the packaging signal for nodaviral RNA2 is a bulged stem-loop. *Proceedings of the National Academy of Sciences* **89**, 11146-11150.

70. Lago, H., Parrott, A. M., Moss, T., Stonehouse, N. J. & Stockley, P. G. (2001). Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *Journal of Molecular Biology* **305**, 1131-1144.

71. Hohn, T. (1969). Role of RNA in the assembly process of bacteriophage fr. *Journal of Molecular Biology* **43**, 191-200.

72. Hung, P. P., Ling, C. M. & Overby, L. R. (1969). Self-assembly of Qβ and MS2 phage particles: possible function of initiation complexes. *Science* **166**, 1638-1640.

73. Matthews, K. S. & Cole, R. D. (1972). Shell formation by capsid protein of f2 bacteriophage. *Journal of Molecular Biology* **65**, 1-15.

74. Kassanis, B. (1962). Properties and behaviour of a virus depending for its multiplication on another. *J Gen Microbiol* **27**, 477-88.

75. Babos, P. & Kassanis, B. (1963). Serological relationships and some properties of tobacco necrosis virus strains. *Journal of General Microbiology* **32**, 135-144.

76. Rees, M. W., Short, M. N. & Kassanis, B. (1970). The amino acid composition, antigenicity, and other characteristics of the satellite viruses of tobacco necrosis virus. *Virology* **40**, 448-61.

77. Jones, T. A. & Liljas, L. (1984). Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J Mol Biol* **177**, 735-67.

78. Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Skoglund, U. & Strandberg, B. (1982). Structure of satellite tobacco necrosis virus at 3.0 Å resolution. *J Mol Biol* **159**, 93-108.

79. Ban, N., Larson, S. B. & McPherson, A. (1995). Structural comparison of the plant satellite viruses. *Virology* **214**, 571.

80. Unge, T., Montelius, I., Liljas, L. & Öfverstedt, L.-G. (1986). The EDTA-treated expanded satellite tobacco necrosis virus: Biochemical properties and crystallization. *Virology* **152**, 207.

81. Bunka, D. H., Lane, S. W., Lane, C. L., Dykeman, E. C., Ford, R. J., Barker, A. M., Twarock, R., Phillips, S. E. & 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. *J Mol Biol* **413**, 51-65.

82. Durham, A. C., Hendry, D. A. & Von Wechmar, M. B. (1977). Does calcium ion binding control plant virus disassembly? *Virology* **77**, 524-33.

83. Montelius, I., Liljas, L. & Unge, T. (1988). Structure of EDTA-treated satellite tobacco necrosis virus at pH 6.5. *Journal of Molecular Biology* **201**, 353.

84. Montelius, I., Liljas, L. & Unge, T. (1990). Sequential removal of Ca^{2+} from satellite tobacco necrosis virus : Crystal structure of two EDTA-treated forms. *Journal of Molecular Biology* **212**, 331.

85. Bentley, G. A., Lewit-Bentley, A., Liljas, L., Skoglund, U., Roth, M. & Unge, T. (1987). Structure of RNA in satellite tobacco necrosis virus. A low resolution neutron diffraction study using $1H_2O/2H_2O$ solvent contrast variation. *J Mol Biol* **194**, 129-41.

86. Sjoberg, B. (1977). A small-angle X-ray investigation of the satellite tobacco necrosis virus. *Eur J Biochem* **81**, 277-83.

87. Van Emmelo, J., Ameloot, P., Plaetinck, G. & Fiers, W. (1984). Controlled synthesis of the coat protein of satellite tobacco necrosis virus in Escherichia coli. *Virology* **136**, 32.

88. Sangita, V., Lokesh, G. L., Satheshkumar, P. S., Vijay, C. S., Saravanan, V., Savithri, H. S. & Murthy, M. R. N. (2004). *T*=1 Capsid Structures of Sesbania Mosaic Virus Coat Protein Mutants: Determinants of *T*=3 and *T*=1 Capsid Assembly. *Journal of Molecular Biology* **342**, 987-999.

89. Lane, S. W., Dennis, C. A., Lane, C. L., Trinh, C. H., Rizkallah, P. J., Stockley, P. G. & Phillips, S. E. (2011). Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* **413**, 41-50.

