RNA packaging and uncoating in simple single-stranded RNA viruses

Saskia Emilie Bakker

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

The University of Leeds

Institute of Molecular and Cellular Biology

September 2011

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

The work described in Chapter 4 has formed part of a jointly-authored publication

Details of the publication are as follows :

Saskia E. Bakker, Robert J Ford, Amy M. Barker, Janice Robottom, Keith Saunders, Arwen R. Pearson, Neil A. Ranson and Peter G. Stockley

Structural and Functional Studies of the Turnip Crinkle Virus Virion Reveal the Roles of Capsid during Genome Uncoating

Submitted

The contributions by each author to the publication are as follows:

Saskia E. Bakker (the candidate) carried out all EM data collection and image processing and most biochemical experiments in the manuscript, provided all the figures in the manuscript and contributed to writing and editing of the manuscript.

Robert J. Ford provided assistance with biochemical experiments (Figure 7).

Amy M. Barker carried out the analytical ultracentrifugation experiments and data analysis (Figure 7) and assisted with the purification of virus.

Janice Robottom provided assistance with biochemical experiments (Figure 6).

Keith Saunders provided the leaves from which the virus was purified.

Dr Arwen R. Pearson contributed to writing and editing of the manuscript.

Dr Neil A. Ranson provided advice, supervision and tuition on all aspects of electron microscopy and contributed to writing and editing of the manuscript.

Prof. Peter G. Stockley provided advice with and supervision of biochemical experiments and contributed to writing and editing of the manuscript.

Acknowledgements

First of all I want to thank all my supervisors for their patience with me. More specifically, I thank Dr Neil Ranson for his help and teaching in EM, Dr Arwen Pearson for her help with crystallography, and Professors Peter Stockley and David Rowlands for their advice and samples. I thank the people in charge of the Leeds Wellcome Trust PhD scheme for putting their trust in me and the Wellcome Trust itself for their generous funding.

I thank Keith Saunders and Dr George Lomonossoff for supplying the TCV infected leaves, Dr James Ault for performing mass spectrometry and Dr Jeff Keen for carrying out the N-terminal protein sequencing. I thank Peiyi Wang and Martin Fuller for keeping the EM lab running on a day-to-day basis, and for their help with more specific problems, and Sean Killen, Mike Wallis and Aidan Richmond for keeping the computers running.

For their help in the lab I would like to thank: Robert Ford, for his help with the gel filtration experiments and other help in the lab; Amy Barker for doing the AUC and Janice Robottom for her help with things during her summer project. James Ross helped me with the ERAV interaction calculations, and Kat Toropova taught me how to do icosahedral reconstructions during my lab rotation in 2007. I would like to thank Miriam Walden for teaching me how to make SDS-PAGE gels and David Bunka for generally answering my questions. Furthermore, I would like to thank everyone in the Pearson lab, Ranson lab and Stockley lab, the Garstang 8.53 office, the EM office, and cake club, for their support, company and valuable discussions. I would like to especially thank my friends Marko Noerenberg, Lucy Chappell, and Briony Yorke for putting up with me.

On a more personal note, I thank my parents for their support throughout school, university and my PhD, and for putting up with me moving abroad and having to take all their holidays in sunny Britain.

I have had much support from my friends in the Netherlands (or elsewhere) at the other end of an email, especially Sita and ‘schaapjes’ Annelies, Marieke, and Mariette.

Lastly, I want to thank Chris for supporting me throughout my entire PhD on the personal and professional level, for helping me through the moments where motivation was low (or absent) and for generally keeping me (mostly) sane.

Abstract

Simple (non-enveloped) small, positive-sense single-stranded RNA viruses infect hosts from all kingdoms of life. However, their assembly and uncoating processes remain poorly understood. For turnip crinkle virus (TCV), 3D reconstructions by cryo-electron microscopy (cryo-EM) are shown for the native and the expanded form. The expanded form is a putative disassembly intermediate and exhibits pores that are large enough to allow exit of single-stranded RNA. Biochemical experiments revealed the expanded form is protease-sensitive, although the RNA genome remains protected from ribonuclease. Virus particles complexed with ribosomes are shown by negative stain EM. Proteolysis causes release of some coat protein from the capsid, while the capsid remains largely intact. Proteolysed particles have lost their icosahedral symmetry and show a protuberance in negative stain EM. Taken together, these results suggest expansion and subsequent proteolysis are essential steps in the uncoating process of TCV, and that the capsid plays multiple roles consistent with ribosome-mediated genome uncoating to avoid host anti-viral activity. Similarly, 3D cryo-EM reconstructions are presented for native equine rhinitis A virus (ERAV) an expanded particle containing no RNA. The native virus fits well with the ERAV crystal structure. The empty particle is a putative disassembly intermediate representing a stage after the release of the RNA genome. A mechanism is suggested that is consistent with the RNA release from the endosome without exposure to the endosomal contents. A crystal structure is presented of satellite tobacco necrosis virus (STNV) virus-like particles containing a small RNA fragment. The coat protein structure is identical to that of native STNV. Although density internal to the coat protein shell has been observed in the experiment that corresponds to earlier experiments, no unambiguous RNA structure can be built into the density. Together, the results presented here shed some light on the life cycle of three of these viruses.

Table of contents

Table of figures

List of tables

Abbreviations

+ssRNA	positive-sense single-stranded RNA
Å	Ångström (10^{-10} m)
bp	base pair
BPMV	bean pod mottle virus
CCD	charge-coupled device
CCMV	cowpea chlorotic mottle virus
CNV	cucumber necrosis virus
CP	coat protein
CPMV	cow pea mosaic virus
(cryo-)EM	(cryo-)electron microscopy
CTF	contrast transfer function
DNA	deoxyribonucleic acid
ds	double-stranded
DYMV	desmodium yellow mottle virus
ERAV	equine rhinitis A virus
FHV	Flock House virus
FMDV	foot and mouth disease virus
ICAM	intercellular adhesion molecule
IRES	internal ribosome entry site
LDL	low-density lipoprotein
m7G	7-methylguanosine
mRNA	messenger RNA
nt	nucleotide
PaV	Pariacoto virus
PSF	point spread function
PvR	poliovirus receptor

RdRp	RNA-dependent RNA polymerase
RGD	arginine-glycine-aspartate
RNA	ribonucleic acid
S	Svedberg
SBMV	southern bean mosaic virus
SNR	signal-to-noise ratio
ss	single-stranded
SBMV	southern bean mosaic virus
STMV	satellite tobacco mosaic virus
STNV	satellite tobacco necrosis virus
T	triangulation number
TBSV	tomato bushy stunt virus
TCV	turnip crinkle virus
TR	translational repressor
TYMV	turnip yellow mosaic virus
VLPs	virus-like particles
VP1	viral protein 1
VP2	viral protein 2
VP3	viral protein 3
VP4	viral protein 4

1 Introduction to virus structure

1.1 Viral capsid architecture

1.1.1 Virus morphologies and classification

Viruses are obligate parasites that can only replicate within the cells of other organisms, as they do not encode their own machinery for protein synthesis or energy production. They infect hosts in all kingdoms of life. They occur in many different shapes and sizes, ranging from small and simple round or rod-like particles, consisting just of protein and nucleic acid, to large and complex assemblies that include multiple nucleic acid molecules, a multitude of different proteins and lipid layers (Figure 1.1). The smallest viruses are the plant satellite viruses (about 17 nm diameter), which only encode their own coat protein and need the host to be simultaneously infected with the appropriate 'helper virus' in order to replicate (Ban *et al.*, 1995). The largest virus discovered to date is *Acanthamoeba polyphaga* mimivirus (APMV), better known as mimivirus, with a diameter of 750 nm (Xiao *et al.*, 2009).

The common components of all viruses are a genome and its protecting protein capsid, which together form the (nucleo)capsid. These can be further protected by a lipid envelope, which can display additional viral proteins for attaching to and entering host cells. In the Baltimore taxonomic system, viruses are classified hierarchically by their genome chemistry (DNA or RNA), capsid morphology (helical or icosahedral), presence of membrane (enveloped or non-enveloped), genome architecture (single- or double-stranded DNA or RNA), virus family, genus, and subtype (Baltimore, 1971). However, the similarity between herpes simplex virus (HSV) and bacteriophage Hong Kong-97 (HK97), or human adenovirus and bacteriophage PRD-1, suggests structural similarities and evolutionary relationships between viruses cross the boundaries between kingdoms of life (Bamford *et al.*, 2005).

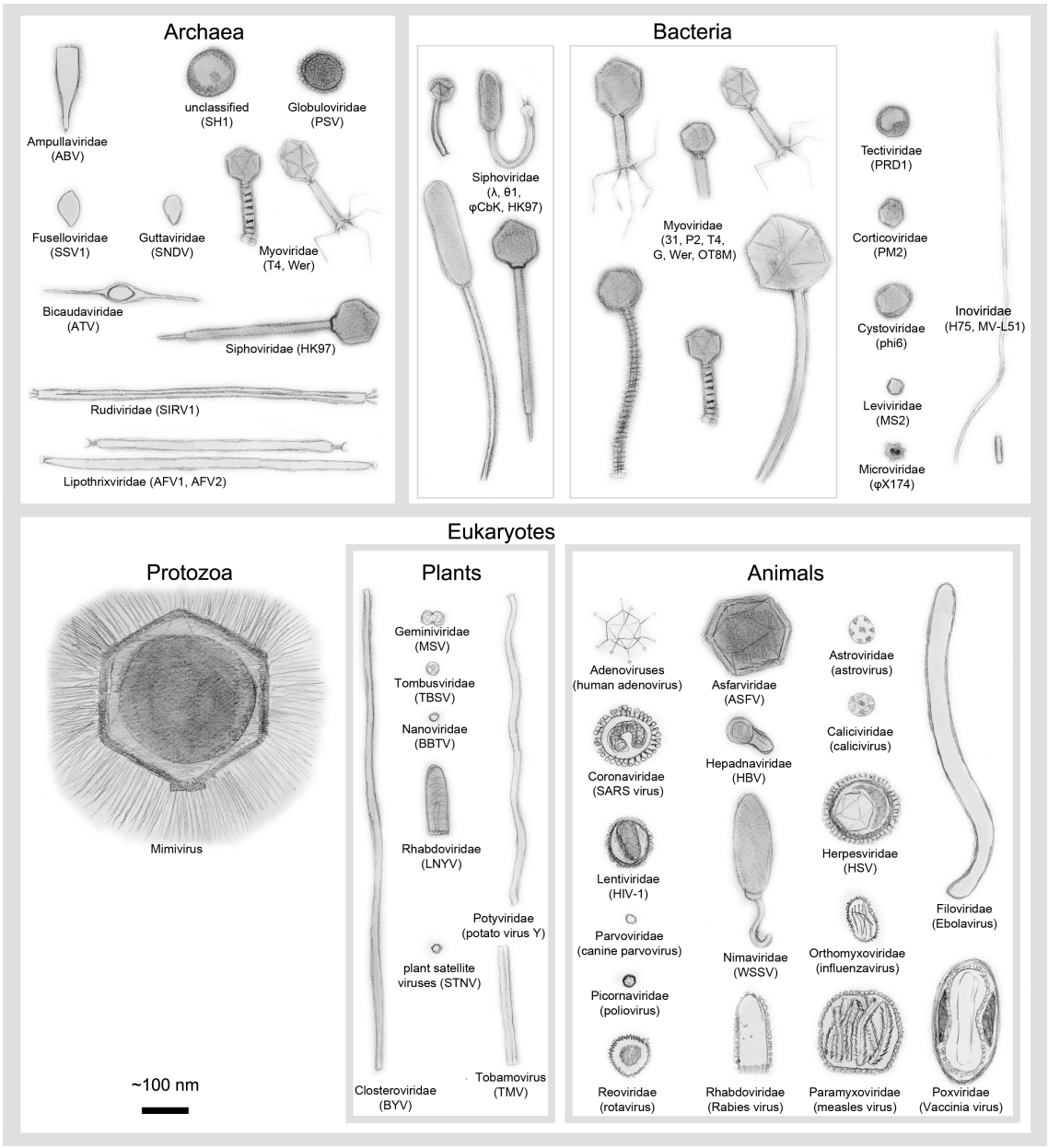


Figure 1.1 Diversity of virus structure
The viruses are shown approximately to the same scale. Archaeal viruses: *Acidianus* bottle-shaped virus (ABV), *Haloarcula hispanica* virus (SH1), *Pyrobaculum* spherical virus (PSV), *Sulfolobus* spindle-shaped virus 1 (SSV1), *Sulfolobus neozealandicus* droplet-shaped virus (SNDV), the Myoviridae bacteriophage T4 and Gluconobacter virus Wer, *Acidianus* two tailed virus (ATV), Siphovirus HK97, *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1), *Acidianus* filamentous virus 1 (AFV1), *Acidianus* filamentous virus 2 (AFV2). Bacteriophages: Siphoviridae colivirus λ, *Bacillus* virus θ1, *Caulobacter* virus ϕCbK, Enterobacteriophage HK97; Myoviridae *Aeromonas* virus 31, *Enterobacteria* virus P2, bacteriophage T4, *Bacillus* phage G, *Gluconobacter* virus Wer and *Serratia* phage OT8M; bacteriophage PRD1, *Alteromonas* phage PM2, *Pseudomonas* phage phi6, bacteriophage MS2, Enterobacteria phage ϕX174, *Thermus* phage H75 and *Acholeplasma* phage MV-L51. Protozoal virus: *Acanthamoeba polyphaga* Mimivirus (mimivirus). Plant viruses: beet yellows virus (BYV), maize streak virus (MSV), tomato bushy stunt virus (TBSV), banana bunchy top virus (BBTV), lettuce necrotic yellows virus (LNYV), satellite tobacco necrosis virus (STNV), potato virus Y and tobacco mosaic virus (TMV). Animal viruses: human adenovirus, severe acute respiratory syndrome (SARS) coronavirus, human immunodeficiency virus (HIV), canine parvovirus, poliovirus, rotavirus, African swine fever virus (ASFV), hepatitis B virus (HBV), white spot syndrome virus (WSSV),

Rabies virus, astrovirus, calicivirus, herpes simplex virus (HSV), human influenzavirus, human paramyxovirus, ebolavirus and vaccinia virus.

The genome length of viruses is limited by their small size compared to bacteria and eukaryotic cells. Although viruses do not, by their nature, possess genes for the production of ribosomes, generation of energy or many other processes essential for the functioning of cells, they still face a variety of challenges. To overcome these challenges, such as cell entry and innate immunity, the virus needs to use coding capacity in a fundamentally limited 'genome space'. These challenges are extraordinarily diverse and depend on the properties of a virus and the host cell type that it infects. They are best described by their place in the viral life cycle (Figure 1.2). Different viruses have evolved a variety of ways to overcome these challenges, and an exhaustive account is not feasible here. Rather I will illustrate the fascinating diversity present in the virus world. Many viruses are still poorly characterised; the best-studied viruses are those infecting bacteria (bacteriophage), some simple plant viruses and viruses that infect humans or livestock.

The method of cell entry each virus employs depends strongly on the host cell type. Cell walls are a significant barrier to overcome for a virus. For DNA bacteriophage, in particular, intricate machineries to 'inject' the genome into the cell have evolved (Ivanovska *et al.*, 2007). Most plant viruses, on the other hand, rely on damage to the cell wall by insect or fungal vectors that feed on the plant, or are transmitted through seeds or vegetative propagation (Poranen *et al.*, 2002). For viruses infecting animal hosts, the barrier to the cytoplasm consists of just the cell membrane and many animal viruses are taken up into the cell by endocytosis (Smith *et al.*, 2010, Brandenburg *et al.*, 2007, Odegard *et al.*, 2009). They then have to deal with the harsh conditions of the endosome and deliver their genome into the cytoplasm. For enveloped viruses the cell entry process can involve fusion of the cellular and viral membranes, initiated by proteins on the membrane surfaces (Lee, 2010). Viral uncoating is discussed in more detail in section 1.3.

Once the genome and any additional viral proteins have entered the cytoplasm, the replication process can be initiated. The viral replication process consists of two parts, the replication of the genome and the production of new virus particles, and depends on the type of genome. The proteins required for the new virus particles are produced by the host ribosomes using templates encoded by the viral genome, while the genome itself is replicated by viral polymerases or with the cellular genome.

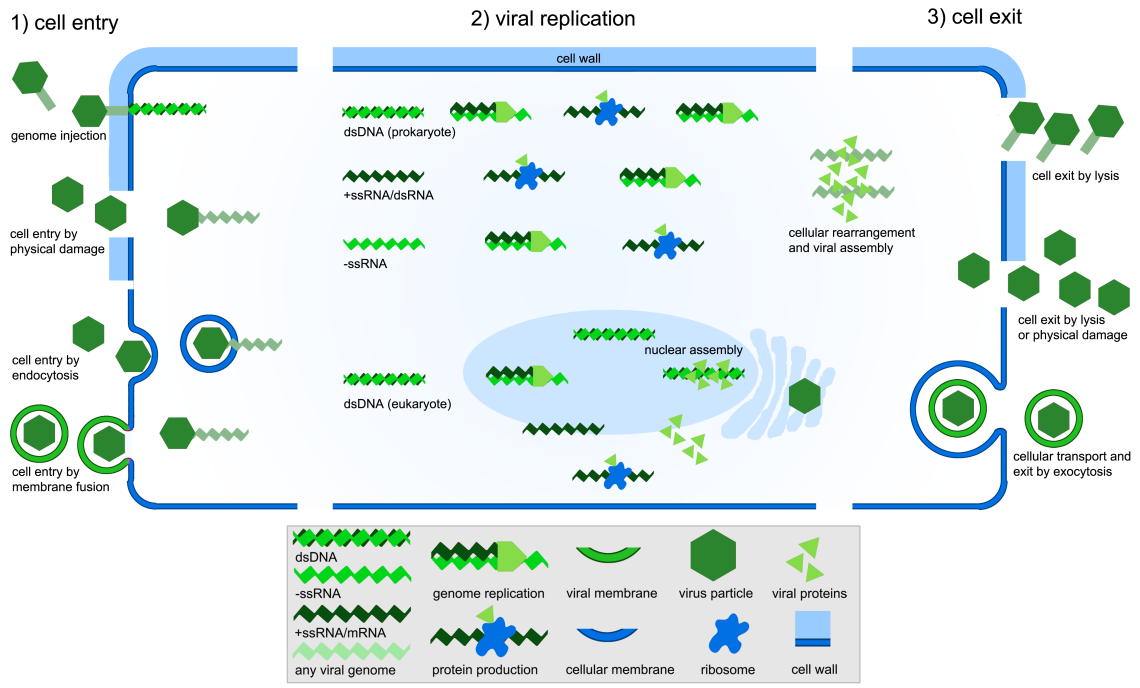


Figure 1.2 A generalised life cycle for viruses
1) Cell entry and viral uncoating. From top to bottom: tailed bacteriophage inject their DNA genome into the cell while the capsid remains outside; plant viruses commonly use mechanical damage for cell entry; unenveloped animal viruses bind cell surface receptors and trigger endocytosis; enveloped animal viruses can fuse directly with the cell membrane. 2) Viral replication. From top to bottom: in prokaryotes, double-stranded (ds) DNA is transcribed to mRNA and translated into viral proteins, and assembly occurs in the cytoplasm; in prokaryotes and eukaryotes, positive-sense single-stranded (+ss) RNA viruses can use the positive strand for the production of viral proteins such as the replicase which duplicates the genome; negative-sense single-stranded (-ss) RNA viruses carry viral proteins to produce the mRNA template for protein production; in eukaryote hosts, dsDNA viruses transcribe and multiply their genome in the nucleus, from where mRNA is exported to the cytosol for protein production. The protein is imported to the nucleus where the virions assemble. 3) Cell exit, from top to bottom: tailed bacteriophages usually exit the cell through lysis; plant viruses use physical damage; animal viruses may use lysis; animal viruses can also use the cellular export machinery.

The replication process is simplest for positive-sense, single-stranded RNA (+ssRNA) viruses. In these viruses, the genome can be used directly as messenger RNA (mRNA) and the replication process starts with protein translation. Viral proteins include the capsid proteins and proteins that disrupt host cell function. One of the viral

proteins is an RNA-dependent RNA polymerase (RdRp), which replicates the genome. The capsid proteins and genomes together assemble into new virus particles, ready to infect another host cell (Sun *et al.*, 2010).

For viruses with other genome types, the genome cannot be directly used as the template for protein production, and a separate transcription step is required to produce mRNA. For viruses with a genome other than +ssRNA, mRNA has to be produced before protein production can commence. Negative-sense, single-stranded RNA (-ssRNA) viruses carry an RNA polymerase within the virion (Bhella *et al.*, 2004). Double-stranded RNA (dsRNA) viruses export single-stranded RNA from the virion, which is then translated and copied (Kobayashi *et al.*, 2009).

Single- and double-stranded DNA viruses replicate their genomes in the cell nucleus of eukaryotic hosts, using a mixture of host and viral enzymes. As in cellular protein production, mRNA for viral protein is exported from the nucleus and proteins are produced in the cytoplasm. The proteins are transported back into the nucleus where viral assembly takes place (Mettenleiter *et al.*, 2009).

Progeny virions self-assemble from the newly-produced genomes and proteins, and obtain any other components they need from the cell. During infection, the cellular chemistry and architecture is often remodelled to produce 'virus factories', where high concentrations of the viral components are present (Laliberté and Sanfaçon, 2010). Herpesviruses and tailed bacteriophages, which are evolutionarily related, assemble the viral capsid (or 'head') of protein around a scaffold protein, then use a packaging motor to fill the head with DNA (Baker *et al.*, 2005). By contrast, most ssRNA viruses appear to assemble their capsid around the genome (Sun *et al.*, 2010).

After assembly, the virus has to exit the cell. Some viruses use the cellular transport and secretion machinery to exit the cell; others can use simple budding from the membrane, cause the host cell to lyse or rely on vectors to transport them (Poranen *et al.*, 2002).

An additional challenge for all viruses is the defence mechanisms of their hosts, which are discussed further in section 1.4. To use the coding capacity of the genome most efficiently, viral components commonly perform multiple roles in the life cycle of the virus.

1.1.2 Viral symmetry and quasi-equivalence

In order to use the limited amount of genome capacity most efficiently, all known viruses form a capsid from multiple copies of a small number of proteins. If these proteins all have to make the same contacts with their neighbours, i.e. be in equivalent environments, this can limit the capsid size or shape. This, in turn, limits the size of the viral genome and therefore the ability of the virus to encode proteins.

The first virus for which the basic composition and symmetry was unequivocally established was tobacco mosaic virus (TMV), which consists of approximately 2,500 identical 20 kDa proteins in a helical arrangement (Watson, 1954; Franklin 1955). A helical arrangement of coat proteins always results in an equivalent environment and does not pose a limit on genome size, as for a longer genome the helix can simply be extended (Figure 1.3a). Unenveloped helical viruses mostly infect plants and bacteria. Many enveloped viruses, such as measles virus and Ebolavirus, also have a helical capsid around the genome, within the membrane envelope. The envelope protects the nucleocapsid and contains further proteins for cell entry and other viral functions (Bhella *et al.*, 2004).

Of the 'spherical' plant viruses, tomato bushy stunt virus (TBSV) and turnip yellow mosaic virus (TYMV) were the first to be characterized. It was discovered early that these both crystallised with a single virus per unit cell in a cubic space group, limiting the possible symmetry of the virus particle itself (Bernal *et al.*, 1938; Carlisle and Dornberger, 1948; Bernal and Carlisle, 1948). Crick and Watson first proposed that viruses with cubic symmetry are likely to have a multiple of 12 as the number of subunits. They also suggested that the asymmetric unit making up the capsid could

consist of several protein subunits that did not obey strict symmetry (Crick and Watson, 1956). Around the same time, both of these viruses, as well as poliovirus, were further characterized and confirmed to have icosahedral symmetry in the protein, but not the RNA component (Caspar, 1956; Klug *et al.*, 1957a; Klug *et al.*, 1957b; Finch and Klug, 1959). Additionally, the individual subunits were estimated to be about 20 kDa, or at least not larger than 50 kDa (Harris and Hindley, 1961; Harris and Knight, 1955; Yamazaki and Kaesberg, 1961). Improvements in electron microscopy methods, especially the development of negative staining techniques (Brenner and Horne, 1959), allowed assembly units called 'capsomeres' to be observed for many viruses, confirming the prediction that viruses would look "polyhedral or with bumps on, like a rather symmetrical mulberry" (Crick and Watson, 1956). In a review, Horne and Wildy showed how these capsomeres were consistent with icosahedral symmetry (Horne and Wildy, 1961). In a seminal paper, however, Caspar and Klug brought together these distinct observations to define the theoretically allowed numbers of coat protein subunits making up viruses of various sizes, and how these subunits could be arranged within the capsid. This provided a theoretical framework to describe how very large capsids can be built from a single protein subunit (Caspar and Klug, 1962).

The largest enclosed structure that can be built with a single subunit in a single identical conformation is the icosahedron. When each of the 20 faces of the icosahedron is occupied by three copies of the same protein, 60 equivalent positions are created. This limits the size of the capsid, and therefore the size of the genome that can be stored in it (Figure 1.3b). Larger icosahedral capsids can only be created by introducing more coat proteins in each icosahedral face. In this case, however, the protein-protein interfaces can no longer be fully equivalent, as the same side of the protein is now required to make three-fold (or quasi-six-fold) and five-fold contacts (Figure 1.3c). The number of coat proteins for each of the 60 equivalent positions is called the triangulation number *T*. Adding progressively more subunits per asymmetric

unit leads to ever-larger capsids with a larger number of different but quasi-equivalent orientations for the coat protein (Caspar and Klug, 1962).

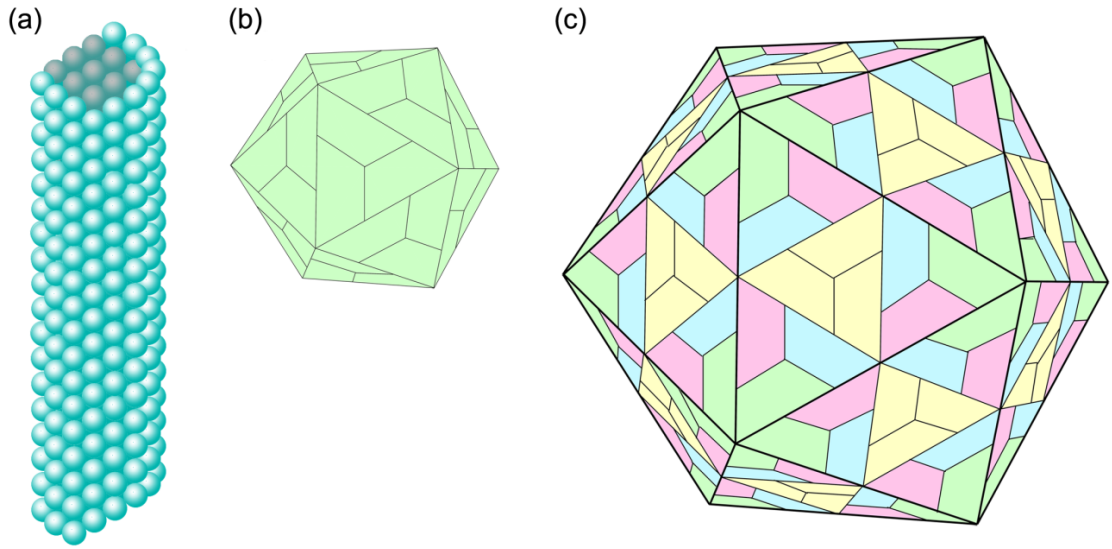


Figure 1.3 Schematic representations of helical and icosahedral virus capsids
(a) A helical capsid. Every coat protein subunit makes identical contacts with its neighbours. The length of the capsid is unlimited, but the shape is always helical. (b) An icosahedral T=1 capsid. Again, each coat protein subunit, shown as a green wedge, makes identical contacts with its neighbours. The capsid size is limited to 60 protein subunits. (c) An icosahedral T=4 capsid. Four types of coat protein subunit (coloured pink, green, cyan and yellow) can be distinguished based on the contacts with neighbouring subunits. Larger capsids can be obtained by adding more coat proteins, but the number of coat proteins is always a multiple of 60.

It seems this method is less straightforward than producing a larger capsid by using a larger coat protein. However, three additional bases are needed in the genome to encode each additional amino acid in the coat protein. This means that using larger, equivalent coat proteins provides no net gain in 'genome space'. However, in some cases, such as in the *Picornaviridae* and Comoviruses, the virus encodes multiple coat proteins that together form an equivalent (*T*=1) capsid (Chen *et al.*, 1989). This strategy allows the capsid proteins to evolve independently of each other. This provides greater flexibility in surface of the capsid, permitting changes in the receptor binding site and surface antigens without affecting capsid stability (Liljas *et al.*, 2002).

Some viruses, e.g. bean pod mottle virus (BPMV), use segmentation as a strategy to expand their genome. These viruses package different parts of the genome in

separate virus particles (Chen *et al.*, 1989). The disadvantage of this strategy is that the host cell has to be simultaneously infected with all the segments, otherwise viral replication cannot take place. Influenzavirus, which has a segmented genome in helical capsids, has overcome this challenge by packaging all segments within a single membrane envelope (Muramoto *et al.*, 2006).

Crystal structures of many viruses are now available. The earliest structures determined were those of tomato bushy stunt virus (TBSV) in 1978, southern bean mosaic virus (SBMV) in 1980, satellite tobacco necrosis virus (STNV) in 1982 and poliovirus in 1985 (Harrison *et al.,* 1978; Abad-Zapatero *et al.*, 1980; Liljas *et al.*, 1982; Hogle *et al.,* 1985). The largest viruses for which crystal structures are available are pseudo-*T*=25 viruses: bacteriophage PRD1 and human adenovirus (Abrescia *et al.*, 2004; Reddy *et al.,* 2010). Additionally, many structures of viral component proteins such as replicases, proteases and integrases have been published. These structures have increased our understanding of how the viral proteins contribute to the life cycle, and of the protein-protein interactions that give rise to the self-assembly of virus capsids. However, for non-icosahedral viruses and flexible structures, crystallography is of very limited use.

Electron microscopy and especially cryo-electron microscopy (cryo-EM) have therefore become a very valuable technique in virus research. For large icosahedral viruses, cryo-EM is approaching the resolution of crystallographic data (Zhang *et al.*, 2010). Cryo-EM also enables the elucidation of genome structures (see section 1.2.1) and the study of pleiomorphic virus structure. Moreover, cryo-electron tomography can be used to study various processes, such as membrane fusion and viral release, in cells or model systems (Maurer *et al.*, 2008).

This general introduction has been intended to place this work in its scientific and historical context. However, the rest of the introduction, like the experimental sections of this work, will focus on three small, icosahedral, positive-sense RNA viruses: turnip

crinkle virus (TCV), satellite tobacco necrosis virus (STNV) and equine rhinitis A virus (ERAV), and their related viruses.

1.2 Viral RNA structure

Through crystallographic studies, the capsid protein and the interactions between the protein subunits are the best-understood component of the virus. As such, these were long seen as the driving force in virus assembly (Zlotnick *et al.*, 2000). The complete, infectious virus particle, or virion, was considered to be relatively inert. Through biochemical experiments, however, we have now learned more about the role of the genome in virus assembly, the dynamic properties of the virus capsid, and the other proteins involved in viral replication. It has become increasingly clear that all parts of the virus play their own unique role(s) in the viral life cycle, even if this is not immediately clear from the static pictures of the protein structure revealed by X-ray crystallography. For example, the structure of the genome within the virion is not often seen in virus crystal structures (see section 1.2.1) and has therefore often been neglected in models of virus assembly, yet it has been shown for several RNA viruses that capsid assembly cannot take place in the absence of RNA (Stockley *et al.*, 1995; Lane *et al.*, 2011).

1.2.1 Icosahedral averaging in virus structures

The structure of the genome is not often seen in virus crystal structures, because only the protein component of the virus exhibits strict icosahedral symmetry at high resolution (Klug *et al.*, 1957). During the crystallization process, each virus that is added to the growing crystal packs randomly in one of 60 orientations, as the capsid has 60-fold symmetry. For the elucidation of the capsid structure, this is not a problem as there is no difference between one orientation and the next. The RNA inside the virus, however, does not follow the same symmetry and therefore the random packing leads to 60 different orientations of RNA within the same crystal (Figure 1.4a). This means the resulting structure of the RNA is averaged over the 60 orientations, showing

strong density only in those positions where the nucleic acid is ordered, i.e. bound in a significant number of the capsid equivalent units (Figure 1.4b, c). This is similar to the averaging that occurs in crystallography for flexible loops or ligands that occupy only a fraction of the binding sites.

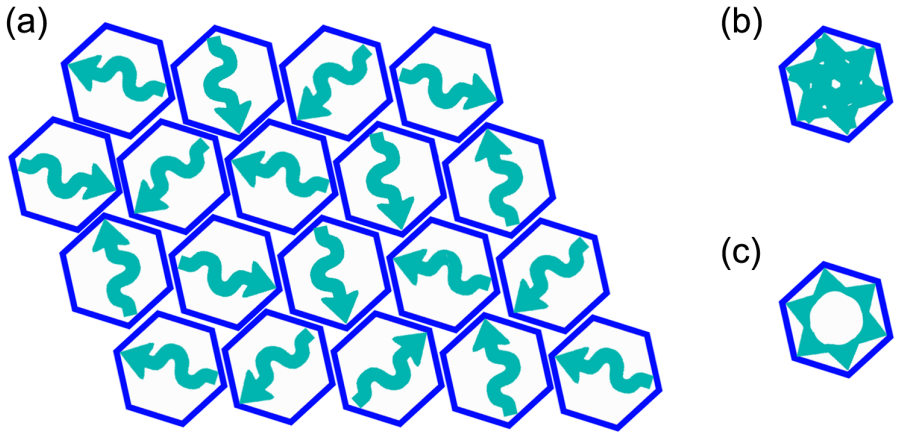


Figure 1.4 Effect of crystal packing of an icosahedral virus on the visibility of the RNA. (a) A schematic representation of viruses in a crystal packed in different orientations regarding the RNA. The viral capsid is represented as a blue hexagon with the RNA as a cyan arrow. (b) The resulting RNA structure is the average of the RNA in the different orientations. (c) The averaged structure if parts of the RNA (i.e. the tail of the arrow) are disordered as well as asymmetric, with another part of the RNA (the arrowhead) bound tightly to the capsid.

A similar problem occurs in single-particle cryo-electron microscopy reconstructions of viruses where icosahedral averaging is also often used. The signal from the RNA in the electron micrograph is not strong enough to drive asymmetric reconstruction against the strongly symmetric signal of the coat protein. However, in cryo-EM reconstructions, more density for RNA is usually seen than in viral crystal structures. As EM is an image-based technique, the data includes information at all resolutions and is not dependent on phasing, so although averaged, density for RNA will be present in the reconstruction as long as it is not completely disordered. It is possible to visualise less icosahedrally ordered RNA by crystallography, but only if the diffraction experiment itself is optimised to permit the accurate collection of very low-resolution reflections. If low-resolution reflections are not included in the datasets, only the parts of the RNA that are ordered to a similar degree as the protein can be resolved (Tsuruta *et al.*, 1998).

The first virus for which RNA was observed crystallographically was bean pod mottle virus (BPMV). A trefoil-shaped electron density, into which 7 nucleotides (nt) per coat protein subunit could be modelled, could be seen on the three-fold symmetry axis. Three of the bases have weaker density, suggesting partial occupancy (Chen *et al.*, 1989). Since this report, some structured RNA has been observed in a number of other virus structures. In Flock House virus (FHV), a 10 base pair (bp) segment of double stranded RNA was observed on the two-fold symmetry axis, such that the two strands were symmetry-related (Fisher and Johnson, 1993). A similar arrangement of RNA was seen in Pariacoto virus (PaV), with a 25 bp RNA duplex along the two-fold symmetry axis (Tang *et al.*, 2001). A nine bp duplex along the two-fold symmetry axis and a single nucleotide on the coat protein of satellite tobacco mosaic virus (STMV) suggests the RNA forms base-paired sections connected by disordered loops (Larson *et al.*, 1993, Larson and McPherson, 2001). In turnip yellow mosaic virus (TYMV), extensive icosahedrally ordered RNA was seen, forming single stranded loops in the capsomeres and double-stranded helices positioned along the five-fold and perpendicular to the three-fold icosahedral symmetry axes. The combined RNA visible in the crystal structure accounted for as much as 70% of the genome (Larson *et al.*, 2005). In desmodium yellow mottle virus (DYMV), two nucleotides were found associated with the A subunit of the capsid and 9 nucleotides with the B/C dimer. Binding over several subunits suggests that either the binding site and RNA-protein interaction only forms after the subunits assemble, or the RNA acts as a nucleation site for the assembly of the capsid (Larson *et al.*, 2000).

Cryo-EM reconstructions of many viruses have also been calculated. As mentioned above, these generally show much larger amounts of the genetic material inside the capsid, albeit at lower resolution that is not sufficient to see detailed interactions.

The bacteriophage MS2 has been extensively studied by cryo-EM as well as X-ray crystallography and, although the X-ray structure of the wild-type MS2 virion showed no density for the genome, in cryo-EM reconstructions extensive density was seen

inside the capsid (Valegard *et al.*, 1990; Koning *et al.*, 2003; Toropova *et al.*, 2008). Similarly, the crystal structure of poliovirus did not reveal any electron density corresponding to the packaged genomic RNA, but in cryo-EM a densely packed genome was observed (Hogle *et al.,* 1985; Bubeck *et al.*, 2005a). Density additional to that seen in the crystal structure was also seen for TBSV and TYMV, among others, in cryo-EM reconstructions (Aramayo *et al.*, 2005; Boettcher and Crowther, 1996, Opalka *et al.*, 2000).

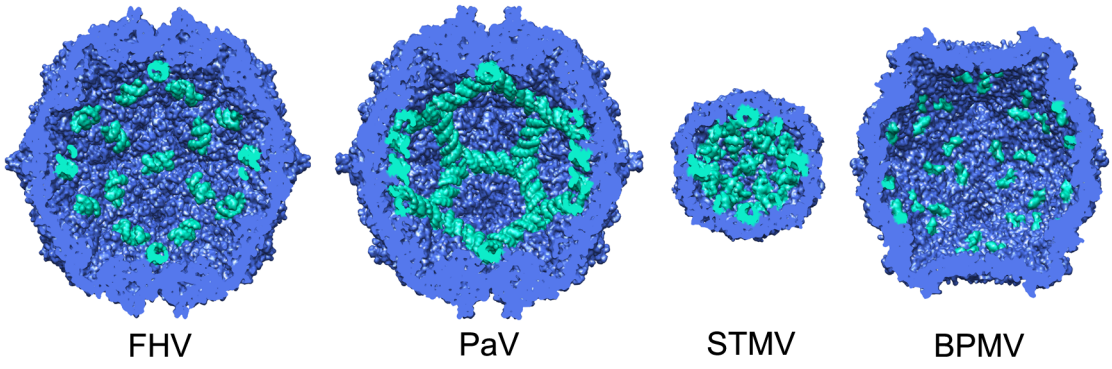


Figure 1.5 Viral RNA from crystal structures
From left to right: Flock House virus (FHV) with its 10 base pair dsRNA, pariacoto virus (PaV) with a dodecahedral cage of 25 base pair dsRNA, satellite tobacco mosaic virus (STMV) with a 7 base pair dsRNA and bean pod mottle virus (BPMV) with its trefoils of 21 bases.

Single-particle virus reconstructions by cryo-EM usually use icosahedral averaging as there is no sufficiently large asymmetric feature to distinguish the different orientations of the genome. As such, the resulting structures, including the genome, are still icosahedrally averaged. For poliovirus and bacteriophage MS2, however, 3D reconstructions have been calculated using reduced symmetry. Poliovirus particles, bound to receptor-decorated liposomes, were imaged by cryo-electron microscopy, and a reconstruction was made with 5-fold symmetry using the receptors as a marker (Bubeck *et al.*, 2005a). Bacteriophage MS2 was bound to bacterial pili and also reconstructed with 5-fold symmetry. These reconstructions show that the genome is not evenly distributed in the virus particle (Toropova *et al.*, 2011).

1.2.2 RNA-protein interactions and RNA recognition

RNA-protein interactions are mediated by the same forces as protein-protein interactions, and can be broadly separated into two categories: sequence-specific RNA recognition and non-sequence specific interactions.

Non-specific RNA binding interactions predominantly involve electrostatic interactions and hydrogen bonds to the ribose-phosphate backbone of the RNA as well as some stacking interactions between aromatic amino acid side chains and unpaired RNA bases. Many viruses contain positively charged residues on the inner surface of the capsid protein, or positively charged protein domains thought to interact non-specifically with the RNA (Sacher and Ahlquist, 1989).

Sequence-specific interactions rely on the different hydrogen bonding capabilities of the four bases that make up RNA. Single-stranded RNA is more structurally flexible than double-stranded DNA, and protein-RNA interactions are therefore more diverse. In addition to Watson-Crick base-pairing (G-C and A-T) RNA can also form non-Watson-Crick base pairs, exposing the face usually involved in hydrogen bonding for protein interactions (Jones *et al.*, 2001).

1.2.2.1 Non-specific protein-RNA interactions in viruses

Examples of non-specific protein-RNA interactions have been observed in various viruses. In the crystal structure of satellite tobacco mosaic virus (STMV), double-helical segments of RNA are seen at the two-fold symmetry axes, modelled as G-C. The RNA electron density was weaker towards the three-fold axes, with a single unconnected base visible. (Larson *et al.*, 1993; Larson *et al.*, 1998). A 10-base pair-long section of double-stranded RNA is also seen at the 2-fold icosahedral symmetry axis of Flock House virus. These were modelled preferentially as pyrimidines, but purines were included to conserve Watson-Crick base pairing (Fisher and Johnson, 1993).

A similar but longer (25 bp) RNA duplex is found on the 2-fold axis of Pariacoto virus, modelled as A-U pairs (Tang *et al.*, 2001). Pariacoto virus coat protein expressed

in a baculovirus expression system forms virus-like particles (VLPs) that package cellular or baculovirus RNA. A similar amount is packaged as in the native virus and the resulting structure looks much like that of the indigenous RNA, indicating the protein-RNA interactions in this case are non-specific (Johnson *et al.*, 2004).

In these examples, it seems the position of the duplexes on the 2-fold icosahedral symmetry axes causes the differences between the individual segments to be averaged out in the crystal, allowing no determination of a consensus sequence. It is likely that these interactions do not rely on a specific RNA sequence, but rather on the three-dimensional conformation of the RNA.

1.2.2.2 Specific protein-RNA interactions in viruses

One of the best-understood specific protein-RNA interactions in viruses is that of bacteriophage MS2. The binding to a high-affinity RNA stem-loop, which also serves as the translational repressor (TR) of the replicase gene, causes a conformational change of the coat protein dimer from the symmetrical C/C conformation to an asymmetrical A/B conformation. This results in one of the protein loops becoming more flexible, removing a steric clash at the five-fold symmetry axes that would otherwise make capsid formation less favourable (Stockley *et al.*, 2007). Although the MS2 genome only contains one copy of the TR sequence, each of the 60 A/B dimers in the capsid requires the binding of a stem-loop to obtain the appropriate conformation. It is thought that although in the genome the TR sequence has the highest affinity and serves as an initiation and recognition site for the packaging of viral RNA, other stem-loop structures in the RNA genome are also capable of inducing the conformational switch required for the formation of a complete capsid (Basnak *et al.*, 2010). This requirement, together with the fact that each end of the RNA is bound to a single copy of the maturation protein, restricts the number of possible RNA structures within the virion to three, closely related, options that fit with *in vitro* reassembly data and the density from the five-fold averaged MS2 cryo-EM reconstruction (Dykeman *et al.*, 2011;

Toropova *et al.*, 2011). Once assembly is initiated by the first CP dimer binding to the genomic TR, more dimers bind to the genome and capsid protein where this is most energetically favourable until the virus particle is complete (Dykeman *et al.*, 2011). Crystal structures and *in vitro* binding studies of the MS2 capsid protein with several RNA stem-loops confirm that RNA binding is not limited to the TR stem-loop (Grahn *et al.*, 1999; Valegard *et al.*, 1997; Stockley *et al.*, 1995).

In a 2.8 Å crystal structure for bean pod mottle virus (BPMV), density was seen for RNA-containing particles that was absent in that of empty particles. Both the bases and the backbone are involved in the protein-RNA interactions and the RNA binding increases order in the first 19 residues of the L subunit. The density was fit as AGUCUC, although the density does not really allow for a distinction beyond that between purine and pyrimidine. The genome contains 200 copies of the motif adenine-purine-pyrimidine-pyrimidine-pyrimidine-X with no regular spacing between them (Chen *et al.*, 1989; Lin *et al.*, 2003). The lack of regular spacing, and therefore regular structure, between these motifs is probably the reason the intermediate segments of the genome are averaged out and not observed in the crystal structure. Because its genome is segmented, the packaging strategy of BPMV needs to be somewhat flexible to accommodate both RNA molecules.

Similar to MS2, STNV coat protein (CP) can be recombinantly expressed in *E.coli* and forms virus-like particles (VLPs) resembling native virus. A crystal structure of STNV VLPs shows electron density for the RNA, but only when low-resolution data are taken into account. This indicates extensive RNA binding must occur in the VLP, even though it packages the CP mRNA rather than the genomic RNA (Lane *et al.*, 2011). Virus and VLPs can be disassembled and reassembled in the presence, but not the absence, of RNA. An RNA oligomer has been identified that binds the CP with high affinity and allows RNA-free CP to assemble into VLPs (Bunka *et al.*, 2011). Although STNV lacks a single strong initiation signal like that of MS2, the large number of redundant weaker interactions will be energetically favourable and therefore the STNV

genome (or recombinant CP mRNA) is more likely to be packaged compared to background RNA (Bunka *et al.*, 2011).

For TCV, a 28-nucleotide hairpin loop has been identified as the essential packaging signal, and this has successfully been used to reassemble virus-like particles (Qu and Morris, 1997; Oleksandr Boredavka and Peter Stockley, personal communication).

In many small ssRNA viruses, it appears that interactions between the genomic RNA and the coat protein are the driving force behind the differentiation between the coat protein conformers, as well as providing the template for protein production. Even the equivalent, *T*=1 particle of STNV requires RNA binding for assembly (Bunka *et al.*, 2011). Relying solely on a certain set of sequence-specific interactions at each equivalent binding site in the capsid would restrict the virus' ability to encode functional proteins, but a degree of sequence recognition is required to ensure the virus specifically packages its genome rather than other RNA present in the cell.

Additionally, predicted secondary structures for the RNA genome of icosahedral viruses are more compact than those of random RNA sequences for any given length, whereas the structures of helical viruses do not show this compaction (Yoffe *et al.*, 2008). The differences between the many examples studied so far suggest that there is not one fixed RNA structure for small ssRNA viruses, but rather that the evolutionary selection for RNA structure, protein sequence and the RNA-protein interactions at the same time resulted in different compromises for different viruses.