90. Lee, K. K. (2010). Architecture of a nascent viral fusion pore. *The EMBO Journal* **29**, 1299-311.

91. Mellman, I. (1996). Endocytosis and molecular sorting. *Annual Review of Cell and Developmental Biology* **12**, 575-625.

92. Mercer, J., Schelhaas, M. & Helenius, A. (2010). Virus Entry by Endocytosis. *Annual Review of Biochemistry* **79**, 803-833.

93. Doherty, G. J. & McMahon, H. T. (2009). Mechanisms of Endocytosis. *Annual Review of Biochemistry* **78**, 857-902.

94. Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. (1980). On the entry of semliki forest virus into BHK-21 cells. *The Journal of Cell Biology* **84**, 404-420.

95. Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. (2004). Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* **11**, 567-573.

96. Smith, J. G., Silvestry, M., Lindert, S., Lu, W., Nemerow, G. R. & Stewart, P. L. (2010). Insight into the Mechanisms of Adenovirus Capsid Disassembly from Studies of Defensin Neutralization. *PLoS Pathog* **6**, e1000959.

97. Poranen, M. M., Daugelavicius, R. & Bamford, D. H. (2002). Common principles in viral entry. *Annual review of microbiology* **56**, 521-538.

98. Zaitlin, M. & Hull, R. (1987). Plant Virus-Host Interactions. *Annual Review of Plant Physiology* **38**, 291-315.

99. Brisco, M., Hull, R. & Wilson, T. M. A. (1986). Swelling of isometric and of bacilliform plant virus nucleocapsids is required for virus-specific protein synthesis in vitro. *Virology* **148**, 210-217.

100. Shaw, J. G., Plaskitt, K. A. & Wilson, T. M. A. (1986). Evidence that tobacco mosaic virus particles disassemble contranslationally in vivo. *Virology* **148**, 326-336.

101. Wilson, T. M. A. (1984). Cotranslational disassembly of tobacco mosaic virus in vitro. *Virology* **137**, 255-265.

102. Vossen, M., Westerhout, E., Söderberg-Nauclér, C. & Wiertz, E. (2002). Viral immune evasion: a masterpiece of evolution. *Immunogenetics* **54**, 527-542.

103. Tock, M. & Dryden, D. (2005). The biology of restriction and anti-restriction. *Current opinion in microbiology* **8**, 466-472.

104. Hay, A. J., Gregory, V., Douglas, A. R. & Lin, Y. P. (2001). The evolution of human influenza viruses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **356**, 1861-1870.

105. Wyatt, R. & Sodroski, J. (1998). The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science* **280**, 1884-1888.

106. Hamilton, A. J. & Baulcombe, D. C. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* **286**, 950-952.

107. Mello, C. C. & Conte, D. (2004). Revealing the world of RNA interference. *Nature* **431**, 338-342.

108. Baulcombe, D. (2004). RNA silencing in plants. *Nature* **431**, 356-363.

109. Cao, M., Ye, X., Willie, K., Lin, J., Zhang, X., Redinbaugh, M. G., Simon, A. E., Morris, T. J. & Qu, F. (2010). The Capsid Protein of Turnip Crinkle Virus Overcomes Two Separate Defense Barriers To Facilitate Systemic Movement of the Virus in Arabidopsis. *Journal of Virology* **84**, 7793-7802.

110. Hogle, J. M., Maeda, A. & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus: I. X-ray crystallographic structure analysis at 3.2Å resolution. *Journal of Molecular Biology* **191**, 625-638.

111. Golden, J. S. & Harrison, S. C. (1982). Proteolytic dissection of turnip crinkle virus subunit in solution. *Biochemistry* **21**, 3862-3866.

112. Stockley, P. G., Kirsh, A. L., Chow, E. P., Smart, J. E. & Harrison, S. C. (1986). Structure of turnip crinkle virus: III. Identification of a unique coat protein dimer. *Journal of Molecular Biology* **191**, 721-725.

113. Bakker, S. E., Ford, R. J., Barker, A. M., Robottom, J., Saunders, K., Pearson, A. R., Ranson, N. A. & Stockley, P. G. (2011). Isolation of an Asymmetric RNA Uncoating Intermediate for a Single-Stranded RNA Plant Virus. *Journal of Molecular Biology* **417**, 65-78.