1.3 Viral uncoating

Virus particles, as seen by cryo-EM and X-ray crystallography, are often presented as rigid and inert structures. However, the demands of the viral life cycle require the capsid to be a dynamic structure. Natural selection ensures the capsids must efficiently self-assemble and protect their genomes from the environment. These pressures are counteracted by the requirement that the genome must be released after cell entry - a

virus that is too rigid and does not release its genome cannot reproduce. Additionally, the capsid must target the virus to a suitable host cell.

Enveloped viruses often enter the cell by membrane fusion, either directly or after endocytosis (Lee, 2010). However, for non-enveloped viruses this is not possible and they have evolved other methods of cell entry and genome uncoating. Many animal viruses, such as poliovirus, ERAV and adenovirus, are first taken up into the cell by receptor-mediated endocytosis (Hogle, 2002; Groppelli *et al.*, 2010; Smith *et al.*, 2010). Endocytosis is a process used by cells to internalise molecules from the cell surface or the environment that are too large to cross membranes via diffusion or through channels (Mellman, 1996). In the case of clathrin-mediated endocytosis, the best understood endocytotic pathway, the cell membrane contains regions rich in receptors on the outside of the cell and clathrin on the cytoplasmic side. The clathrin causes the membrane to curve and bud off into the cell, resulting in a vesicle within the cytoplasm called an endosome (Doherty and McMahon, 2009). Endocytosis can also be mediated by clathrin-independent pathways, such as caveolin-dependent endocytosis (Doherty and McMahon, 2009). Once the vesicle has formed, it is processed further in the early endosome, which serves as a sorting station (Mellman, 1996). It can then be recycled to the cell surface in a recycling endosome, or be targeted for degradation in the late endosome and lysosome. The pH in the early and late endosome is lower than in the cytosol (Mellman, 1996). This pH change can be used by viruses to drive conformational changes in proteins, resulting in the externalisation of hydrophobic parts of the protein that can aid membrane disruption. In poliovirus, it is thought a combined action of the myristylated VP4 and the N-terminal region of VP1 forms pores in the endosomal membrane, allowing the genome to enter the cytosol and leaving behind a stable empty particle (Brandenburg *et al.*, 2007). FHV has internal amphiphilic helices, called the γ-peptides. On acidification of the endosome, the γ-peptides are externalised via conformational changes of the particle, and released by autocatalytic cleavage. The peptides are then capable of membrane disruption (Odegard *et al.*,

2009). Adenovirus also enters the cell via endocytosis, and undergoes conformational changes on acidification, but its DNA genome is not released directly into the cytoplasm. Instead, a partially uncoated virus particle enters the cytoplasm and exploits cellular transport machinery to reach the nucleus (Smith *et al.*, 2010).

Plant viruses, on the other hand, experience a major hurdle to cell entry in the form of the cell wall. Unlike bacterial viruses with their specialised injection mechanisms, the cell entry process of plant viruses is often reliant on physical damage to the cell wall. Plant viruses are transmitted either through seed or vegetative propagation, or by insect or fungal vectors, which damage or weaken the cell wall (Poranen *et al.*, 2002). The relation between virus and vector is often very specific and it seems the host specificity of plant viruses is primarily determined by the vector (Zaitlin and Hull, 1987). The uncoating of some plant viruses has been shown to be co-translational, i.e. the genome is mechanically uncoated by the ribosome, resulting in virus-ribosome complexes that are known as striposomes; these have also been observed *in vivo* (Shaw *et al.*, 1986). For some icosahedral plant viruses, translation and therefore uncoating is dependent on particle expansion (Brisco *et al.*, 1986). Those viruses that exhibit empty particles and are primarily stabilised by strong protein-protein interactions, such as CCMV and TYMV, do not expand but still release their RNA efficiently *in vivo*, losing 5-8 coat protein subunits in the process (Matthews and Witz, 1985).

1.4 Host defence mechanisms

As viruses infect organisms in all kingdoms of life, anti-viral defence mechanisms are found in all organisms. Bacterial defences within the cell wall include the restriction enzymes, which recognise specific nucleotide sequences or modifications as 'alien' and destroy the nucleic acid in question (Tock and Dryden, 2005). Higher organisms have evolved more elaborate defences, such as the vertebrate immune system and RNA interference.

RNA interference is a process found in all eukaryotes for regulation of gene expression and protection against transposons and viruses. It depends on short fragments (21-26 bp) of complementary RNA (Hamilton and Baulcombe, 1999). A precursor dsRNA is processed into the small interfering RNA fragments by an enzyme called Dicer. The double-stranded siRNA is then separated into two strands (Baulcombe, 2004). The guide strand is bound to an RNA-induced silencing complex (RISC) or RNA-induced transcriptional silencing complex (RITS), while the complementary passenger strand is degraded (Mello and Conte, 2004). RISC is involved in post-translational silencing by binding to and causing degradation of the mRNA or viral genomes, while RITS is involved in chromatin silencing by methylation (Baulcombe 2004; Mello and Conte, 2004). Although dsRNA is the strongest trigger, in lower eukaryotes RNA silencing can also be triggered by ssRNA or DNA through a dsRNA intermediate produced by a cellular RNA-dependent RNA polymerase (Mello and Conte, 2004).

In addition to RNA interference, vertebrates have elaborate innate and adaptive immune systems capable of distinguishing between self- and non-self, healthy and infected cells. Circulating antibodies can recognise pathogens and trigger the production of more antibodies, while the presentation of antigens on the cell surface of infected cells marks these cells for destruction by killer cells (Summerfield *et al.*, 2009). Additionally, innate recognition mechanisms such as Toll-like receptors can trigger antiviral responses such as inflammation and the interferon response and also activate the adaptive immune system (Kawai and Akira, 2006). The vertebrate immune system is complex, and a detailed description of its components is beyond the scope of this thesis.

Some viruses, in turn, have evolved mechanisms to disable the host anti-viral defences. Bacteriophages encode proteins that mimic the structure and charge distribution of DNA to block the active site of restriction enzymes (Tock and Dryden, 2005), while plant viruses produce proteins that disable RNA silencing (Cao *et al.*,

2010). Many animal viruses, such as HIV, have variable regions or glycosylation sites on the viral surface to hide the more conserved functional domains of the viral proteins from the immune system (Reitter *et al.*, 1998). Some, including FMDV, can also interfere more directly with the innate immune system by interfering with cellular transcription and/or translation (Summerfield *et al.*, 2009). The rapid evolution of viruses compared to the host cell often allows viruses to successfully evade the immune system (Summerfield *et al.*, 2009).

1.5 Overview of thesis

In this work the RNA structure and uncoating mechanism of two viruses, TCV and ERAV, are studied by cryo-EM and biochemical methods. In addition, the structure of STNV virus-like particles is probed by X-ray crystallography. In Chapter 2 the theory behind the methods used is explained, and Chapter 3 gives detailed materials and methods for all experiments in this thesis.

In Chapter 4, the three-dimensional cryo-EM reconstructions of native and expanded TCV are discussed and the properties of the expanded particle are explored by biochemical studies. The results of these experiments lead to a new model for TCV uncoating when placed in the context of the viral life cycle and virus-host interactions.

In Chapter 5, a three-dimensional cryo-EM reconstruction of expanded, empty particles of ERAV is described alongside that of native ERAV. A model is proposed for the transition between these states and for the uncoating process of ERAV.

In Chapter 6, RNA-protein interactions of STNV are further studied in virus-like particles which, instead of the asymmetric genome, contain 60 copies of an RNA stem-loop that binds with high affinity to the coat protein. By using low-resolution X-ray data, density in the RNA region could be seen.

Chapter 7 gives a general summary of the results and places the experimental findings from this work in the context of available data from other viruses.

2 Theory and techniques

Many processes in biology happen on a scale that is too small to see with the naked eye. Since the Enlightenment, the aspiration to see smaller things has driven the development of microscopes. However, light microscopy is limited by the wavelength of visible light to objects larger than about 200 nm. To image smaller objects, theoretically a shorter wavelength is required. As photons with a shorter wavelength are not visible, electron microscopy and X-ray diffraction are used to image smaller objects (Figure 2.1). Electron microscopy and X-ray crystallography have been used in this work and will be discussed in section 2.1 and 2.2, respectively.

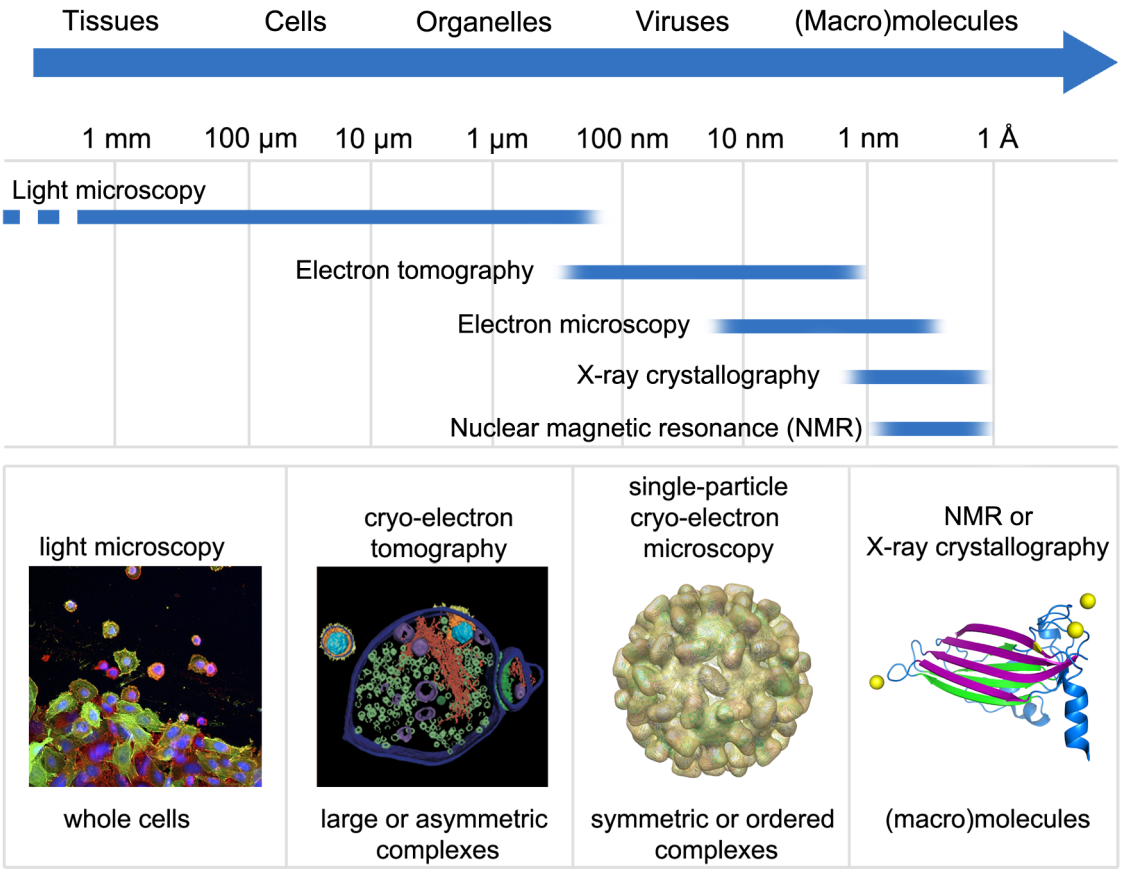


Figure 2.1 Resolution ranges in structural biology
The light microscopy image was taken from the Wellcome Image Library (B0006844). The tomography image is from Maurer *et al.*, 2008.

2.1 Transmission electron microscopy

Electron microscopy directly images the sample by recording the scattering of electrons by different areas of the sample, like light microscopy records the scattering

of visible light by the sample. Electrons, however, are so easily absorbed or scattered by matter, including atmospheric gases, that the electron source, sample and detector must be kept under high vacuum. The sample also has to be kept very thin. Electrons are scattered by the sample, so that where the sample is thicker or denser, fewer electrons reach the detector (Von Borries and Ruska, 1939). Because electrons are charged, they can be focussed by magnetic lenses, in much the same way as light is focussed by glass lenses in a light microscope (Figure 2.2).

Ernst Ruska and Max Knoll built the first prototype electron microscope in 1931, followed in 1933 by the first electron microscope that gave better definition than a light microscope. During the machine development by Ernst Ruska, his brother Helmut Ruska worked on the medical and biological applications. Both brothers were funded by Siemens, who held the patent (Ruska, 1986). Tobacco mosaic virus and potato virus X were imaged as early as 1939 (Kausche *et al.*, 1939), despite the low contrast of biological samples and their susceptibility to damage by the electron beam (Ruska, 1939). The susceptibility of biological samples to damage by the high energies imparted by the electrons means only a limited electron dose can be used before the sample degrades, and this results in a noisy image. At the same time, the low atomic number of the atoms in biomolecules means they scatter electrons only weakly, and the scattering of the molecules is similar to that of the plastic, carbon or vitreous ice support. Taken together, this results in low contrast, noisy images of biological samples in an electron microscope.

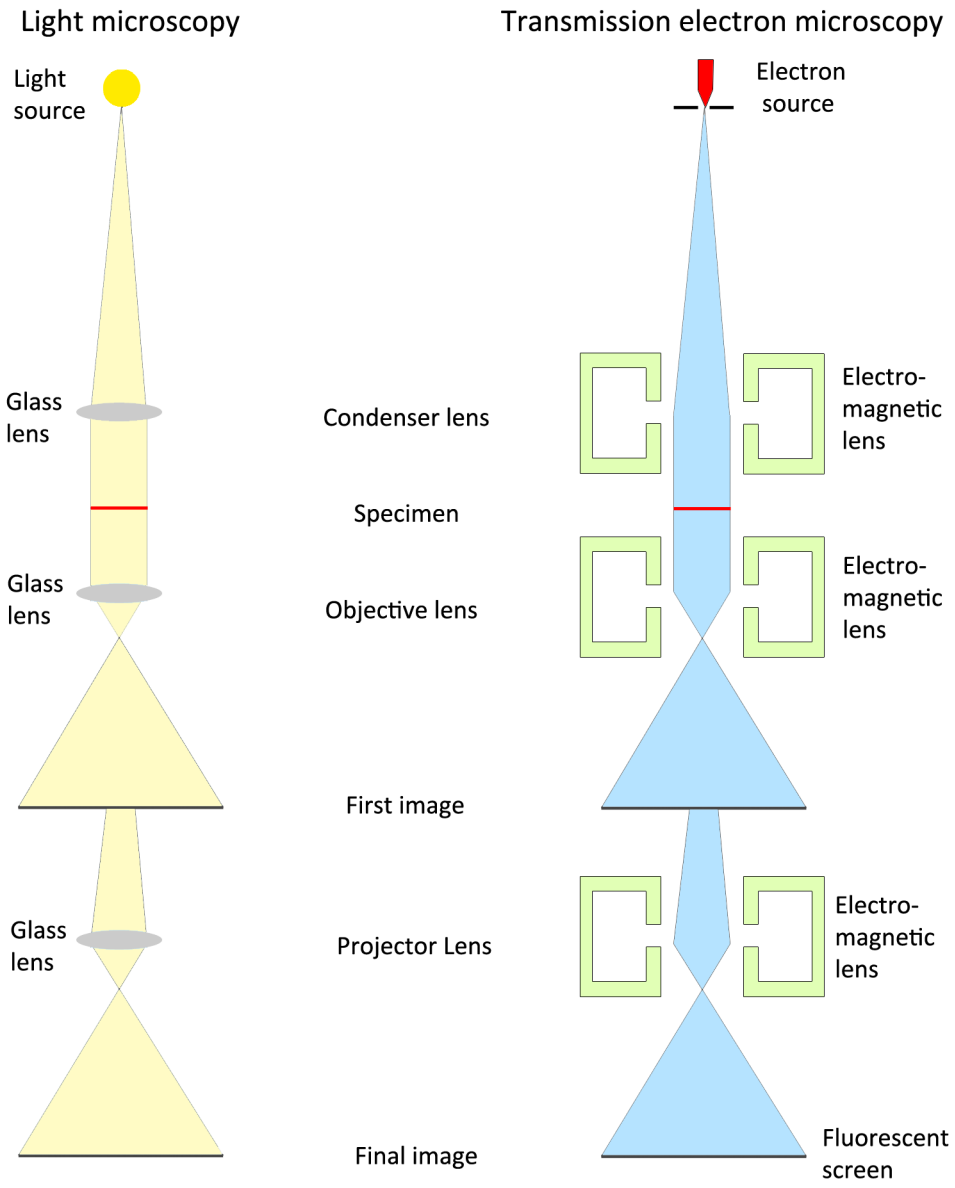


Figure 2.2 Comparing electron microscopy and light microscopy
Schematic representations of a light microscope (left) and transmission electron microscope (right) showing the similarity in components. The light source emits visible light (400-750 nm), which is focussed by glass lenses. The resolution is limited to about 200 nm. In the transmission electron microscope, the electrons are emitted from the electron source and accelerated to 80-200 keV (depending on the type of microscope). The beam is made parallel by the condenser lens, and passes through the specimen. After the specimen the beam passes through the objective and projector lenses to generate the final image. Image adapted from Chescoe and Goodhew, 1990.

In the remainder of this section, I will first describe the principles of image formation in the electron microscope, followed by the two main methods of sample preparation, and finally the image processing and 3D reconstruction process.

2.1.1 The electron source

Two main types of electron source are in use for electron microscopes: thermionic cathodes and the field emission gun (FEG). The thermionic cathode works on the principle that metals allow electrons to exit if they are heated to a sufficiently high temperature. The simplest form is a hairpin-shaped tungsten filament, but using a crystal of lanthanum hexaboride (LaB_6) enhances the escape of electrons and gives a brighter electron beam (Frank, 2006).

Electrons with different energies behave differently in a magnetic field (similar to light with different wavelengths behaving differently in a lens), meaning they get focussed in a different plane. Therefore, an electron beam with a small spread in the energies of the electrons, or monochromatic beam, results in sharper images. Similarly, electrons emitted from slightly different points on the source will behave differently, so a small source size where all electrons apparently exit from the same point, or coherent beam, will give better images (Frank, 2006).

With a FEG electron source, the electrons are abstracted from the FEG tip by an electric field, instead of heating. This results in electrons with a much narrower spread of energies, creating a much more coherent and monochromatic beam. The routine use of electron microscopes equipped with a FEG has dramatically improved the resolutions obtainable for 3D reconstructions of macromolecules (Frank, 2006).

2.1.2 Image formation

Image formation in an electron microscope is a complicated process. In brief, the beam leaves the electron source and passes through the condenser lens, which makes the beam parallel, and condenser aperture. It then passes through the sample, followed by the objective lens and aperture, which focus the inverse image and eliminate electrons that have been very widely scattered, respectively. The final magnified image is generated by the projector lens and projected onto a fluorescent screen or detector.

Two types of contrast are generated in the electron microscope (Figure 2.3). Amplitude contrast is generated when the electrons are inelastically scattered by the sample and the sample absorbs some of the energy of the electrons. The deposited energy causes changes in the sample, such as breaking of bonds, release of atoms or charging. Amplitude contrast can be enhanced by removing the electrons that were scattered to large angles by using a small objective aperture.

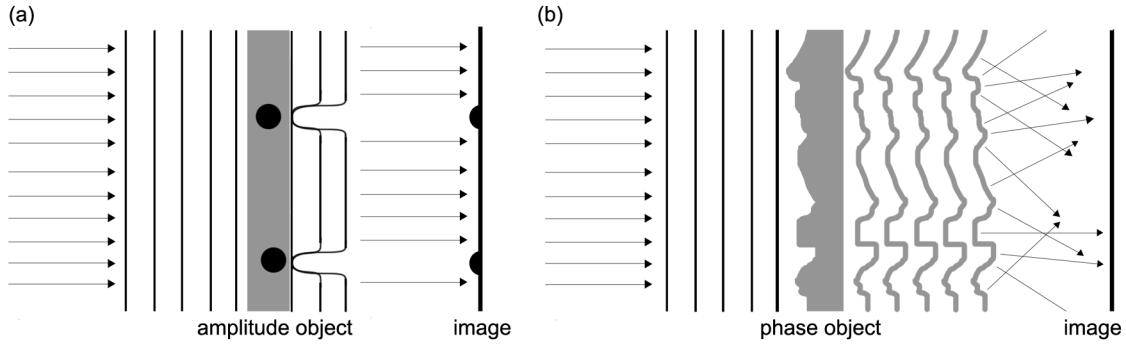


Figure 2.3 Image formation in the electron microscope
(a) Formation of an amplitude image. Parallel waves fall on the specimen, which is not transparent in all places, and the light is stopped, resulting in dark and light spots on the image. (b) Formation of an image from a phase object. Parallel waves fall on the specimen, which has variations in density (shown as variations in thickness). The phase of the emerging wave is changed and leads to interference with the waves that passed through the sample unchanged. This interference results in contrast in the image.

Light atoms, such as carbon, nitrogen and oxygen, do not have a large electron scattering factor and therefore do not scatter the electron beam strongly. This means macromolecules and the carbon support film show very little amplitude contrast. Therefore, a more useful image in macromolecular electron microscopy is created using phase contrast. Phase contrast is formed by interference of elastically scattered electrons with unscattered electrons. It arises from small imperfections in the lenses, such as spherical aberration, which causes peripheral parts of the beam to be focussed in a plane other than the focal plane. Phase contrast is also created if the image is not exactly in focus. In macromolecular electron microscopy, underfocus (generally referred to as defocus) is generally used. Increasing the defocus generates more contrast at the cost of high-resolution information, as the information at higher spatial frequencies is occluded by the more frequent phase shifts through inaccurate determination of the contrast transfer function (see section 2.1.5.1). Increased sample

thickness increases the chance of multiple interactions with the sample or inelastic scattering of the electron, and therefore leads to lower contrast.

2.1.2.1 Low-dose electron microscopy

Electrons interact strongly with matter, so the electron beam causes changes in the object during imaging, especially if the sample is a biological specimen embedded in ice. On exposure to the electron beam, the ice warms up and can change or melt, while the embedded molecules move and become damaged. This is evident in micrographs that have been taken using too high a dose, defined as total accumulated number of electrons per unit surface area of sample. Higher electron doses result in a better signal-to-noise ratio (SNR), but cause damage to the specimen, while low doses of electrons give a low SNR, which makes the particles hard, or even impossible, to see. A balance has to be struck at an electron dose where one can see the particles, but not damage them. A sensible upper limit is usually taken to be 10-20 $e^-/Å^2$. This means more complicated imaging protocols are needed for data collection in cryo-EM than the relatively straightforward point-and-shoot approach used in negative stain EM, where the heavy metal stain is less sensitive to beam-induced damage and change. A lower magnification results in a lower electron dose as the same beam (number of electrons) is spread over a larger physical area. Therefore, in the low dose protocol, a low (~3000-5000x) magnification is used to identify areas of the grid worth imaging. Once an area has been identified, the appropriate defocus is selected at a magnification high enough to clearly see the amorphous carbon film (>100,000x). During this process the beam is diverted away from the direct optical path, which would otherwise have crossed the sample and delivered a high dose of electrons to a small area. The beam is interrupted (blanked) during adjustment of the beam, lifting of the fluorescent screen and preparation of the camera (see below) to prevent alteration of the area to be recorded. Once the preparation is complete, the beam is allowed to go

through the sample and to the recording device at the desired magnification for the duration of the exposure time only.

2.1.3 Image recording

2.1.3.1 Film

Traditionally, electron micrographs have been recorded on photographic film, which consists of a cellulose layer covered with a photosensitive emulsion of silver halide salts. When electrons hit the emulsion, a latent image is formed that is made visible by the process of development, which turns the exposed silver halide into metallic silver that appears black on the film. After development, the film has to be digitized before it can be further processed. Digitization is done using a microdensitometer, which measures the relative density of the silver on the film and saves the image as a digital file.

2.1.3.2 CCD

A CCD camera for electron microscopy consists of a charge-coupled device (CCD) covered with a phosphor layer. The CCD does not detect electrons directly: the phosphor layer emits light when excited by electrons and this light is detected by the CCD and saved as a digital image.

The main advantage of using a CCD camera is that the resulting images are digital and therefore the need for film digitisation are avoided. This means a great increase in convenience and speed and a decrease in price. Unlike film cameras, CCDs have an unlimited capacity so they are well-suited to automated data collection and large datasets.

Current CCD systems, however, have a number of limitations. The pixel size of CCD cameras is 15 μm, corresponding to 3 Å/pixel in an image recorded at 50,000x magnification. A scanned film micrograph recorded at 50,000x and digitised using a scanner with a 7 μm pixel size results in a pixel size of 1.4 Å/pixel in the digital image.

Therefore, a higher magnification needs to be used to obtain a CCD image with an equivalent pixel size, reducing the number of images per micrograph.

Additionally, there is a larger amount of noise in the micrographs. The energy of the electrons is not completely absorbed when they first hit the detector. Instead, they are scattered within the phosphor layer and cause further excitation in neighbouring pixels, resulting in noise in the image and limiting the resolution (Faruqi and Henderson, 2007).

2.1.3.3 Direct electron detectors

The indirect way electrons are detected causes additional noise in the resulting images. Because of this, the development of direct electron detectors for use in electron microscopy has been a priority in the field for a long time. These detectors are thinner than CCD cameras, which means there is less electron scattering within the detector and therefore a better SNR. They are now commercially available (FEI Company, 2009).

2.1.4 Sample preparation

2.1.4.1 Heavy metal staining

Contrast was initially improved by heavy metal shadowing, where the sample was coated with a heavy metal by evaporation. Additionally, early support films made of Formvar or similar materials degenerated rapidly in the electron beam. Carbon films, produced by evaporating carbon onto mica in a vacuum and floating it onto EM grids, proved less fragile (Bradley, 1954). The subsequent development of negative stain, which uses a solution of heavy metal ions, improved contrast and allowed resolution of features that could not be distinguished using metal shadowing (Brenner and Horne, 1959). These procedures, virtually unchanged, are still used widely today.

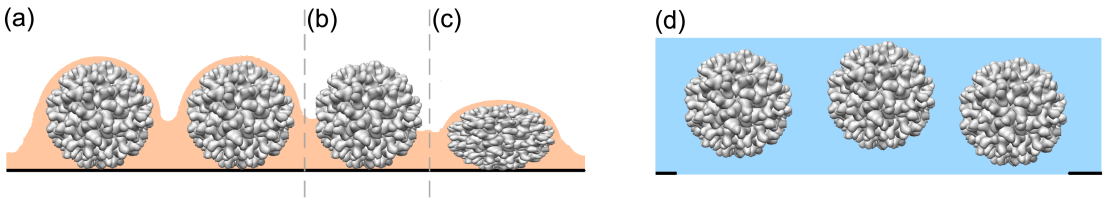


Figure 2.4 Heavy metal staining and embedding in vitreous ice of virus particles
(a) Virus particles embedded in heavy metal stain (orange shading) supported on continuous carbon film. (b) Partially stained virus particles on continuous carbon film. (c) Flattened negatively stained virus particle on continuous carbon film. (d) Unstained virus particles embedded in vitreous ice (blue shading) supported on holey carbon film.

To obtain a negatively stained sample, the biomolecules are deposited on a grid that has been coated with a thin layer of amorphous carbon. Excess liquid is blotted away and the grid is then flooded with a solution of a heavy metal salt, such as 1-4% (w/v) uranyl acetate or phosphotungstic acid and dried. The metal salt occupies the spaces on the support film where there is no biomolecule, leaving a low-density imprint of the biomolecules against a high-density metal background (Figure 2.4a). This has the advantage that, even if radiation damages the biomolecule, its outline is preserved in the dried heavy metal salt. There are two main artefacts that can be introduced using this method. Firstly, especially for larger biomolecules such as viruses, it is possible that the stain only partially covers the molecule (Figure 2.4b). This results in an incomplete projection image and limits the use of the images for three-dimensional reconstructions. Secondly, negative staining and air-drying frequently results in flattened molecules (Figure 2.4c), which will give a distorted reconstruction. Additionally, molecules may adopt preferred orientations on the carbon film, which potentially limits the quality of reconstruction, especially for asymmetric molecules.

2.1.4.2 Vitreous ice

Biomolecules naturally occur in a highly crowded but basically aqueous environment. However, the volatility of water makes it unsuitable for use in the vacuum environment of the electron microscope. Various substitution methods have been tried, for example embedding in glucose, but the similarity in scattering of glucose and macromolecules results in very low contrast. A revolutionary advance was made when the equipment and methodology for producing EM specimens in vitreous ice were developed. The majority of the technological advances were made in the pursuit of electron diffraction of catalase crystals (Taylor and Glaeser, 1974), including adapted specimen holders and holder insertion mechanisms (reviewed in Taylor and Glaeser,

2008). This was shortly followed by the development of ultrafast freezing methods, which enabled the embedding of macromolecules in vitreous ice (Adrian *et al.*, 1984). Vitreous ice does, however, pose several challenges. Firstly, the aqueous solution must form thin enough films to allow observation in the electron microscope. Secondly, the grid containing the thin films must be frozen quickly enough to prevent formation of crystalline ice. Thirdly, during any subsequent storage, handling and examination in the microscope, the grid must be kept cold to prevent the vitreous ice turning crystalline, while preventing contamination by moisture from the surrounding air. Finally, a more concentrated sample is usually required for cryo-EM than for negative stain. However, the challenges are easily outweighed by the advantages of higher resolution, visibility of internal features, a more natural environment, fewer artefacts and minimal effect of preferred orientations.

2.1.5 Image pre-processing

2.1.5.1 Contrast transfer function (CTF) correction

Image formation in the electron microscope and the generation of phase contrast were described in section 2.1.2. The contrast transfer function (CTF) describes the transfer of information in the microscope as a function of resolution (Equation 2.1). The real-space effect of the CTF is to act as a point-spread function (PSF), which is convoluted with the image.

Equation 2.1 CTF by weak phase approximation

$$CTF(s) = \sin \gamma(s)$$

$$\gamma(s) = 2\pi \left[(C_s\lambda^3/4)s^4 - (\Delta z/2)s^2 \right]$$

Where C_s is the spherical aberration coefficient, ΔZ the defocus, s the spatial frequency (resolution) and λ the electron wavelength (Downing and Glaeser, 2008).

As the shape and periodicity of the CTF, and hence the PSF, is dependent on the defocus, images taken at different defocus are difficult to compare directly before CTF correction has been applied (Figure 2.5).

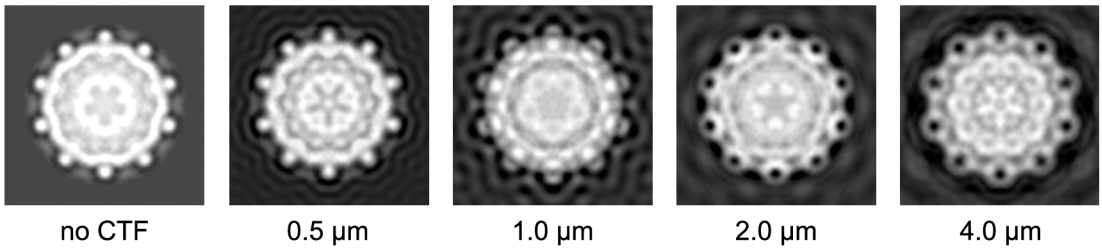


Figure 2.5 CTF at different defocus and its effect on the image
A projection image without CTF (far left) and the same image convoluted with a CTF with increasing defocus. The image is less recognisable with increasing defocus.

Additionally, as the value of the CTF is zero at certain spatial frequencies (resolutions), no information is available for those spatial frequencies. This means a reconstruction made from images with the same defocus will have information missing at certain resolutions. Therefore, information from multiple micrographs is combined, not only to improve the SNR but also to ensure there is information at all spatial frequencies.

As the effect of the CTF is a physical property of the microscope, it should ideally be calculated from many micrographs taken at the same defocus. In practice, power spectra (Fourier transforms) are calculated for sections of a micrograph and averaged. The average power spectrum is assessed for regularity, as drift or axial astigmatism can cause the power spectrum to be irregular or ellipsoid, and this can be used as an initial measure of image quality. The power spectrum is rotationally averaged and a CTF curve, using standard values (C_s, λ) for the microscope in question with the defocus (ΔZ) as variable, is fitted to the resulting pattern.

Once the CTF has been determined, there are several methods of CTF correction. The quickest approach is phase flipping, in which the phase of those parts of the CTF with negative information transfer is changed by 180° so information transfer is greater than or equal to zero at all frequencies (Figure 2.6). More elaborate methods, e.g. Wiener filtering, correct not only for the phase of the CTF but also the amplitude. However, at the frequencies where the CTF is zero, the information cannot be restored (Frank, 2006).

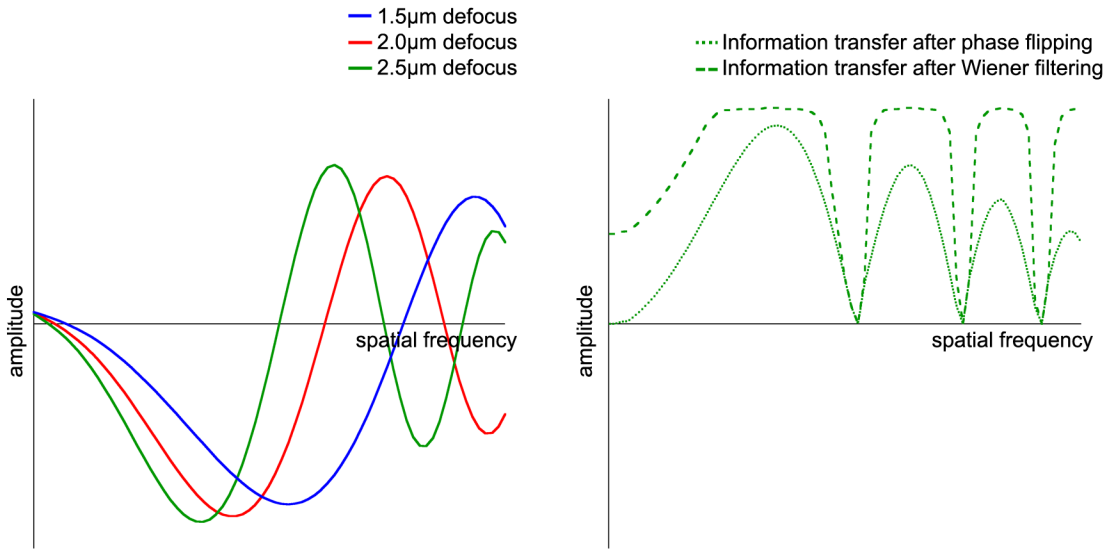


Figure 2.6 CTF correction by phase flipping and Wiener filtering
Left, three CTF functions are plotted with amplitude against spatial frequency, with defocus values of 1.5 µm, 2.0 µm and 2.5 µm. The right-hand panel shows the information transfer after CTF correction by phase-flipping (dotted line) and Wiener filtering and amplitude correction (dashed line) for a defocus of 2.5 µm.

2.1.5.2 Image selection and centring

For further processing, individual particle views have to be selected and cut out from the micrograph. Various programs exist that help the particle selection by computational methods, such as the detection of particle edges or correlation with a reference particle set (Zhu *et al.*, 2004; Roseman, 2004), although for some heterogeneous or low-contrast samples, manual or semi-automated particle selection can be required.

To obtain good results in classification (see below) and speed up alignment steps, it is useful if the particles are well centred within the image boundaries.

2.1.6 3D reconstruction

The fundamental problem in obtaining a three-dimensional reconstruction from cryo-EM images is determining the correct orientations for noisy particle views. Raw micrographs have a low SNR, but as the noise is random, averaging many images strengthens the signal while reducing the noise. As more raw images are averaged, the SNR will rise. The average image is then used for further processing. Once images have been selected, centred and divided into classes, 3D reconstruction of the

molecule can be started. Three-dimensional reconstruction can be carried out in either real or Fourier space.

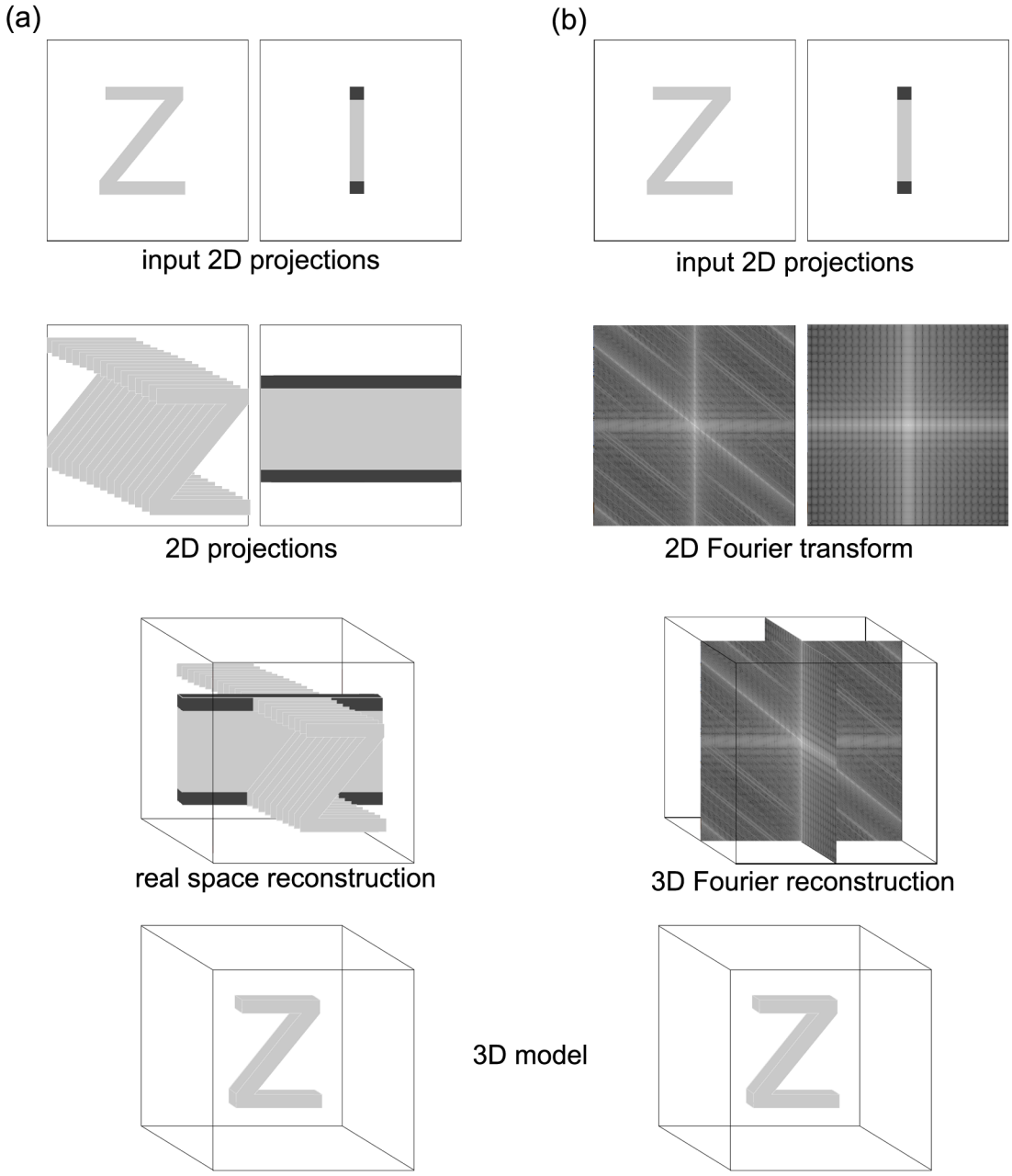


Figure 2.7 Fourier space and real-space based 3D reconstruction methods
(a) Real space 3D reconstruction. The two-dimensional (2D) projections are smeared out and combined at the appropriate angles. Parts of the reconstruction where more of the smeared 2D projections coincide form part of the final 3D model. (b) Fourier-based 3D reconstruction. Fourier transforms of the 2D projections are combined at the appropriate angles. A Fourier transform of the combined 2D reconstructions gives the 3D reconstruction in real space. Only 2 projections per model are shown for clarity, but an ideal reconstruction would include 2D projections for all orientations and therefore maximally sample the 3D transform of the object.

Fourier space reconstruction uses Projection Theorem, which states that the Fourier transform of a projection of a 3D volume is the same as a central section of the Fourier

transform of that 3D volume. Therefore, the volume can be reconstructed from projections as long as the orientation of the particles is known, and the SNR is sufficiently high.

Real-space refinement takes the averages of classified images, which represent projections of the molecule, from a number of angles and works out what arrangement of electron density would lead to that combination of projections. This is best done in an iterative way: the program makes an initial guess of the density and projections are calculated, then the experimental images are compared to the projections and the density is calculated again. This process is analogous to a puzzle game where a grid is given with a series of numbers denoting the number of filled squares in the rows and columns (Figure 2.7).

Both methods benefit from a large number of evenly spaced orientations. More orientations lead to a smaller difference in angle between the orientations, or angular spacing ($\Delta\theta$, Figure 2.8). Fourier space 3D reconstruction is generally less computationally intensive than 3D reconstruction in real space.

2.1.6.1 The use of symmetry

If the molecule of interest exhibits symmetry, this will simplify the 3D reconstruction process greatly. When an *n*-fold symmetrical model is rotated in x steps, the x/*n*th orientation will appear identical to first. This means that for the same number of images, *N,* with *n*-fold symmetry, the effective number of particles will be *n*-fold higher, or *N***n*, because each image orientation is used *n* times in the 3D reconstruction (Figure 2.8).

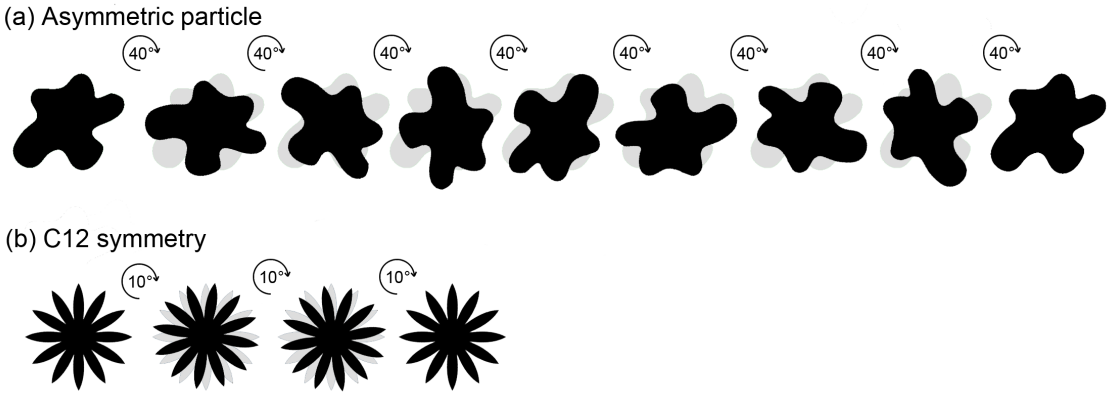


Figure 2.8 Effect of symmetry in 3D reconstruction
The figure shows a 2D 'molecule' for clarity. (a) For an asymmetric molecule, a full 360° rotation is required to arrive at an identical orientation. (b) For a molecule with C12 symmetry, 30° rotation results in an identical orientation. This means the same class average can be used twelve times in the 3D reconstruction process and therefore fewer particles need to be collected for a molecule with high symmetry to get the same SNR.

This means a symmetrical molecule can have a smaller angular spacing (difference in orientation between class averages) than an asymmetrical molecule and still have the same number of images, and therefore SNR, per class average.

2.1.7 Obtaining a starting model

The first step in 3D reconstruction is obtaining a starting model. After a starting model has been obtained, it can be refined by including more of the dataset. The starting model is projected to obtain 2D reference images using known orientations, and the experimental images are aligned to the reference images, a process called reference-based alignment. The orientation of each experimental image is assigned as the orientation of the reference image that it resembles most. A new model is made using average images for each orientation and this model is used as the starting model for the next round of refinement.

The starting model can be obtained by one of the following methods.

2.1.7.1 Starting model from an existing structure

If a crystal structure or previous microscopic reconstruction is available, this can be converted to a suitable format and used directly as a starting model for refinement. To

reduce the risk of the final model being biased by the starting structure, high-resolution information is removed from the model by application of a Fourier filter (Figure 2.9a).

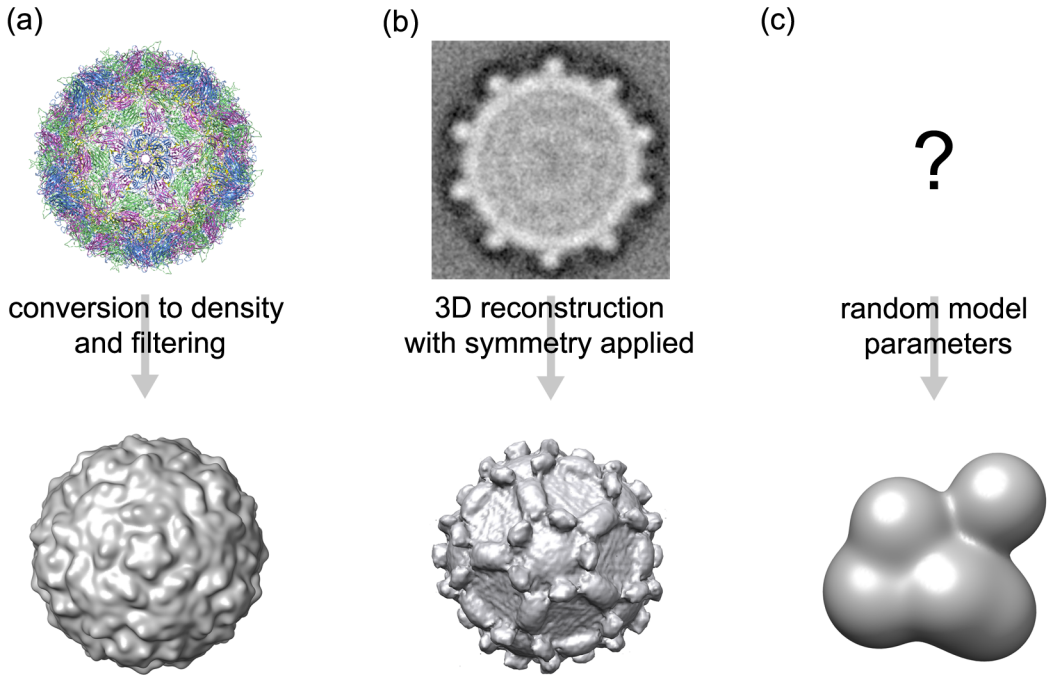


Figure 2.9

Figure 2.9 Obtaining a starting model
(a) From a known (crystal) structure. Shown is the crystal structure of poliovirus (PDB entry 1hxs) in ribbon representation, and the same structure converted to density and filtered to 30 Å. This model was used as a starting model for the ERAV reconstructions discussed in Chapter 5. (b) A starting model with high (icosahedral) symmetry obtained from a single class average. This model was used as a starting model for the TCV reconstructions discussed in Chapter 4. (c) A starting model generated from random parameters using EMAN (Ludtke *et al.*, 1999).