114. Wei, N., Heaton, L. A., Morris, T. J. & Harrison, S. C. (1990). Structure and assembly of turnip crinkle virus: VI. Identification of coat protein binding sites on the RNA. *Journal of Molecular Biology* **214**, 85-95.

115. Bringloe, D. H., Gultyaev, A. P., Pelpel, M., Pleij, C. W. & 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**, 1539-46.

116. Rolfsson, O., Toropova, K., Ranson, N. A. & Stockley, P. G. (2010). Mutually-induced Conformational Switching of RNA and Coat Protein Underpins Efficient Assembly of a Viral Capsid. *Journal of Molecular Biology* **401**, 309-322.

117. Durchschlag, H. (1986). *Thermodynamics data for Biochemistry and Biotechnology*, (Hinz, H. J., ed.) Springer Verlag, New York, NY.

118. Schuck, P. (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal* **78**, 1606-1619.

119. Thomas, C. L., Leh, V., Lederer, C. & Maule, A. J. (2003). Turnip crinkle virus coat protein mediates suppression of RNA silencing in Nicotiana benthamiana. *Virology* **306**, 33-41.

120. Groves, W. E., Davis Jr, F. C. & Sells, B. H. (1968). Spectrophotometric determination of microgram quantities of protein without nucleic acid interference. *Analytical Biochemistry* **22**, 195-210.

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

122. Ellis, R. J. & van der Vies, S. M. (1991). Molecular Chaperones. *Annual Review of Biochemistry* **60**, 321-347.

123. Lane, S. (2003). *Structural and Functional Studies of Recombinant STNV capsids,* PhD Thesis. (University of Leeds)

124. Brissette, J. L., Russel, M., Weiner, L. & Model, P. (1990). Phage shock protein, a stress protein of Escherichia coli. *Proceedings of the National Academy of Sciences* **87**, 862-866.

125. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. (1991). Characterization and sequence of the Escherichia coli stress-induced psp operon. *Journal of Molecular Biology* **220**, 35-48.

126. Mastico, R. A., Talbot, S. J. & Stockley, P. G. (1993). Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *Journal of General Virology* **74**, 541-548.

127. Chu, C. T., Howard, G. C., Misra, U. K. & Pizzo, S. V. (1994). α2-Macroglobulin: A Sensor for Proteolysis. *Annals of the New York Academy of Sciences* **737**, 291-307.

128. Sottrup-Jensen, L. (1989). Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *Journal of Biological Chemistry* **264**, 11539-42.

129. Krimbou, L., Tremblay, M., Davignon, J. & Cohn, J. S. (1998). Association of apolipoprotein E with {alpha}2-macroglobulin in human plasma. *J. Lipid Res.* **39**, 2373-2386.

130. Murray, D. H., Prokop, K. (1968). in: Zorbach, Tipson (Eds.). *Synthetic Procedures in Nucleic Acid Chemistry*, 193-197.

131. Li, T. C., Yamakawa, Y., Suzuki, K., Tatsumi, M., Razak, M. A., Uchida, T., Takeda, N. & Miyamura, T. (1997). Expression and self-assembly of empty virus-like particles of hepatitis E virus. *Journal of Virology* **71**, 7207-13.

132. Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M. & Gelbart, W. M. (2011). Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *Journal of Virology* **86**, 3318-3326.

133. Lomonossoff, G. P. (1995). Pathogen-Derived Resistance to Plant Viruses. *Annual Review of Phytopathology* **33**, 323-343.

134. Harvey, J. J. W., Lewsey, M. G., Patel, K., Westwood, J., Heimstadt, S., Carr, J. P. & Baulcombe, D. C. (2011). An Antiviral Defense Role of AGO2 in Plants. *PLoS ONE* **6**, e14639.

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

136. Robinson, I. K. & Harrison, S. C. (1982). Structure of the expanded state of tomato bushy stunt virus. *Nature* **297**, 563-568.

137. Toropova, K., Stockley, P. G. & Ranson, N. A. (2011). Visualising a Viral RNA Genome Poised for Release from Its Receptor Complex. *Journal of Molecular Biology* **408**, 408-419.