2.1.7.2 High-symmetry starting model

When the molecule of interest has high symmetry, such as an icosahedral virus, a starting model can be obtained from a single image class. The images are classified as described in section 2.1.6.2 and divided into a number of classes. A suitable molecular view is then selected and used for back-projection with the appropriate symmetry applied. This results in a starting model with the required symmetry, but without model bias, that can be used for refinement. Even a sphere with the required symmetry applied will be featured enough to be used as a starting model for icosahedral viruses. Similar techniques are applied when creating models of helical structures (Figure 2.9b).

2.1.7.1 Random starting models

A third method of generating a starting model is to use a random arrangement of density with the desired size and symmetry. If several independent starting models result in similar final models, one can be confident the final model is unbiased (Figure 2.9c).

2.1.7.2 De novo *starting model generation*

If the model does not have high symmetry and there is no existing structure, or model bias is to be averted, a starting model can be obtained directly from the data using averaged images with appropriate angles. In order to do this, the orientations of the images have to be determined with sufficient accuracy, which can be difficult. The relative orientations of two views can be determined using correlation between 1D projections of the images in a process called sinogram correlation (Frank, 2006). Alternatively, the data collection strategy can be adjusted: one can generate known differences in orientation by tilting the sample in the microscope and taking images of the same area in different orientations.

2.1.8 Alignment and classification

As described in section 2.1.6, a single particle image does not have a high enough SNR to accurately determine the orientation for use in 3D reconstruction. Alignment and classification are therefore key processes in all single-particle electron microscopy applications. During alignment, the image is rotated and translated in two dimensions to ensure its orientation corresponds to that of the other particles to be averaged. Classification can be used to distinguish between different views of the same particle within the dataset, or to separate different conformations of similar views. It can also be used to discriminate malformed or degraded particles within the dataset, so they can be excluded from further processing.

When reference images are available, the experimental images can be classified by calculating their correlation with the reference images. All images are aligned to all

available reference images and a correlation coefficient is calculated for each image with each reference image. Each reference image represents a class of particles, and each test image is classified based on the reference image with which it has the highest correlation coefficient.

2.1.8.1 Multivariate statistical analysis

A useful way of classifying different views without the bias of reference images is by multivariate statistical analysis. For this analysis, the total variance between the images is decomposed to mutually exclusive components that are ordered by decreasing magnitude (Frank, 2006). In more human terms, this means the average of all images is calculated. Then eigenimages are generated for each area where there is variation in the image (such as particle size). These are made in such a way that the images either have or lack the feature described by the eigenimage (Figure 2.10b).

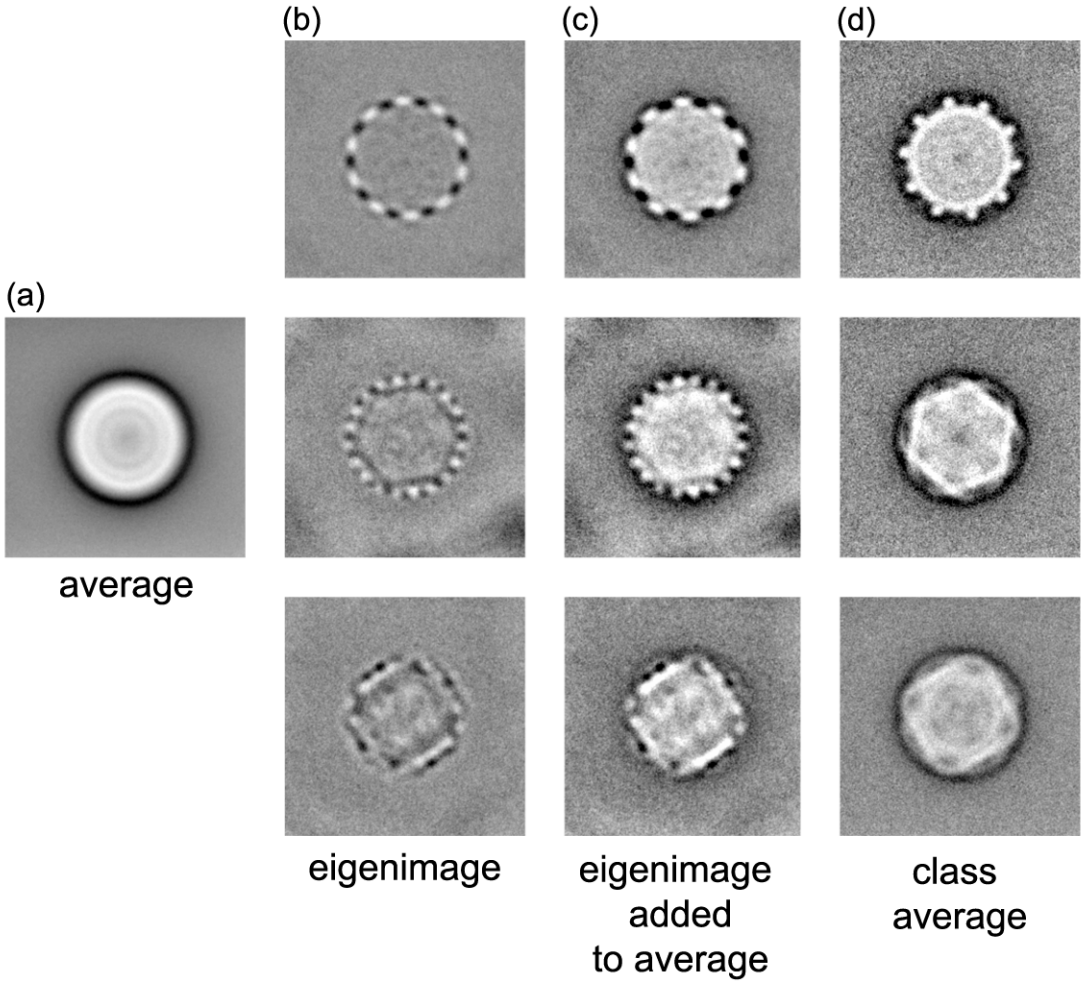


Figure 2.10 Multivariate statistical analysis
Showing the results of MSA classification for a TCV dataset. (a) The averages of all images in the dataset after centering. (b) Three of the 69 eigenimages. (c) The result of adding the eigenimages in (b) to the global average of the dataset. (d) class averages showing the features of the eigenimages.

Therefore, if the eigenimage is added to the average image, the resulting image is what the average would look like if all images had that property, while the other features remain blurred (Figure 2.10c). If all images lacked the feature, the average image would look like the average image with the eigenimage subtracted. The entire dataset can then be divided into classes of similar images, by determining per image whether it shows or lacks the features described by the eigenimages (Figure 2.10d).

2.1.9 Resolution estimation and model validation

There has been extensive discussion in the EM field about how best to calculate the resolution of cryo-EM reconstructions, and indeed how resolution is best defined. A commonly used method is Fourier shell correlation (FSC). Here, two models are made,

each using half the images from the dataset, and the correlation between these models is calculated as a function of resolution. The resolution where the correlation equals a certain cut-off value is quoted as the resolution of the model. Cut-off values of FSC = 0.5 (Böttcher *et al.*, 1997; Malhotra *et al.*, 1998) and FSC = $3\sigma^*\sqrt{n}$ (where n is the symmetry; Orlova *et al.*, 1997) have been used. The FSC = $3\sigma^*\sqrt{n}$ cut-off results in significantly higher resolution estimates; for the same reconstruction of keyhole limpet hemocyanin, the resolution was 12.8 Å using FSC = $3\sigma^*\sqrt{n}$, but only ~20 Å if FSC = 0.5 was used as the cut-off value (Orlova *et al.*, 1997). It has, however, been suggested that using FSC = 0.5 as the cut-off results in an underestimation of the resolution, as only half the data is used for each model, thereby decreasing the signal-to-noise ratio of the models as fewer images are averaged (Rosenthal & Henderson, 2003).

In addition to resolution testing, the quality of the reconstruction can be assessed by comparison with known structures, determined by other methods, such as a crystal structure. If no crystal structure is available, agreement with biochemical experiments and/or fitting of available structures of domains or related proteins may help validation and interpretation of the map. For instance, a known antibody binding site should be accessible to antibody in the final model or residues shown to be close by cross-linking studies should be close together in the model.

2.2 X-ray crystallography

X-ray crystallography uses the diffractive properties of crystals when placed in an X-ray beam to determine the structure of the material within the crystal. The crystal is placed between an X-ray source and a detector (Figure 2.10) and rotated to record information from different angles. The resulting set of diffraction patterns can be used to determine the structure of the molecules in the crystal.

The history of macromolecular X-ray crystallography can be told through a series of Nobel prizes. The first Nobel Prize in Physics in 1901 was awarded to Wilhelm Conrad Röntgen for the discovery of X-rays (Odhner, 1901). The nature of X-rays as a form of

electromagnetic radiation and the diffraction of X-rays by crystals resulted in a Nobel Prize for Max von Laue in 1914, only two years after the publication of his seminal paper (Von Laue, 1915; Friedrich *et al.*, 1912). The next Nobel Prize, in 1915, was for the father-son team of William Henry and William Lawrence Bragg, for extensive application of X-ray crystallography and the definition of Bragg's Law, which is described in section 2.2.1 (Bragg, 1922). The size of the molecules that could be tackled grew steadily until the determination of the first protein structures led to Nobel Prizes in Chemistry: one was awarded to Max Perutz and John Kendrew in 1962, for solving the crystallographic phase problem (see section 2.2.6.2) and the structures of haemoglobin and myoglobin (Perutz, 1962). The Nobel Prize for Physiology or Medicine in the same year went to Francis Crick, James Watson and Maurice Wilkins for the structure of DNA, which was based on X-ray fibre diffraction pictures taken by Rosalind Franklin (Crick, 1962). Another Nobel Prize was awarded in 1964 to Dorothy Crowfoot Hodgkin for the structures of many biologically relevant molecules, which we would now consider 'small molecules', including penicillin (fewer than 40 atoms) and insulin (Hodgkin, 1964). The 1988 Nobel Prize in Chemistry was awarded to Hartmut Michel, Robert Huber and Johann Deisenhofer for the determination of the three-dimensional structure of a photosynthetic reaction centre (Huber, 1988). In 2009, the Nobel Prize in Chemistry was awarded for the crystal structure of the ribosome, with nearly 400,000 atoms, to Venkatraman Ramakrishnan, Thomas Steitz and Ada Yonath (Ramakrishnan, 2009).

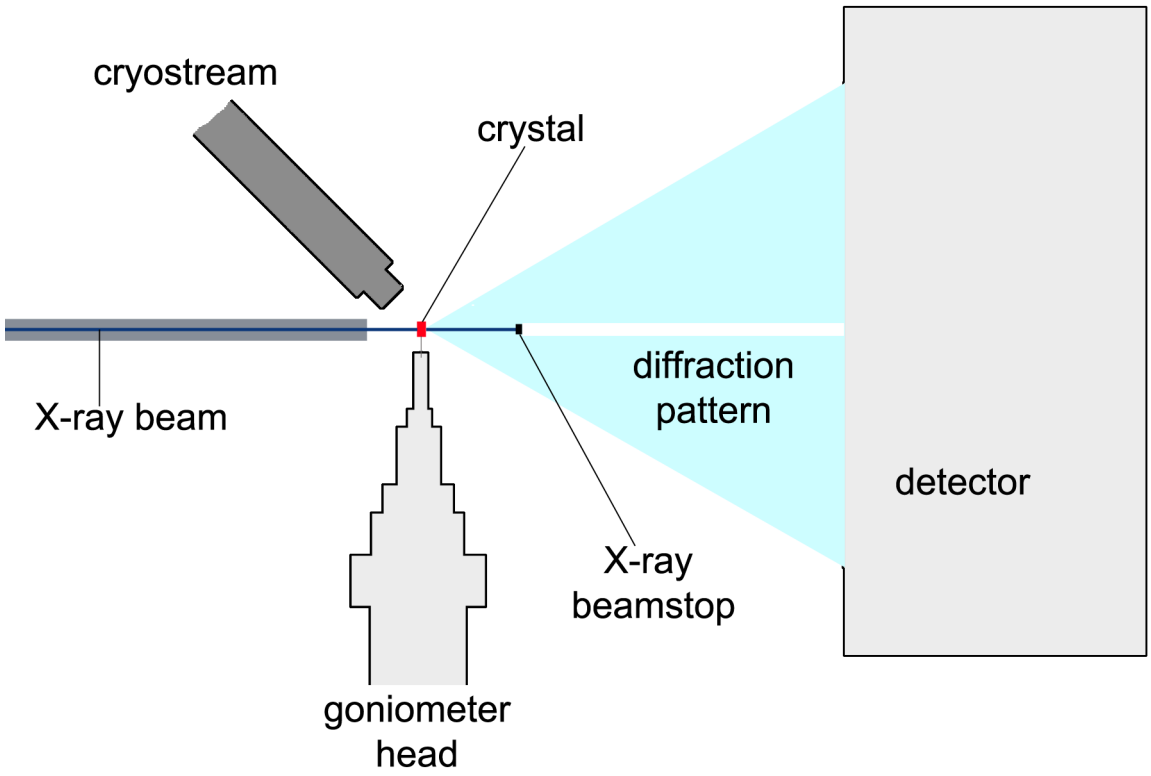


Figure 2.11 Schematic representation of an X-ray diffraction setup
The crystal (red rectangle) is mounted on the goniometer head, and cooled by cold nitrogen gas from the cryostream. Part of the X-ray beam is diffracted by the crystal; the part that is not diffracted hits the beam stop. The diffraction pattern is recorded by the detector.

2.2.1 X-ray diffraction

A crystal consists of molecules packed into an ordered lattice. Such an array of molecules will scatter a beam of appropriate wavelength, such as X-rays or electrons, in the same way that an ordered 2-dimensional pattern will diffract visible light. Constructive and destructive interference of the scattered photons results in a discrete pattern of spots, known as the diffraction pattern, with the spatial arrangement of the spots determined by the crystal lattice and their intensity by the contents of the repeating unit of the lattice (the unit cell). In other words, the diffraction pattern is the Fourier transform of the unit cell contents multiplied by the Fourier transform of the lattice (see section 2.2.6.2). The diffraction pattern, and therefore the resulting structure, shows the average of all the unit cells in the crystal, in the same way as a cryo-EM reconstruction shows the average of the particles used.

Most of the X-ray beam does not interact with the atoms in the crystal at all, and exits unchanged. A beam stop is used to prevent damage to the detector from this strong direct beam (Figure 2.10). The part of the beam that does interact with the cell can be scattered elastically or inelastically, like the electron beam in EM. Inelastic scattering causes energy to be transferred to the molecule and causes radiation damage and noise. If an X-ray is scattered elastically within the crystal, it is diffracted, producing a new wave with the same frequency and a different phase. When all the diffracted waves are added, phase differences cause interference generating the diffraction pattern. If the waves are in phase ($\Delta\varphi = n\lambda$), the peaks are reinforced (constructive interference), but if the waves are out of phase, they cancel out (destructive interference).

Diffraction can be thought of as the reflection of the incoming X-ray beam by a series of planes separated by a distance *d*. Maximum constructive interference, and therefore diffraction, is obtained when $n\lambda = 2d\cdot\sin\theta$, i.e. when the phase difference between two interfering electromagnetic waves is a multiple of the wavelength (Bragg's Law). In other words, diffraction spots are seen where the scattering vector **S** and the distance vector between two identical atoms in different unit cells **r** multiply to an integer *n.* In a crystal the distance vector is identical to the unit cell length. More repeats of the unit cell in the scattering volume, i.e. a larger crystal, will result in a clearer diffraction pattern as the signal-to-noise ratio of the diffraction spots increases.

The highest-resolution data are scattered with the widest angle and appear at the edge of the detector. Low-resolution reflections are near the centre of the detector and may be masked by the beam stop. Considerations of resolution are further discussed in section 2.2.5.

2.2.2 Sample preparation

2.2.2.1 Crystallisation

Many molecules can be crystallised. The crystallisation of small molecules is often used as a purification procedure and usually involves making a supersaturated solution of the substance, by heating or change of solvent, which results in nucleation and crystal growth. Proteins and other macromolecules generally degrade when such methods are used, so an alternative strategy has been developed. The molecule of interest is purified and dissolved in an appropriate buffer. The solubility is then altered by adding precipitants, such as salts or polymers, and removal of water to saturate the solution. This results in nucleation and crystal growth. The water has to be removed in a controlled way to allow formation of ordered crystals while preventing precipitation. Vapour diffusion is a commonly used technique and involves equilibration of the protein/precipitant droplet against a reservoir without protein containing a higher concentration of the precipitant. Alternatively, the protein/precipitant droplet can be covered with mineral oil to reduce evaporation (microbatch crystallisation), crystals can be grown in a gradient of precipitant, or the concentration of the precipitant can be slowly changed by dialysis (Figure 2.11).

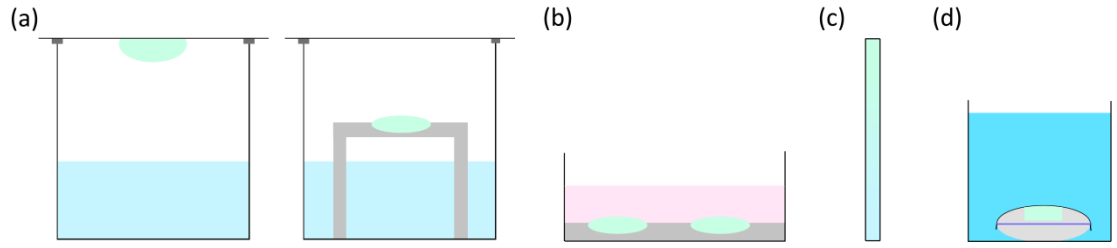


Figure 2.12 Protein crystallisation techniques
(a) Vapour diffusion: a drop containing protein and precipitant (green) is equilibrated against a reservoir (blue shading) with higher precipitant concentration, resulting in a net movement of water from the drop to the reservoir. This can be done as with the drop suspended over the solution on the coverslip (hanging drop, left) or with the drop on a plastic support (sitting drop, right). (b) Microbatch crystallisation: drops containing protein and precipitant are kept under mineral oil (pink shading) to slow evaporation. (c) Protein solution is layered with precipitant solution, resulting in a concentration gradient (green-blue). Crystals will form at the optimum conditions. (d) Dialysis of the drop (green) containing protein against a buffer containing precipitant (blue) slowly increases the precipitant concentration in the protein drop.

Viruses are crystallised using the same methods as for other macromolecules. However, because of their size and shape, the contact area between neighbouring molecules in the crystal is small relative to their mass. Virus crystals are therefore typically more fragile than those of other macromolecules.

The relevance of crystal structures to proteins in their cellular environment has been questioned because of the seemingly unnatural environment of the crystal. These concerns, however, are largely unfounded, as protein crystals contain a large amount of solvent between protein molecules, up to 70%, and the protein concentration in the cell approaches that in the crystal. Furthermore, various proteins occur naturally in near-crystalline arrays (Smith *et al.*, 1982; Subramaniam and Henderson, 2000), and many enzymes have been shown to be active in the crystalline environment (Makinen and Fink, 1977), showing the crystalline environment is not necessarily prohibitive to protein function.

2.2.2.2 Crystal mounting

Once a crystal has been obtained, it needs to be held in the correct position in the X-ray beam. To minimise incoherent scattering of the beam (which increases noise), it should be surrounded by as little material as possible. Stable crystals, such as those of salts and small molecules, can be glued directly to a holder, which is then placed in the X-ray beam. However, crystals of proteins and other macromolecules, including viruses, contain higher amounts of solvent and are correspondingly more fragile and sensitive to dehydration. Up to the 1990s, such crystals were mounted in sealed capillaries with a drop of mother liquor, not touching the crystal, on both sides to maintain humidity. This procedure was extremely time-consuming, especially when considering that radiation damage by the beam also limits the diffraction lifetime of the sample, so many crystals were needed to collect a complete dataset. A better approach is to mount the crystal in a suitably sized nylon loop and cool it to ~100 K (Hope, 1988). The crystal is held at this temperature during data collection using a

stream of dry nitrogen gas. The low temperature limits the movement of free radicals and thereby reduces secondary radiation damage processes (Garman, 2010). However, the freezing process can damage the crystals if the water in the crystals freezes to crystalline ice. This can be reduced by rapid freezing, as in cryo-EM, or by using cryo-protectants: molecules that prevent the formation of crystalline ice. Commonly used cryoprotectants include salts, sugars, polymers and alcohols. Finding the optimal cryoprotectant and concentration for the crystallisation conditions of a particular crystal can be challenging; on the other hand the protein may crystallise in a conditions that already contain enough cryoprotectant.

2.2.3 The X-ray source

2.2.3.1 Rotating anode X-ray source

A rotating anode X-ray source consists of a vacuum tube with a cathode, emitting electrons, and an anode receiving them. When the electrons collide with the anode, they excite the anode material and some of the energy is emitted as X-rays. The main advantages of a rotating anode X-ray source are that they are relatively compact. This means they are available as the 'home source' at many universities for crystal screening and data collection. The most common type of rotating anode for in-house crystallography uses a copper anode, which emits X-rays with the fixed Cu Kα wavelength of 1.54 Å.

2.2.3.2 Synchrotron X-ray beamlines

A beam of electrons when bent by a magnet loses energy in the form of photons. This was seen as a by-product in particle accelerators used in particle physics, but purpose-built synchrotron storage rings are now used to generate photons with a wide range of wavelengths and applications. Electrons move around the storage ring at close to the speed of light and are continuously steered by magnets, generating photons. To obtain more photons, multiple magnets can be placed along the electron beam path. Specific photon energies can then be selected by a monochromator.

The main advantages of synchrotron X-ray beamlines over rotating anode X-ray sources are their tunability and the higher intensity of the beam available. The higher intensity means shorter exposure times are needed compared to rotating anode X-ray sources and that smaller crystals can be used for structure determination. The tunability means datasets can be collected at different wavelengths and the difference between such datasets can be used to solve the phase problem (see section 2.2.6.2).

2.2.4 The X-ray detector

Detectors for X-ray crystallography are usually area detectors, i.e. flat detectors mounted some distance away from the crystal that detect the scattered X-rays (as shown in Figure 2.11). Originally, these used to be several layers of photographic film, that then needed developing and digitising, but these are no longer in use. Currently commonly used detectors are the image plate and the CCD detector. Image plate detectors consist of a X-ray sensitive phosphorescent layer. CCD detectors for X-ray crystallography, like those for EM, consist of a fluorescent layer that is coupled to a CCD chip with an optical link. CCD detectors have a much shorter readout step than image plate detectors, which can speed up data collection when short exposure times are used.

2.2.5 Data collection

During data collection, several factors need to be considered to obtain the best possible dataset for a given crystal. The optimal data collection strategy depends on the crystal properties, such as space group, diffracting power and crystal shape, as well as crystal content. Therefore, test images are collected prior to full data collection to determine the maximum resolution of the crystal and the space group.

The recorded dataset should be as complete as possible, as an incomplete dataset can result in density that is difficult or impossible to interpret. For crystals with high-symmetry space groups, less data is required to obtain a complete set of reflections. However, a certain amount of redundancy improves data quality as a high redundancy

in a dataset means there are more equivalent reflections to be averaged, resulting in more precise measurements. If the phases are to be determined using the anomalous scattering of X-rays by heavy atoms, the inverse wedge of reciprocal space will also need to be collected.

The maximum resolution to which a crystal diffracts the X-ray beam determines the distance from the crystal to the detector, as higher resolution reflections are further from the centre of the detector. Having a small detector distance will, however, also result in reflections being closer together and potentially overlapping. This is especially important in crystals with large unit cells, resulting in small distances between reflections, such as virus crystals. Mosaicity, or the imperfect alignment of domains within a crystal, makes individual reflections spread out and can also increase overlap of reflections. In addition, reflections at very low resolution can be obscured by the direct beam stop when the detector distance is small. As discussed above, low-resolution reflections can be extremely informative in virus crystallography as the asymmetric genome diffracts to a lower resolution than the highly symmetric coat protein. This means that no density for the genome can be seen when only high-resolution diffraction data (30-2 Å) is used, but the icosahedrally averaged genome structure can be obtained when low-resolution data (180-30 Å) is included (Tsuruta *et al.*, 1998, Lane *et al*., 2011). In some cases, it may be necessary to collect a high-resolution dataset with small detector distance and a low-resolution dataset with larger detector distance and merge these to obtain the best possible dataset.

A final consideration during data collection is radiation damage. The high-intensity X-ray beams used cause damage, such as reduction of redox centres, breaking of disulphide bonds and decarboxylation of acidic residues, even at liquid nitrogen temperatures, potentially resulting in misinterpretation of a structure or mechanism. This means a balance has to be struck between obtaining high completeness and redundancy of the dataset on one side and minimising radiation damage on the other (Nave, 1995; Ravelli and Garman, 2006).

2.2.6 Data processing

2.2.6.1 Integration

The raw data obtained from X-ray crystallography is a set of diffraction images. In order to get a structure, the images need to be processed. The first step in image processing is the determination of the dimensions and symmetry of the crystal lattice. The spacing of the crystal lattice is inversely correlated with the spacing of the diffraction spots, i.e. for a large unit cell, the spots are close together, whereas a small unit cell has diffraction spots far apart. As such, salt crystals, which have small unit cells, result in a small number of very strong spots and are easily recognised by their diffraction pattern. Once the unit cell has been determined from strong spots in a few images, computer programs can predict the location of all spots in all images and measure the intensity of the pixels within each spot. This process results in a file with the location and strength of each spot in the raw data, or 'unmerged' file.

The next step is to scale and merge the data. Scaling ensures that intensity variations, for example due to differences in the path length through the crystal that the beam experiences as the crystal rotates, are corrected for, so that all measurements on equivalent spots can be compared. The merging procedure then averages these equivalent spots. Equivalent spots arise because symmetry-equivalent reflections are recorded multiple times as the crystal is rotated, similar to the use of symmetry averaging in cryo-EM (see section 2.1.6.1). Scaling and merging is a critical step because the structure is primarily determined from the relative intensities of the spots. During these processes, various measures of data quality are generated. The R_{merge} is an indication of how well the measurements on different equivalent spots agree and should be below 10% (Tickle *et al.*, 1998; Evans, 2006).

2.2.6.2 The phase problem

Additive and destructive interference of the photons diffracted by the different molecules in the crystal lattice results in a diffraction pattern of spots, where diffraction

is additive, separated by 'blank' regions where interference was destructive. In mathematical terms, a diffraction dataset in 3D is the Fourier transform of the electron density of the molecule in the crystal multiplied by the Fourier transform of the crystal lattice. To calculate the electron density of the molecule, therefore, the Fourier transform of the diffraction dataset is calculated. However, during data collection the intensity of the diffracted X-rays is measured, but it is unknown what phase of the scattered wave gave rise to the measured intensity. Both intensity and phase are needed for a Fourier transform calculation and therefore the phases must be established.

Various methods exist for obtaining initial phases. By far the easiest method is molecular replacement (Rossman and Blow, 1962). Phases of a known structure, for a homologue or part of the complex for instance, can be combined with the measured intensities to obtain an initial map. This means that the model must be placed in the correct orientation in the experimental unit cell, by determining the orientation and position of the model where the correlation between the known structure and experimental data is maximised. Multiple copies of the same protein, or components of a complex, can be found this way. However, sometimes molecules with large conformational differences cannot be found and care must be taken with model bias using these methods. Molecular replacement is commonly used to solve structures of protein-protein, protein-nucleotide or protein-ligand complexes, structures of protein mutants, and structures of proteins where a homologous structure is already available.

Phases can also be obtained by experimental methods. Elements heavier than those typically found in macromolecules scatter X-rays inelastically as well as elastically at certain wavelengths, resulting in differences between otherwise equivalent reflections. These differences can be used in a variety of ways. Calculating these differences for one wavelength only is called single anomalous diffraction (SAD); if datasets are collected at multiple wavelengths this is known as multiple anomalous diffraction (MAD). Typically MAD and SAD are used for proteins that naturally contain

anomalously scattering atoms, such as metal ions, or if the protein has been expressed using selenomethionine instead of methionine. If the protein does not naturally contain anomalous scatterers, one or more types of heavy atoms can be co-crystallised with the protein or soaked into the crystals. In this case an alternative phasing approach, where the Bragg intensities in the presence of the heavy atoms are compared to those in a native dataset without the heavy atoms, can be used. This is known as single or multiple isomorphous replacement (SIR or MIR). Combinations of these methods, i.e. single or multiple isomorphous replacement with anomalous scattering (SIRAS or MIRAS) are also used, usually if one of the above methods was tried first, but did not yield an interpretable map. For small molecules, phases can be obtained using direct methods, which rely on phase relationships between atoms. This method can also be used for (small) proteins if the resolution is better than about 1.2 A.

2.2.6.3 Refinement

Once initial phases have been obtained, an electron density map can be calculated using the experimental amplitudes and initial phases. If the initial map is of sufficient quality, this electron density map can be used to build a model of the structure. The model is then used to generate new phases, which are used to calculate the next map. This iterative process is repeated until the fit of the model to the data no longer improves. The fit is quantified by a residual, commonly called the R-factor. It is essentially a calculation of which fraction of the data remains unexplained by the current model (Equation 2.2).

Equation 2.2 Calculation of the residual

$$R = \frac{\sum_h \left| F_{obs} - kF_{calc} \right|}{\sum_h F_{obs}}$$

Where F_{obs} are the measured structure factors, F_{calc} the calculated structure factors for the model. The values are summed over all reflections *h* and calculated structure factors are scaled by scale factor *k*.

Theoretically, the R-factor should decrease until it reaches the same value as the R_{merge}, the measure for the data quality. In small molecule crystallography this is the case, however, in macromolecular crystallography there usually is a gap between the R_{merge} and the R-factor, indicating there are properties of the protein in the crystal that are not accurately described by the model. After merging the data, an R_{free} set is defined. This is a subset of the data, usually 5%, that is used to prevent over-fitting the model to the data during refinement, i.e. making the model more complicated than the data demands. The 'free' data are not used for model fitting. They are reserved for calculating the free R residual, in the same way as for the regular R-factor. As all atoms in the structure (and therefore in the model) contribute to all diffraction spots, the R_{free} is a measure of how well the model predicts the data that was not used in the refinement.

Typically, the number of parameters that are refined in crystallography is four times the number of atoms in the asymmetric unit; the position of the atom in space (x, y, z) and the temperature factor B. Usually the occupancy, or the fraction of asymmetric units that a certain atom is present in, is assumed to be 1. However, for heavy atoms or ligands, the occupancy may also need to be refined or it will result in an unusually high B-factor. For proteins and other macromolecules at resolution lower than about 2.5 Å, the number of parameters far exceeds the number of measurements, so known chemical properties of protein structure are used to supplement the experimental data in what is called restrained refinement. Refinement using only the experimental data without stereochemical restraints is feasible only for ultra-high resolution datasets (better than 1.2 Å). This use of external restraints may result in overfitting of the model to the data; the R_{free} subset of data is used to detect this (Brunger, 1997).

Many algorithms can be used to optimise the model and reduce the differences between the model and the data. The most commonly used refinement programs, such as Refmac5 (Murshudov *et al.*, 1997), use the maximum likelihood principle. The algorithms maximise the joint probability function of the data, given the model and prior information such as stereochemical restraints.

Non-crystallographic symmetry (NCS), when multiple copies of a protein are present in the asymmetric unit, can be used to improve the observation/parameter ratio by restricting the differences between equivalent proteins. In the case of icosahedral viruses with one copy of the virus in the asymmetric unit, 60-fold NCS both reduces the amount of model building required and provides a powerful tool to improve map quality. However, because of the equivalent packing interactions of viruses in different orientations (section 1.2.1), any differences between copies of the coat protein are averaged out and present as disorder.

2.2.6.4 Validation

Structures of protein and nucleic acid are validated by comparison of their chemical properties, such as bond length and bond angles, to known values derived from small molecule crystallography and very high resolution protein structures. Interactions between subunits or molecules should be plausible, i.e. reasonable distances and interactions and no charge repulsion. Refinement programs and separate web services are available for analysis of structures (Chen *et al.*, 2010).

3 Materials and methods

3.1 Purification of turnip crinkle virus

TCV was purified from frozen leaves of *Nicotiana benthamiana* infected with 35S-TCV (derived from TCV-M cDNA) in an *Agrobacterium tumifaciens* vector, as described previously (Hogle *et al.*, 1986, Thomas *et al.*, 2003, Oh *et al.*, 1995). Leaves were kindly provided by Keith Saunders and Dr George Lomonossoff, John Innes Centre, Norwich, as previously described (Hogle *et al.*, 1986). The leaves were ground using a mortar and pestle under liquid nitrogen. The ground leaves were thawed and strained through muslin under vacuum. Celite (Supelco, 5g per 100 ml of sap) was added and the mixture was centrifuged at 12000 rpm for 30 minutes using a Sorvall SS34 rotor at 4 °C. Per 100 mL supernatant, 33 g of 30% PEG 6000 (Fluka) and 0.8 mL of 5 M NaCl (Fisher Scientific) were added. The mixture was centrifuged for 30 minutes at 12000 rpm as above. The pellets were extracted with phosphate-magnesium buffer (10 mM magnesium sulphate, 10 mM sodium phosphate at pH 7.4) overnight at 4 °C. The extracts were centrifuged in a Beckmann preparative ultracentrifuge with SW55Ti rotor for 2 hours at 35000 rpm at 4 °C. Pellets were resuspended in phosphate-magnesium buffer and 10 mL per 100 mL resuspension of 36 mg/mL bentonite, which is used to adsorb plant material while virus does not adsorb (Sigma, in phosphate-magnesium buffer). The suspension was centrifuged for 30 minutes at 12000 rpm to remove the bentonite. The supernatant was spun for 2 hours at 4 °C at 35000 rpm, pellets were resuspended in phosphate-magnesium buffer and spun for 10 minutes at 10000 rpm. These last two steps were repeated until the ultracentrifuge supernatant was clear. For some applications, the virus was further purified on a 10-40% sucrose gradient, using the Beckmann preparative ultracentrifuge with an SW55Ti rotor, and spun at 35000 rpm for 2 hours at 4 °C. The fractions containing TCV were pooled and dialysed using SnakeSkin Pleated Dialysis Tubing with a molecular weight cut-off of 3,500 Da (Pierce

Thermo Scientific, Wilmington (DE), USA) against phosphate-magnesium buffer for 24 hours, to remove the sucrose.

3.1.1 Measuring TCV concentration

The concentration of the virus was determined by UV/Vis spectrophotometry. Absorption was measured at a wavelength of 260 and 280 nm and virus concentration calculated using Equation 3.1 (Stoscheck, 1990).

Equation 3.1 Virus concentration

$$[\text{ protein (mg/mL) }] = (\ 1.55 \times A_{280} - 0.76 \times A_{260}\ ) \times \text{dilution factor}$$

3.2 Cryo-electron microscopy

3.2.1 Preparation of grids

Prior to applying the sample, Quantifoil grids were glow-discharged for 30 seconds to make the carbon film more hydrophilic. 3 µL virus solution with a concentration of 3 mg/mL was applied to Quantifoil R1.2/1.3 holey carbon films (Quantifoil Instruments GmbH, Jena, Germany). The grid was held in tweezers mounted in an automated pneumatic blotting and plunging device made by Prof. Howard White (Walker *et al.*, 1996). The grid was blotted from both sides for 1.6 seconds and directly plunged in liquid ethane. After freezing, the grid was transferred into a grid holder for storage in liquid nitrogen.

3.2.2 Cryo-transfer procedure

To transfer the frozen grid from the storage holder to the sample holder, a Gatan 626 Single Tilt Liquid Nitrogen Transfer Holder (Gatan Inc., Abingdon Oxon, UK) and transfer station were cooled with liquid nitrogen. The transfer station holds the sample holder in place alongside the storage holder. The grid is secured in the holder with a small clip ring that screws into the holder; to transfer the grid the clip ring is removed using the clip ring tool. The grid is transferred from storage holder to sample holder

using cooled tweezers and the clip ring is returned. The cryoshield is then closed and the holder cooled further before insertion into the microscope.

The microscope is equipped with a CompuStage^2, which has a 'closed' position at about 135˚ (where upright is defined as 0˚). In order to minimise the loss of nitrogen from the sample holder and therefore the risk of the sample warming up, the CompuStage was tilted to -60 degrees, moving the 'closed' position to about 75˚. The airlock was pre-evacuated before inserting the holder. The airlock was evacuated for a further 90 seconds to prevent air and nitrogen vapour being introduced into the column before rotating the holder to the 'open' position and final insertion of the sample into the column.

3.2.3 Data collection

All data for this thesis were collected using a FEI Tecnai F20 electron microscope operating at an acceleration voltage of 200 kV. This microscope is equipped with a 4,096 x 4,096 pixel CCD camera (Gatan). As frozen-hydrated biological samples are extremely sensitive to radiation damage by high-energy electrons, a low-dose procedure was employed to limit the number of electrons passing through the samples. The low-dose procedure operated at three different magnifications. The search mode used a 5,000x magnification and was employed to identify areas of high sample quality. Then focus mode was used to focus the beam on the carbon film at 150,000x magnification. In focus mode, the beam is directed a small distance away from the target area, so the target area is not exposed to the large electron dose required for focusing. Then exposure mode at 50,000x nominal magnification was used to record the image of the target area. The image was recorded either on the CCD camera or on photographic film; in either case an exposure time of 1 s was used.

3.2.4 Developing film

To prevent introduction of moisture into the column of the microscope, Kodak SO-163 electron film (Agar Scientific, Stansted, UK) was dessicated in an pumping unit

(Edwards, Crawley, UK) for at least two hours but preferably overnight. After dessication, transfer into the microscope was done as quickly as possible. After data collection, film was taken out of the camera and developed by immersing in Kodak D19 developer (Agar Scientific) at 20 °C for 12 minutes, followed by tap water (in lieu of stop solution) for 1 minute, and fixed in Ilford Hypam rapid fixer (Ilford Photographic, Cheshire, UK) for 5 minutes before rinsing in running tap water for 20 minutes. Developed film was stored in transparent sleeves for protection.

3.2.5 Film digitisation

To be able to process information recorded on photographic film, it has to be recorded as a digital image. Two types of scanner have been used in this thesis.

3.2.5.1 Scanning with the Zeiss SCAI microdensitometer

The Zeiss SCAI microdensitometer holds the film between two glass plates to keep the film within the focus of the camera. The step size of the scanner is 7 µm, i.e. 7 µm of film is recorded as one pixel in the digital image file. The final image sampling is calculated using Equation 3.2; with a real magnification of 52,911x this gives a sampling of 1.323 Å/pixel.

Equation 3.2 Image sampling

sampling (Å/pixel) = step size of scanner (Å) / real magnification

3.2.5.2 Scanning with the Imacon FlexTight 848

The Imacon FlexTight 848 (Imacon A/S, Denmark) bends the film to force the film to the focal point of the scanner. The step size of this scanner is 10 µm. Micrographs were digitised as 16-bit grey-scale TIFF images. For a real magnification of 52,911x the sampling calculated using equation 3.2 is 1.89 Å/pixel. Because of the larger pixel size, the Imacon scanner was only used on smaller datasets where a high resolution was not expected. A summary of the samples, magnifications, scanners and final sampling is given in Table 3.1.

Table 3.1 Summary of samples examined and magnification used

Sample	Medium	Magnification (real)	Scanner	Final sampling
TCV native	film	50,000 (52,911)	Zeiss	1.323
TCV expanded	film	50,000 (52,911)	Imacon	1.89
TCV recontracted	CCD	50,000 (69,000)	-	2.17
ERAV	CCD	62,000 (87,209)	-	1.72

3.2.6 Image pre-processing

3.2.6.1 Image conversion

The micrographs (both scanned and CCD) were converted from TIFF format to SPIDER format using the Imagic procedure em2em (Van Heel *et al.*, 1996). The contrast in the SPIDER images was then rescaled to more readily interpretable values by automatically narrowing the histogram to four times the standard deviation.

3.2.6.2 CTF correction and particle selection

The contrast transfer function (CTF) was calculated for each micrograph using the program CTFFIND3 (Mindell & Grigorieff, 2003). Particles were interactively selected using the program Boxer (Ludtke *et al.*, 1999). Selected particles were cut out and CTF correction was applied in SPIDER (Frank *et al.*, 1996). A band-pass filter was applied with a lower limit of 550 Å and upper limit of 6 Å. All images were combined in a single image stack. All subsequent image-processing steps were performed in SPIDER. The combined stack was normalised using the Spider command AR.

3.2.6.3 Image centring

Images were centred by iterative translational alignment using the Spider command AP MS to a rotationally averaged global sum of the image data, which was obtained by summing the imagestack with AS R, then rotationally averaging the summed stack with RO I. The routine was automated in the script *icos_centre.spi* (see Appendix CD) and executed twice before proceeding with image processing.

3.2.7 Image processing and 3D reconstruction

3.2.7.1 Obtaining a starting model

A starting model for TCV was obtained using the high symmetry of the virus. The centred particles were subjected to multivariate statistical analysis (MSA) and classified using Imagic. Classes appearing to exhibit five-fold symmetry were used for back-projection with icosahedral symmetry applied, using the SPIDER command BP 3F, and the best-looking model was chosen as the starting model.

3.2.7.2 Refinement of starting model

The starting model was reprojected using the SPIDER command PJ 3Q. All images were aligned to the references, using the SPIDER command AP SH. To prevent interpolation errors due to repeated movement of images, the aligned imagestack from the previous iteration was used for alignment. The shift and rotation were added to the total shifts and rotations from the previous iterations and applied to the raw images, using the SPIDER commands SA P and RT SQ, respectively. The shifted and rotated images were then averaged, using the SPIDER command AS R, according to which reference image they most resembled. To prevent bias of the model because of differences in numbers of images per class, only the *N* images with the highest correlation coefficient to the reference image are used for each average, where *N* is three times the number of images in the class with the lowest number of images. The model was visually inspected and reprojected to provide the reference images for the next iteration.

Icosahedral symmetry was used during the refinement for all samples. Because of the 60-fold symmetry, the reference model was only projected over the asymmetric triangle with theta ranging from 0 to 37˚ and phi from 0 to 71.999˚; this covers the entire virus provided the reference model has the 5-fold icosahedral symmetry axis centred along the z-axis. The number of reference images used was increased during the refinement from 53 to 331, corresponding to angular spacings of 3.5˚ and 1.5˚,

respectively (see Table 3.2). The final number of angles used in the refinement depended on the number of images in the dataset.

Table 3.2 Angular spacing ($\Delta\vartheta$) and number of angles

$\Delta\vartheta$	**Number of angles**	**Used for models**
3.5	53	all
3.0	78	all
2.5	110	ERAV,expanded TCV, native TCV
2.0	183	native TCV
1.5	331	native TCV

The refinement process was automated in the script *icos_oneiteration.spi* (see Appendix CD) and repeated until the refinement converged, i.e. until the model no longer significantly changed from one iteration to the next.

3.2.8 Resolution estimate by Fourier shell correlation

The resolution of the refined cryo-EM reconstruction was estimated by Fourier shell correlation. This method compares two models in Fourier space; the resolution where the correlation between the two models is 0.5 is quoted as the resolution of the model. To obtain the two models, instead of averaging the *N* images of each class with the highest correlation coefficient to the reference image, two averages are calculated per class, one with the odd and the other with the even-numbered averages. An 'odd' and an 'even' model are calculated using the 'odd' and 'even' averages for all angles. The models were masked using two different masks to prevent an artificially high resolution estimate because of mask correlation. The two models were normalized and the Fourier shell correlation (FSC) curve was calculated using the SPIDER command RF 3. The resolution estimate was then calculated by dividing the pixel size in Å by the inverse resolution where the FSC is 0.5.

3.2.9 Fitting the crystal structure into the maps

The C-subunit of the TCV crystal structure was kindly made available by Prof. S.C. Harrison (Harvard Medical School, Harvard, MA, USA). To generate a full icosahedral capsid, the C-subunit was duplicated and superposed on the A- and B- subunits of the TBSV crystal structure (2TBV, Hopper *et al.*, 1990). The N-terminal arm, which was described as disordered in both the TCV and TBSV crystal structures, was truncated for the A- and B-subunits at residue 81. Icosahedral symmetry operators were applied to the resulting asymmetric trimer to obtain the full capsid structure.

As the cryo-EM reconstructions were oriented around the 5-fold, the maps were rotated to follow the conventions of the PDB. This was done in Spider using the RT 3D command and angles 0, -32, 90 for phi, theta and psi. The individual domains of the regenerated TCV model were then fitted into each map by eye using Chimera and its 'sym' command and the coordinates saved relative to the map. In order to improve the fit, rigid body refinement in CNS was used. The rotated cryo-EM map was converted to CCP4 map format using the CP TO CCP4 command in Spider, with 32-bit output and 0,0,0 as starting point for the map. The axis order then needed to be changed to z, x, y using MAPMAN (Kleywegt and Jones, 1996), followed by SFALL (Winn *et al.*, 2011) to generate a reflection file from the map. SFTOOLS (Winn *et al.*, 2011) was then used to generate artificial standard deviations for the reflections by dividing the value of the reflections by 10. The resulting file was exported to CNS format using mtz2various. CNS was run using input files *make_cv.inp*, to generate a reflection file; *generate_easy.inp* to generate the molecular topology file and *rigid.inp* for rigid-body refinement of the individual domains (Brunger *et al.*, 1990). This process was not automated, but a summary can be found in the file EMfittingCNS.pdf (see Appendix CD)

The final fit was assessed by generating the full icosahedral capsid using the on-line utility on ViperDB (Carillo-Tripp *et al.*, 2009) and converting the pdb file to density using

the CP FROM PDB command in Spider, then using the CC C command to obtain cross-correlation coefficients.

3.3 Negative stain electron microscopy

3.3.1 Preparation of grids

Carbon-coated grids (Agar Scientific) were treated with UV light for 15 minutes to render the carbon more hydrophilic. 5 µL sample was applied to the grid and left for 60 seconds before blotting excess liquid off from the side. An equal volume of water was applied and left for 30 seconds to wash. Then 5 µL of 1% (w/v) uranyl acetate was applied and blotted off from the side after 10 seconds. The staining was repeated and the grid thoroughly air-dried before examination in the microscope.

3.3.2 Data collection

Negative stain microscopy was carried out using the Jeol 1200 microscope operated at 80 keV. Beam size and brightness were adjusted to cover the entire phosphorescent screen with an even illumination and a measurement of exposure time of between one and two seconds. The actual exposure time of the film was automatically adjusted to this measurement.

3.3.3 Developing and digitising film

After data collection, film was taken out of the camera and developed by immersing in Kodak D19 developer at 20 °C for 4 minutes, followed by tap water for 1 minute, and fixed in Ilford Hypam rapid fixer for 2 minutes before rinsing in running tap water for 20 minutes. Film was air-dried and digitised using the Imacon scanner as described in section 3.2.5.2.