138. Celerino, A.Z., Sherin, S. A.M., John, E. J., Andrew, G. W. L., Ivan, R., Michael, G. R., Dietrich, S. & Tomitake, T. (1980). Structure of southern bean mosaic virus at 2.8 Å resolution. Nature Publishing Group.

139. Rossmann, M. G., Abad-Zapatero, C., Murthy, M. R., Liljas, L., Jones, T. A. & Strandberg, B. (1983). Structural comparisons of some small spherical plant viruses. *Journal of Molecular Biology* **165**, 711-36.

140. Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Current opinion in microbiology* **8**, 451-458.

141. Yuan, X., Shi, K., Young, M. Y. L. & Simon, A. E. (2011). The terminal loop of a 3' proximal hairpin plays a critical role in replication and the structure of the 3' region of Turnip crinkle virus. *Virology* **402**, 271-280.

142. Rao, A. L. N. (2006). Genome Packaging by Spherical Plant RNA Viruses. *Annual Review of Phytopathology* **44**, 61-87.

143. Sherman, M. B., Guenther, R. H., Tama, F., Sit, T. L., Brooks, C. L., Mikhailov, A. M., Orlova, E. V., Baker, T. S. & Lommel, S. A. (2006). Removal of Divalent Cations Induces Structural Transitions in Red Clover Necrotic Mosaic Virus, Revealing a Potential Mechanism for RNA Release. *Journal of Virology* **80**, 10395-10406.

144. Xing, L., Li, T.-C., Mayazaki, N., Simon, M. N., Wall, J. S., Moore, M., Wang, C.-Y., Takeda, N., Wakita, T., Miyamura, T. & Cheng, R. H. (2010). Structure of Hepatitis E Virion-sized Particle Reveals an RNA-

dependent Viral Assembly Pathway. *Journal of Biological Chemistry* **285**, 33175-33183.

145. Savithri, H. S. & Erickson, J. W. (1983). The self-assembly of the cowpea strain of southern bean mosaic virus: Formation of *T*=1 and *T*=3 nucleoprotein particles. *Virology* **126**, 328-335.

146. Harrison, S. C. (1980). Protein interfaces and intersubunit bonding. The case of tomato bushy stunt virus. *Biophysical Journal* **32**, 139-153.

147. Dong, X. F., Natarajan, P., Tihova, M., Johnson, J. E. & Schneemann, A. (1998). Particle Polymorphism Caused by Deletion of a Peptide Molecular Switch in a Quasiequivalent Icosahedral Virus. *Journal of Virology* **72**, 6024-6033.

148. Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V. V. & Estes, M. K. (2002). Structural Requirements for the Assembly of Norwalk Virus-Like Particles. *Journal of Virology* **76**, 4044-4055.

149. Johnson, J. E. (1996). Functional implications of protein-protein interactions in icosahedral viruses. *Proceedings of the National Academy of Sciences* **93**, 27-33.

150. Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. (1995). Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**, 63-78.

151. Laliberte, J.-F. & Sanfacon, H. l. n. (2010). Cellular Remodeling During Plant Virus Infection. *Annual Review of Phytopathology* **48**, 69-91.

152. Chidlow, J. & Tremaine, J. H. (1971). Limited hydrolysis of cowpea chlorotic mottle virus by trypsin and chymotrypsin. *Virology* **43**, 267-278.

153. Tremaine, J. H. & Ronald, W. P. (1978). Limited proteolysis of southern bean mosaic virus by trypsin. *Virology* **91**, 164-172.

9. Appendix

On accompanying CD:

Table 9.1: *E. coli* strain genotypes

Figure 9.1: svAUC analysis of STNV CP under reassembly conditions

Figure 9.2: Mass Spectrum of STNV CP

Figure 9.3: Mass Spectrum for aptamer B3

Figure 9.4: Mass Spectrum for B3 4U

Figure 9.5: Mass Spectrum for B3 short

Figure 9.6: Mass Spectrum for TR

Figure 9.7: Sedimentation equilibrium AUC of STNV ‘half capsids’

Figure 9.8: Size exclusion chromatogram for STNV-TR VLPs

Figure 9.9: Size exclusion chromatogram for STNV-B3 short VLPs

Figure 9.10: SDS-PAGE of EDTA, chymotrypsin treated STNV VLPs

Thesis_RJFord.pdf