3.4 Biochemical analysis of expanded TCV

3.4.1 Native, expanded, dissociated and recontracted TCV

TCV was biochemically analysed in four different states. For the native state, TCV stock solution was diluted to the desired concentration in the appropriate buffer. Native TCV buffer contained 10 mM magnesium sulphate and 10 mM potassium phosphate. Expansion buffer contained 5 mM EDTA and 25 mM Tris at pH 8.5. Dissociation buffer contained 5 mM EDTA, 100 mM Tris at pH 8.5 and 1 M NaCl. For recontracted TCV, the virus was diluted in concentrated expansion buffer and incubated at room temperature. After 5 to 15 minutes, $CaCl_2$ to a final concentration of 10 mM and HEPES at pH 7.5 (final concentration 100 mM) were added to recontract the virion.

3.4.2 Partial chymotrypsin digestion of the TCV coat protein

For the chymotrypsin digest experiment of the TCV coat protein, native, expanded, dissociated and recontracted TCV were prepared as described in section 3.4.1. 5 µg chymotrypsin (from bovine pancreas, Roche Diagnostics GmbH, Penzberg, Germany) was added to each 50 µL of reaction mixture. At each time point (0, 2, 5, 10, 30 and 60 minutes), 5 µL of reaction mixture was added to 20 µL SDS-PAGE loading buffer (10% w/v SDS, 200 mM dithiothreitol, 20% v/v glycerol, 200 mM Tris-HCl at pH 6.8, and 0.05% bromophenol blue) and denatured for 5 minutes in a boiling water bath. The samples were analysed by SDS-PAGE (resolving gel: 10% acrylamide, 375 mM Tris-HCl at pH 8.8, 0.1% w/v SDS, 0.1% w/v ammonium persulphate, 0.01% TEMED; 5% stacking gel: 5% acrylamide, 0.125 M Tris-HCl at pH 6.8, 0.1% w/v SDS, 0.1% w/v APS, 0.003% TEMED; Laemmli, 1970) using a Bio-Rad Mini-Protean electrophoresis apparatus, run at a constant current of 25 mA per gel (Bio-Rad Laboratories, Inc. Hercules (CA), USA). The running buffer contained 25 mM Tris-HCl, 200 mM glycine and 0.1% w/v SDS. Gels were stained in 45% v/v methanol, 10% v/v acetic acid and 0.5% w/v Coomassie brilliant blue for 30-60 minutes, and destained in 30% v/v

methanol and 10% v/v acetic acid with several changes of solution until sufficient contrast was obtained.

3.4.3 Determining the chymotrypsin cleavage sites

To determine the chymotrypsin cleavage sites of the TCV coat protein, cleavage products were analysed by mass spectrometry and N-terminal sequencing. For the N-terminal sequence analysis, native, expanded, dissociated and recontracted TCV were prepared as described in section 3.4.1, then incubated with chymotrypsin for ten minutes and analysed by SDS-PAGE as described in section 3.4.2. The protein was transferred to a PVDF membrane (EMD Millipore, Billerica (MA), USA) using an Invitrogen XCell II Blot Module Kit CE Mark (Invitrogen Life Technologies, Carlsbad (CA), USA) with a transfer buffer containing 10% methanol, 25 mM Tris pH 8.5 and 192 mM glycine. The membrane was stained with 0.1% Brilliant Blue-R, 1% acetic acid and 40% methanol for 60 seconds, followed by destaining with 50% methanol until sufficient contrast was obtained. The blot was then submitted for N-terminal sequencing by Edman degradation using an Applied Biosystems Procise 494 HT N-terminal sequencer.

Samples for mass spectrometry were taken from the same reaction mixtures used for the N-terminal sequencing. All subsequent steps were performed using mass spectrometry grade acetonitrile, water and formic acid. The samples were concentrated, purified and denatured using C4 ZipTip®s (Millipore). For each sample, the ZipTip® was washed three times with 10 µL 95% v/v acetonitrile and 3% v/v formic acid, followed by four washes with 10 µL 3% v/v formic acid in water. The sample was loaded by pipetting up and down in the reaction mixture, after which the tip was washed six times with 10 µL 5% v/v acetonitrile and 3% v/v formic acid. Finally, the sample was eluted with 5 µL 60% v/v acetonitrile and 3% v/v formic acid and submitted for analysis by mass spectrometry.

3.4.4 Coat protein sequencing

To obtain the sequence of the TCV coat protein, RNA was extracted from the virus. 50 µL of TCV at 14 mg/mL was added to a microcentrifuge tube containing 10 µL 1 M Tris pH 8.5, 10 µL 50 mM EDTA, 25 µL 4 M NaCl, 5 µL water and 100 µL phenol:chloroform:isoamylalcohol at pH 6.6 (25:24:1; Applied Biosystems Life Technologies, Carlsbad (CA), USA) and vortexed. The tube was spun for 2 minutes at 17000 *g* and the aqueous phase was added to a second aliquot of phenol-chloroform, mixed and centrifuged a second time. The phenol was extracted with two aliquots of chloroform:isoamylalcohol (24:1; Applied Biosystems). The aqueous phase was added to 200 µL ethanol with 10 µL 4 M NaCl and the mixture left at -20 ^{o}C overnight to precipitate the RNA. The RNA was centrifuged at 4 ^{o}C for 20 minutes at 17000 *g*. The pellet was washed with 75 % v/v ethanol and spun for a further 15 minutes. The pellet was dissolved in 30 µL nuclease-free water and concentration estimated by UV-spectrophotometry at 260 nm using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington (DE), USA). The RNA concentration was calculated with Equation 3.3, assuming an extinction coefficient for RNA of 25 µL/µg/cm.

Equation 3.3 RNA concentration

$$\mu g\ RNA/mL = 40 \times A_{260}$$

To obtain a PCR template, the coat protein gene was reverse transcribed from the TCV genomic RNA using a Transcriptor High Fidelity cDNA Synthesis Sample Kit (Roche) according to the manufacturer's instructions using primers RI and RII. Two PCR reactions per reverse transcription reaction were set up using a KAPA2G Robust PCR Kit (Kapa Biosystems, Boston, USA) according to the manufacturer's instructions. 5 or 15 µL of the reverse transcription reaction mixture was used as the template without further purification, and primers FI, FII, RI and RII were added in the four possible combinations of forward and reverse. The PCR products were purified on a 1.5 % agarose gel using 1x TBE running buffer (89 mM Tris-borate, 2 mM EDTA) and extracted using a QIAgen Gel Extraction Kit (QIAGEN, Washington, USA) according to

the instructions. The products were quantified using the NanoDrop spectrophotometer as above, but using Equation 3.4 to calculate the concentration, assuming an extinction coefficient for DNA of 20 µL/µg/cm.

Equation 3.4 DNA concentration

µg DNA/mL = 50 x A_{260}

The PCR product was submitted to GATC Biotech for automated sequencing using primers FI, RII, MF and MR. Two further primers, SF and SR, were selected to cover the entire gene. All primers are summarised in Table 3.3.

Table 3.3 Primers used for reverse transcription, PCR and sequencing

Name	Sequence
FI	CCT GAA ATC AAA CCG ATT CAC ACA TCC
FII	GGT CCT CCT ACT TTG TCA TCT GAT TCC
RI	GGT CAA AAT AAA GCG ACT GGG GG
RII	CCC TAA CAC AGG TCA AAA TAA AGC G
MF	CGA GAT GCA GCC AAA CCT CC
MR	GCT GAT ACC ATC CGC CAC AAA GC
SF	CCC AAT GGG CGA TAA AGT G
SR	CTG AGT GAC GTG AAT CG

3.4.5 Ribonuclease digestion of the TCV genome

Native and expanded TCV were prepared as in section 3.4.1. Additionally, TCV genomic RNA, isolated as described in section 3.4.4, was dissolved in expansion buffer to a final concentration of 0.5 µg/µL. Ribonuclease A (Applied Biosystems) was added to a final concentration of 4 pg/µL. At each time point, i.e. 0, 10, 30, 60, 120 and 180 minutes, 4 µL of the reaction mixture and 75 µL water were added to 100 µL of phenol:chloroform:isoamylalcohol and phenol-chloroform extracted as described in section 3.4.4. The aqueous phase was collected and added to two volumes of ethanol and one-tenth volume of 4 M sodium chloride and left at -20˚C overnight to precipitate the nucleic acid. It was then spun for 20 minutes at 4˚C at 17000 *g* and the supernatant

was discarded. The pellet was washed with 75 % ethanol. The RNA was dissolved in nuclease-free water. Aliquots were boiled for 5 minutes in an equal volume of Formaldehyde loading buffer (Applied Biosystems) containing 0.1 µL 10 mg/mL ethidium bromide (Sigma) per sample. The samples were analysed on a 1.5 % denaturing agarose gel (1.5 % agarose, 40 mM MOPS pH 7, 10 mM sodium acetate, 1 mM EDTA, 2.214 M formaldehyde).

3.4.6 Virus-ribosome complexes

To obtain striposomes, TCV was expanded as in section 3.4.1 and added to a wheat germ extract prepared according to the manufacturer's instructions (Promega, Madison (WI) USA). Mg^{2+} (as $MgCl_2$) was added to a final concentration of 1 mM to the reaction after mixing in the EDTA-containing expanded TCV. After incubating at 25˚C for 30 minutes, cycloheximide (Sigma) was added to a final concentration of 1 mg/mL to stop translation. The sample was then centrifuged at 4 ^{o}C for 10 minutes and the supernatant concentrated using a microfuge concentrator device with a 300 kDa molecular weight cut-off (Pall Corporation, Port Washington (NY), USA) before application to a UV-treated EM grid and staining with 1% uranyl acetate.

3.4.7 Analytical ultracentrifugation

Analytical ultracentrifugation was carried out by Amy M. Barker. Samples (0.32 mL) were placed in 1.2 cm pathlength 2-sector meniscus-matching epon centrepiece cells built with sapphire windows. These were then centrifuged at 10,000 rpm and 20.0°C in an An50Ti analytical rotor in an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Inc., Palo Alto, California 94304). Changes in solute concentration were detected by interference and absorbance scans at 270 nm. Buffer densities and viscosities were calculated using the program Sednterp version 1.09 (2006, Dr Thomas Laue, Department of Biochemistry, University of New Hampshire, Durham, NH 03824). Radial absorbance plots were used for fitting to sedimentation profiles with the program

Sedfit, version 12.1b (2010), using a continuous distribution c(s) Lamm equation model (Schuck, 2000).

3.4.8 Gel filtration

To see if any protein is released from the capsid after proteolysis, expanded TCV was prepared and treated with chymotrypsin for 30 minutes as in section 3.4.2. Aprotinin (Roche), a specific chymotrypsin inhibitor, was added to the sample and the buffers in all subsequent steps, at a final concentration of 2 μg/L. Proteolysis reactions for gel filtration analysis were stopped by adding Roche Complete protease inhibitor (100 µL of a solution of 1 tablet/mL), 160 µg Pepstatin A in DMSO and 2 mg Aprotinin per mL reaction volume. Particles were separated on a 10/300 GL Superdex S200 column using an ÄKTA Explorer at a flow rate of 0.5 mL/min. 1 mL fractions were collected and concentrated using an Amicon Ultra (3 kDa molecular weight cut-off, Millipore) spin-concentrator device. The concentrated fractions were analysed by SDS-PAGE as in section 3.4.2.

3.5 X-ray crystallography of STNV-B3

3.5.1 Crystallisation conditions

STNV virus-like particles (VLP) reassembled with the B3 RNA aptamer were kindly provided by Robert Ford (University of Leeds). The VLPs were crystallised using the conditions previously established for recombinant STNV (Lane *et al.*, 2011). 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 solution at 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.

3.5.2 Crystal mounting

Crystals were cryo-protected with glycerol in increasing concentrations using a three-step protocol (Table 3.4). Immediately after the last step, the crystal was plunge-frozen in liquid nitrogen and subsequently stored in liquid nitrogen.

Table 3.4 Cryo-protection of STNV-B3 crystals

Solution	**Time**
20% glycerol, 80% reservoir solution	60 s
25% glycerol, 75% reservoir solution	20 s
30% glycerol, 70% reservoir solution	1 s

3.5.3 Data collection

Data were collected at beamline I02 at Diamond Light Source using a wavelength of 0.9795 Å. The best crystal diffracted to 2.29 Å; to ensure less ordered structures could also be examined by including low resolution data, three separate datasets were recorded with different detector distance and optimised beamstop position. These datasets are summarised in Table 3.5.

Table 3.5 Datasets used for STNV-B3

Dataset	**Detector distance**	**Resolution range**
High resolution	236.62 mm	30 - 2.29 Å
Low resolution	617.87 mm	116 - 3.8 Å
Very low resolution	993.81 mm	217 - 7.0 Å

3.5.4 Data processing and integration

The data were processed by the automated data reduction procedure Xia2 (Winter, 2010), using the 3dii option, which forces the program to use all images for automatic indexing with the program XDS (Kabsch, 1988). Although this is slower, it can prove useful in more challenging cases (Winter, 2010). Xia2 is especially designed to process multiple datasets collected from the same crystal. It uses the indexing program XDS for indexing iteratively with POINTLESS to check for the spacegroup and cell parameters

that fit the data best and enforce consistent indexing between the datasets, and integrates the data. It scales the datasets together using XSCALE and writes out a single scaled and merged reflection file.

3.5.5 Initial phasing

As the crystals have the same space group and cell parameters as recombinant STNV, the structure of recombinant STNV was used for calculating initial phases by rigid body refinement of the entire virus in REFMAC (Murshudov *et al.*, 1997). To prevent artificially low R-factors, the R_{free} flags were copied from the recombinant STNV dataset. Waters from the recombinant structure were used in initial phasing but removed after rigid body refinement.

3.5.6 Protein structure refinement

The STNV VLPs crystallised with one VLP in the crystallographic asymmetric unit. This results in 60-fold non-crystallographic symmetry (NCS). Only one, arbitrarily chosen, coat protein subunit (chain A) was subjected to rebuilding during in Coot (Emsley and Cowtan, 2004). The other chains were regenerated after rebuilding by the script makewholecapsid.sh (see Appendix CD), which calls the CCP4 program LSQKAB to first calculate the rotational and translational operators for each chain from the previous round of refinement, and then to copy chain A for each operator. The resulting file is used for maximum likelihood restrained refinement in REFMAC using tight restraints for the full, 60-fold NCS. The VLPs contain 92 calcium ions: one for each CP subunit, 12 on the 5-fold icosahedral symmetry axes, and 20 on the three-fold icosahedral symmetry axes. For refinement, the occupancies of the calcium ions on the 5-fold and 3-fold axes, and any associated water molecules, were manually adjusted to 0.2 and 0.33, respectively, to allow the symmetry-related atoms to occupy the same space.

Map averaging was performed in Coot using the NCS maps function. As additional density was observed around the icosahedral two-fold symmetry axes extending

beyond the extent of the averaged map, coat protein dimers were identified and a 30-fold NCS averaging was performed. This resulted in an averaged map that better allowed inspection of the additional density.

4 Three-dimensional cryo-EM reconstructions of TCV

4.1 Introducing TCV

Turnip crinkle virus (TCV) is a member of the carmovirus genus of the family of Tombusviridae, which was named after tomato bushy stunt virus (TBSV). It was first described as the causative agent of leaf crinkling, mottle and slight stunting in *Brassica rapa* (turnip), and mild mottle in other species of *Brassica* (Broadbent, 1957). It has a wide experimental host range, but host specificity in the wild is determined by the insect vector. TCV is transmitted by small jumping flea-beetles: 9 species of the genus *Phyllotreta* and 2 of the genus *Psylliodes*. When the beetles feed on plants, they cause damage to the plant cell wall, which allows the virus to enter. The virus then uncoats its genome, which acts directly as messenger RNA, and viral proteins are produced. The genome is reproduced by the polymerase and assembles with coat protein to form new virions (Figure 4.1). The details of genome uncoating in TCV are unclear.

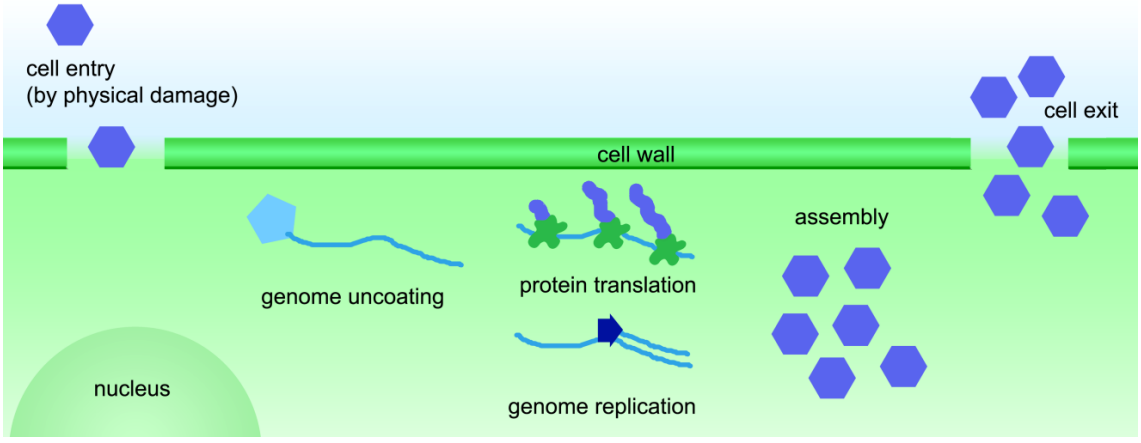


Figure 4.1 TCV life cycle
The life cycle of TCV. The virus enters the cell by physical damage. Once the virus is inside the cell, the genome is uncoated. The genome is then used as the template for both protein production and the replication of the genome by the viral replicase. Progeny virus particles assemble from the new genomes and coat protein, and the virus exits the cell again by physical damage.

The crystal structure of TCV was one of the first virus crystal structures to be solved in the 1980s. The TCV virion has a *T*=3 quasi-equivalent capsid, and both the capsid

and the coat protein are very similar in structure to those of TBSV (Harrison *et al.*, 1978; Hogle *et al.*, 1986). The coat protein consists of 351 amino acids, with two structured domains observed in the X-ray structure: the S-domain forms the continuous coat protein shell, while the P-domain forms the 'spikes' projecting from the exterior of the capsid (Figure 4.2c). The S-domain exhibits the jelly-roll fold common to many virus structures. The N-terminal part of the protein is disordered and was not observed in the crystal structure. For the A and B subunits the N-terminal 80 residues are disordered, but in the C-subunit residues 53 to 80 interact around the three-fold icosahedral symmetry axis to form an intertwined structure called the β-annulus (Figure 4.2b), which stabilises the three-fold contact (Hogle *et al.*, 1986).

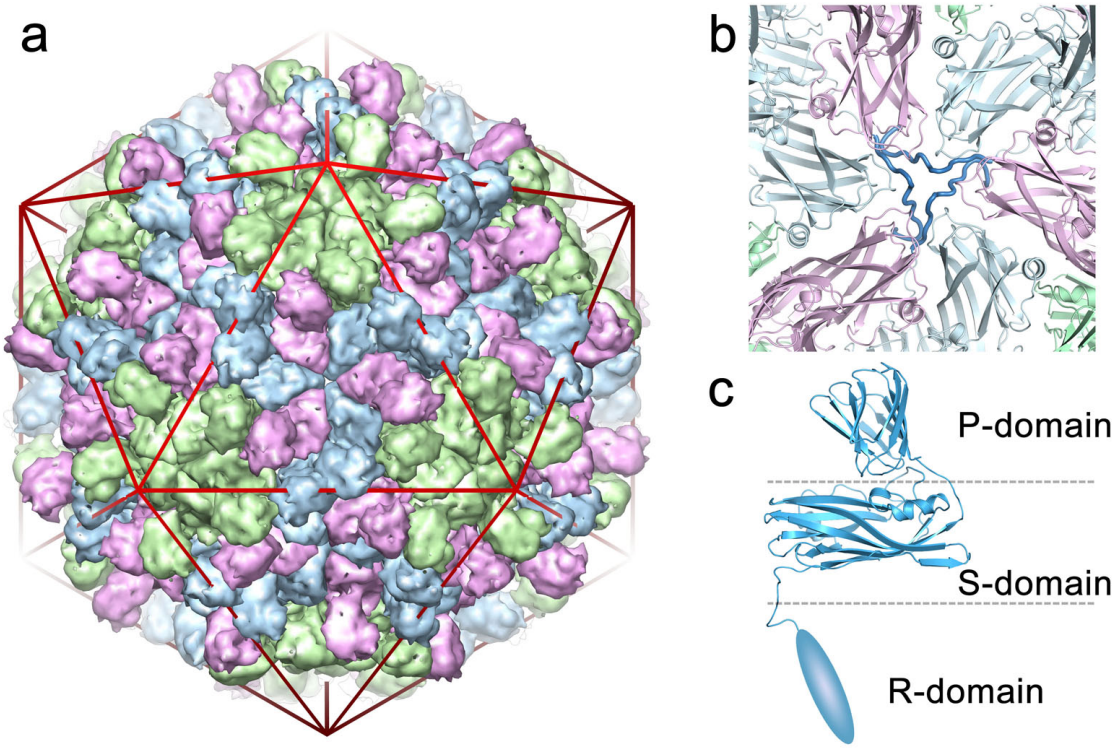


Figure 4.2 Structure of TCV
a) Surface rendering of the capsid showing A-subunits in green, B-subunits in magenta and C-subunits in blue. An icosahedron is shown to indicate the symmetry of the particle. b) The β-annulus shown in ribbon representation. Subunits are coloured as in panel a. The interacting parts of the protein are highlighted in darker blue. c) Domain structure of the TCV coat protein. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

The TCV genome consists of 4051 bases of single-stranded RNA. It encodes 4 proteins: the 88 kDa RNA-dependent RNA polymerase, which contains an amber stop

codon after 28 kDa, the 38 kDa coat protein and an 8 kDa protein that is presumed to be a movement protein and overlaps with the polymerase gene (Carrington *et al.*, 1989). The 3' and 5' untranslated region (UTR) are translational enhancer sequences (Qu and Morris, 2000; Stupina *et al.*, 2008).

TCV has been extensively studied biochemically as well as structurally. It has long been known that TCV can be expanded and dissociated *in vitro.* Expansion takes place when TCV is put in an EDTA-containing buffer at slightly elevated pH (~8.5) and low ionic strength. If the expansion buffer is used in combination with high ionic strength, the virus dissociates into coat protein dimers and a salt-stable RNA-protein complex (rp-complex), which can be separated by gel filtration (Golden and Harrison, 1982). *In vitro* reassembly can take place from the salt-stable rp-complex, with TCV RNA or heterologous RNA, but RNA is required for assembly. If the virus is reassembled with naked genomic RNA, the rp-complex is regenerated (Sorger *et al.*, 1986). Both expanded virus and dissociated coat protein are susceptible to proteolysis (Golden and Harrison, 1982), and proteolysed coat protein cannot form *T*=3 capsids, but forms smaller *T*=1 capsids instead, presumably because it cannot form the β-annulus at the icosahedral three-fold (Sorger *et al.*, 1986).

When TCV is analysed by SDS-PAGE, an 80 kDa protein (p80) is seen as well as the coat protein. There is no separate gene or read-through product encoding the p80 protein and amino acid composition analysis has shown it to contain the same amino acids as the coat protein, suggesting it is a covalent dimer of coat protein monomers. The p80 dimer is present in the rp-complex after disassembly of the virion, and is stable in reducing conditions, acidic, basic and denaturing conditions, but the chemical nature of the cross-link is unknown. It has been shown to be more susceptible to surface labelling than the monomeric coat protein, indicating increased breathing dynamics even in the native capsid (Sorger *et al.*, 1986; Stockley *et al.*, 1986). Similar minor protein components with an apparent mass of about twice that of the coat protein have been observed for other plant viruses, such as TBSV, satellite tobacco necrosis

virus, southern bean mosaic virus and cowpea chlorotic mottle virus (Rice, 1974), as well as different minor protein components consisting of coat protein read-through products (Van den Worm *et al.*, 2006).

As well as being the building block of the capsid, the coat protein of carmoviruses such as TCV also plays a role in the suppression of RNA silencing in the host plant (Cao *et al.*, 2010). RNA silencing is a plant defence mechanism that targets double-stranded RNA. Dicer-like (DCL) proteins cleave double-stranded RNA to produce small-interfering (si)RNA, which is bound by Argonaute (AGO) proteins to target complementary RNA for cleavage or translational suppression (Harvey *et al.*, 2011). The TCV coat protein appears to interact with AGO1 in a way that mimics the host regulatory pathway (Azevedo *et al.*, 2010). In contrast to the carmoviruses, TBSV and the other tombusviruses encode a separate silencing suppressor protein, p19, which binds to siRNA and prevents its unwinding to form the short ssRNA fragments needed for RNAi (Scholthof, 2006; Vargason *et al.*, 2003). These silencing-suppressor proteins are required to enable systemic infection in the plant. The coat protein may also play a role in regulating expression of the polymerase gene (Wei *et al.*, 1990).

This work aims to investigate the structures of native and expanded TCV as well as the uncoating mechanisms of the virus.

4.2 Three-dimensional reconstruction of native TCV

4.2.1 Capsid structure

A three-dimensional reconstruction of native TCV was made using 18681 single-particle views of the virus for the refinement. The final model used the 14841 of these images, spread over 331 angles, that correlated best to the reference images obtained from the previous iteration. A side-by-side comparison of all the averages and corresponding reprojections is given in Appendix A.

The final reconstruction of native TCV, with a resolution of ~11.5 Å, shows the typical spiked appearance of the *Tombusviridae* (Figure 4.3a). The protein shell is

continuous, with no visible holes or pores. The density for the spikes connects, with A/B-dimers at each end connected by a C/C-dimer. This connection is not seen when the crystal structure of TBSV is converted to density.

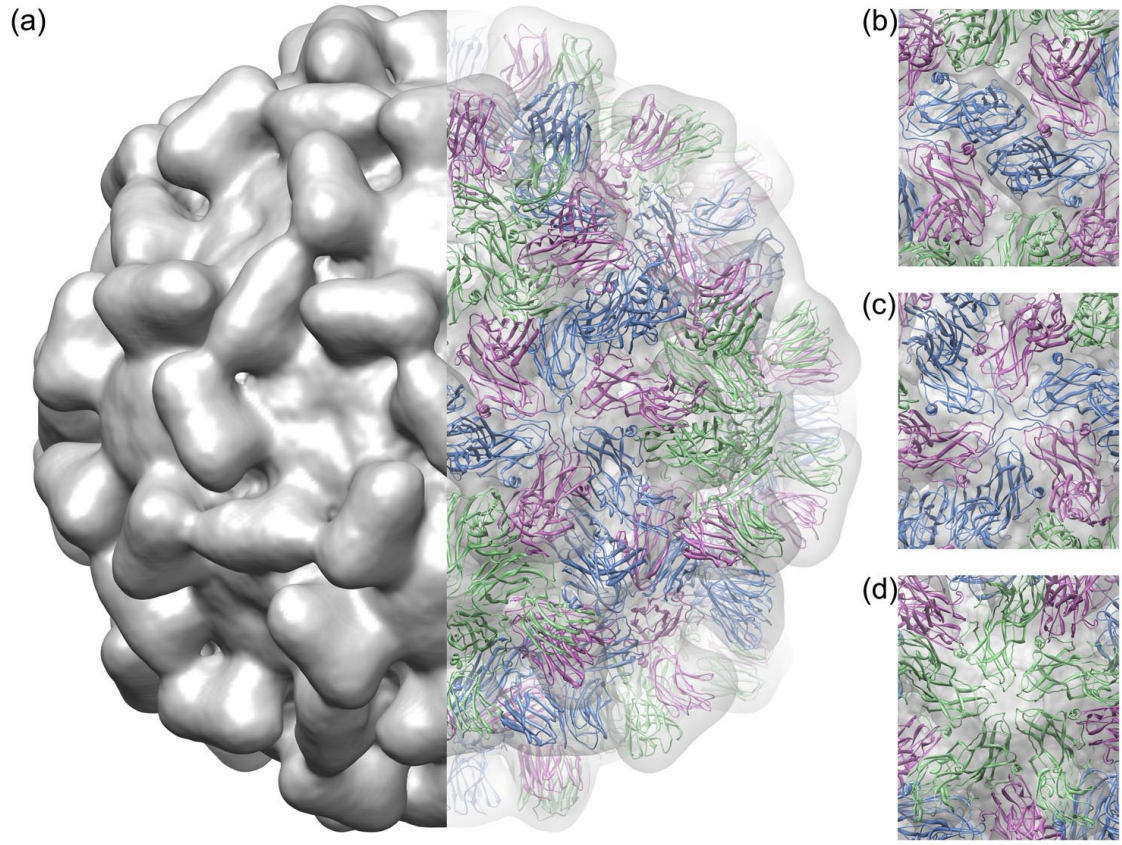


Figure 4.3 3D cryo-EM reconstruction of TCV
(a) Surface view and atomic structure fitting of the native TCV virion with on the left, a solid grey surface representation of the cryo-EM density. On the right, the surface has been made transparent in order to show the fitted coordinates of the TCV crystal structure shown in ribbon representation (coloured as in Figure 4.2a). The crystal structure is coloured as follows: green for the A-subunit, magenta for the B-subunit and blue for the C-subunit. (b)-(d) show close-up views of the fitted coordinates, perpendicular to the (b) 2-fold, (c) 3-fold and (d) 5-fold symmetry axes of the particle. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

The atomic model of TCV, obtained as described in section 3.2.9, fits very well in the cryo-EM density. The S-domains that form the continuous protein shell of the virus did not require any additional fitting. The fit of the P-domains, however, was improved by a small rigid-body movement (Figure 4.4). A difference in the hinge angle between the S- and P-domains of TCV compared to TBSV was described when the TCV structure was first solved (Hogle et al, 1986), indicating the change in angle represents a real difference between the two viruses.

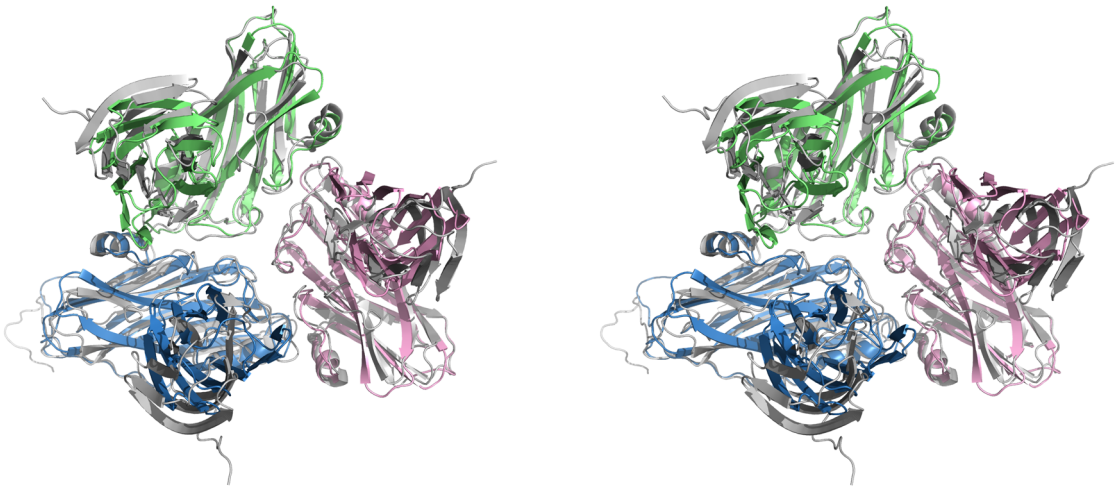


Figure 4.4 Difference between the TCV and TBSV structures
A stereoview of the asymmetric unit of TCV (coloured as in Figure 4.3) and TBSV (PDB entry 2TBV, shown in grey). The S-domains are in a very similar position, but the P-domains have moved slightly in TCV relative to TBSV. Only one asymmetric unit is shown for clarity. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

4.2.2 Pseudo-icosahedrally ordered internal structure

In addition to the density corresponding to the protein capsid, two further shells of density are seen in the reconstruction of native TCV (Figure 4.5a). The density of the innermost shell (pink in Figure 4.5a), which lies between radii of 50 and 72 Å, is weak and does not show any distinctive features. The second shell, which covers radii between 82 and 110 Å (coloured yellow in Figure 4.5a) just under the protein capsid, is more featured. At higher contour levels, i.e. when only the strongest density is shown, the second shell shows a cage-like structure which consists of pentagonal rings of density around the five-fold symmetry axes, surrounded by flattened hexagons centred on the two-fold symmetry axes and spanning the space between the three-fold symmetry axes. The additional density within the virion corresponds to the viral RNA genome and the coat protein N-terminal regions, the parts of the virus that were not seen in the crystal structure. We cannot, at this resolution, distinguish which parts of the density are attributable to RNA or protein.

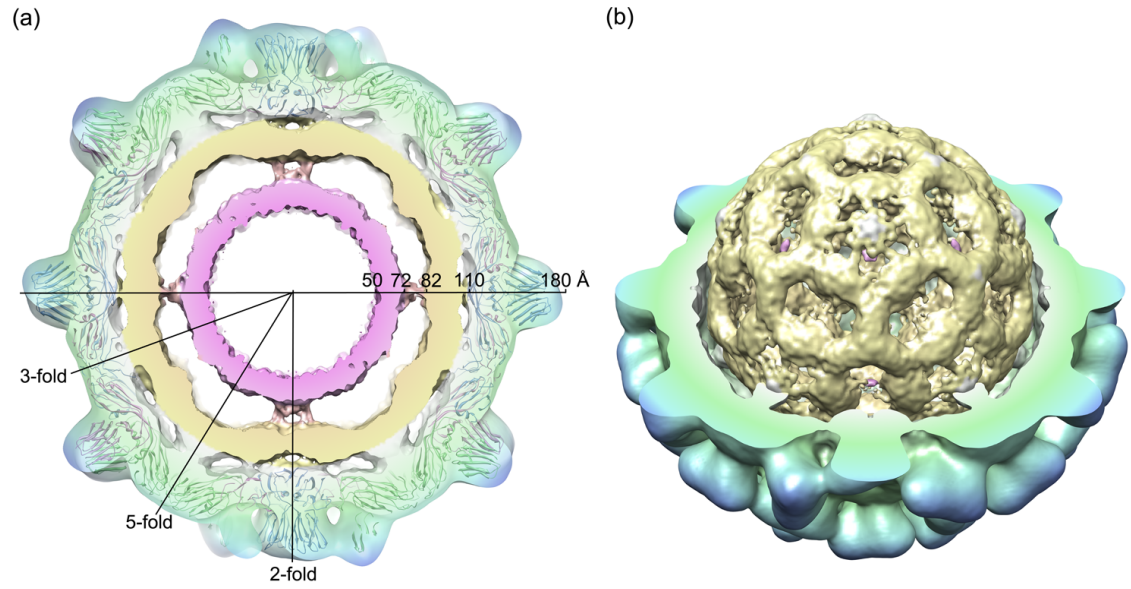


Figure 4.5 Internal structure of native TCV
(a) A central section through the cryo-EM structure of the native TCV virion. The EM density is transparent in the protein capsid region, revealing the fitted atomic coordinates coloured as in figure 4.1. The density is coloured using a radial colour scheme, ranging from pink at a radius of 50 Å to blue at 180 Å (see horizontal scale bar). The icosahedral symmetry axes are also indicated. (b) Cutaway view of the virus. The protein capsid density has been partially removed to show the cage-like structure of the internal density at higher contour levels. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

A similar multi-shell structure has been seen in cryo-EM reconstructions of other plant viruses and bacteriophages (Aramayo *et al.*, 2005; Katpally *et al.*, 2007; Toropova *et al.*, 2008). The plant viruses TBSV and CNV are both very similar to TCV and cryo-EM reconstructions have shown similar arrangements of RNA, although the details of the coat protein-RNA interactions at this resolution are different. In the TCV reconstruction shown here, the main contacts between the protein density and the middle shell of density occur at the three-fold and five-fold icosahedral symmetry axes, with the main contact between the innermost and middle shells at the two-folds. For TBSV, the main contacts between the middle shell and the capsid are on the two-fold symmetry axes, while the interior shells connect at the five-fold axes (Aramayo *et al.*, 2005). In CNV, on the other hand, the main contacts between the coat protein and the middle shell are at the five-fold axes, while the two inner shells connect at the two-folds axes (Katpally *et al.*, 2007). These differences suggest there may not be a conserved interior structure for members of the *Tombusviridae* family. A similar two-shell structure

was observed in a 3D cryo-EM reconstruction of native bacteriophage MS2 (Toropova *et al.*, 2008).

Neutron small-angle scattering studies of TBSV indicates a structure with four shells. The outermost shell is formed by the coat protein, followed by a shell consisting of predominantly RNA, a shell of predominantly protein and an innermost shell with predominantly RNA (Aramayo *et al.*, 2005). This was also assumed to be the case in CNV where a fourth density shell can be seen (Katpally *et al.*, 2007). On the other hand, the TCV reconstruction presented here shows no signs of an fourth shell. The genomes of TBSV and CNV with 4776 and 4701 nucleotides, however, are longer than that of TCV with 4051 bases (Hearne *et al.*, 1990; Rochon and Tremaine, 1989; Carrington *et al.*, 1989). Previous work on the MS2 genome has shown that when a shorter RNA is packaged, it preferentially occupies the outer shells of density, presumably because of repulsion of the negatively charged RNA, while no ordered density is seen for the inner shells when shorter RNA is present (Rolfsson *et al.*, 2010). It should also be emphasised that all these reconstructions have been obtained with icosahedral symmetry applied, and all internal structure is therefore averaged. It is not possible to say from the averaged structure what the fold of the RNA is.

4.2.3 Resolution of the native TCV reconstruction

The resolution of the cryo-EM reconstruction was determined as only about 11.5 Å, despite the large number of particle images used. The particles exhibit a good spread of defocus and micrographs exhibiting drift or astigmatism, judged by the power spectrum of the micrograph, were removed from the dataset. The particles are also evenly spread across the angular spacing used. No averaged image was allowed to contain more than three times the number of images in the smallest class, which is usually the 5-fold view.

To assess whether components of the model have different degrees of order, various masks were used during the calculation of the FSC. This shows that, although

the inner component of the model has a slightly lower resolution, the P-domain spikes are most flexible and this results in the lower than expected resolution (Figure 4.6).

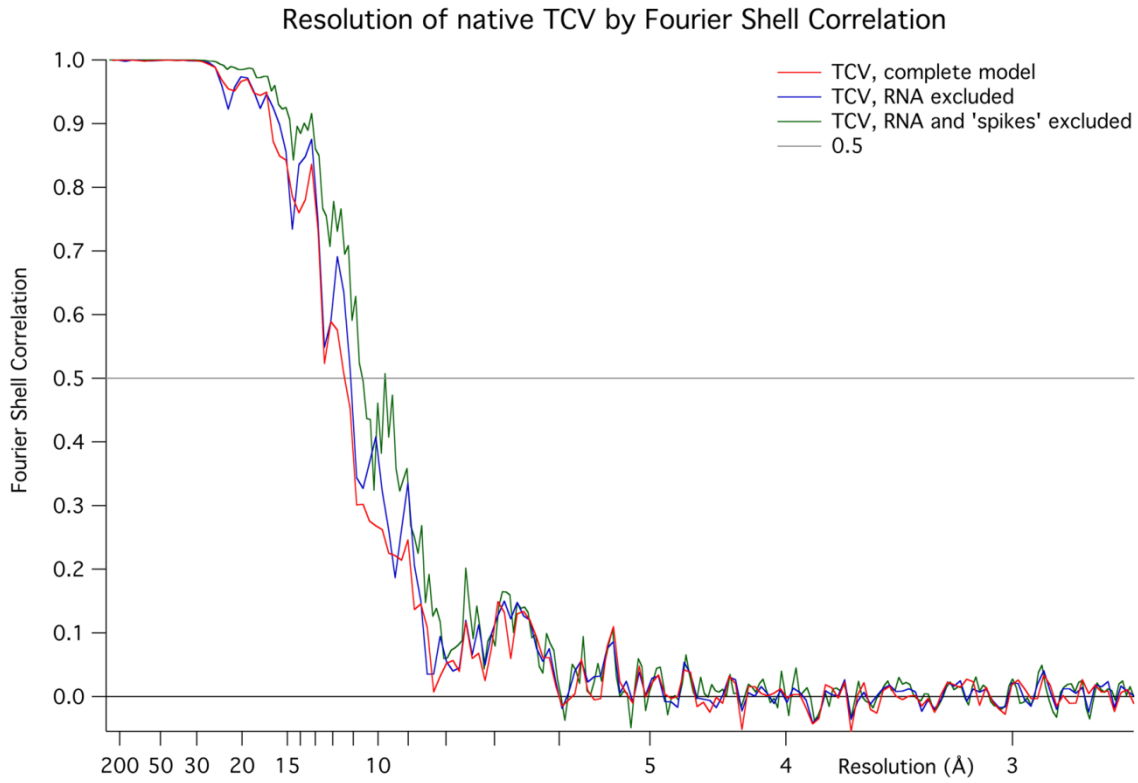


Figure 4.6 Fourier shell correlation curves for native TCV
Showing the Fourier shell correlation (FSC) curves for the full reconstruction in red. This indicates the model has a resolution of approximately 11.5 Å. If the RNA is not taken into account, the resolution improves marginally to about 11 Å (blue curve). When, however, both RNA and coat protein P-domains or 'spikes' are ignored and only the coat protein shell is included in the calculation, the resolution of the model improves to just below 10 Å. This indicates the P-domains may be flexible and limiting the resolution of the model.

Analysis of individual classes by multivariate statistical analysis (MSA) did not yield any obvious discrete states or conformations of the spikes - this was to be expected as averaging will eliminate any distinct conformations and it is unlikely multiple virus particles happen to have the same arrangement of spike conformations as well as having been imaged in the same orientation on the grid. The linked, Z-like formation of the spikes around the icosahedral 2-fold axis also indicates a degree of disorder; cryo-EM reconstructions of TBSV show individual spikes for each CP dimer (Aramayo et al., 2005)

As the spikes are key features in the alignment of TCV particles, the flexibility may have negatively influenced the resolution by introducing alignment errors as well as the

lower resolution of the spike region resulting in a lower average resolution for the model.

4.3 Three-dimensional reconstruction of expanded TCV

4.3.1 Capsid structure

TCV was expanded by dilution into a buffer containing 5 mM EDTA at pH 8.5. The three-dimensional reconstruction of expanded TCV was made using 5121 single-particle views of the virus. For the final model, 2610 particle images in 110 classes were used, selected as described in section 3.2.7.2. The resolution was determined to be 17 Å. A side-by-side comparison of all the averages and corresponding reprojections is given in Appendix B.

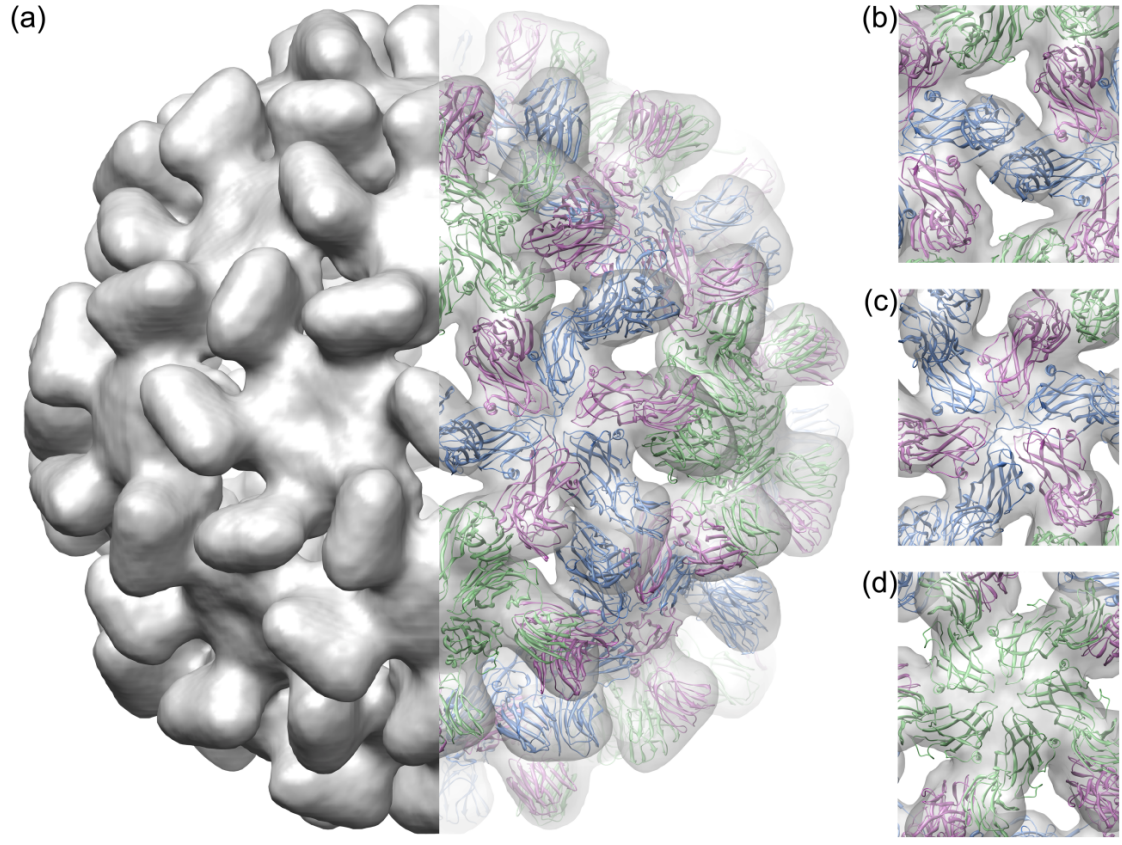


Figure 4.7 Cryo-EM reconstruction of the expanded TCV virion
(a) Surface view and atomic structure fitting of the expanded TCV virion. On the left, the cryo-EM density is shown as a solid grey surface. On the right, the density is transparent to show the fitted coordinates for the TCV crystal structure in ribbon representation. The ribbons are coloured as in Figure 4.1. (b)-(d) Close-up views of the atomic structure fit, perpendicular to the (b) 2-fold, (c) 3-fold and (d) 5-fold symmetry axes of the particle. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

The capsid of expanded TCV is subtly different from the native virion (Figure 4.7). Although it still clearly has a spiked appearance, the density for the P-domains is well-separated. The protein shell is not continuous, as in the native particle, but shows pores either side of the two-fold symmetry axes. The expanded particle is about 5% larger than the native virion, but the different features of the capsid indicate it is genuinely a different species rather than a magnification error.

The atomic coordinates of TCV were fitted into the cryo-EM density using rigid-body fitting, as for the reconstruction of native virus. The pores in the capsid density fall between the three subunits that make up the icosahedral asymmetric unit.

4.3.2 Internal structure

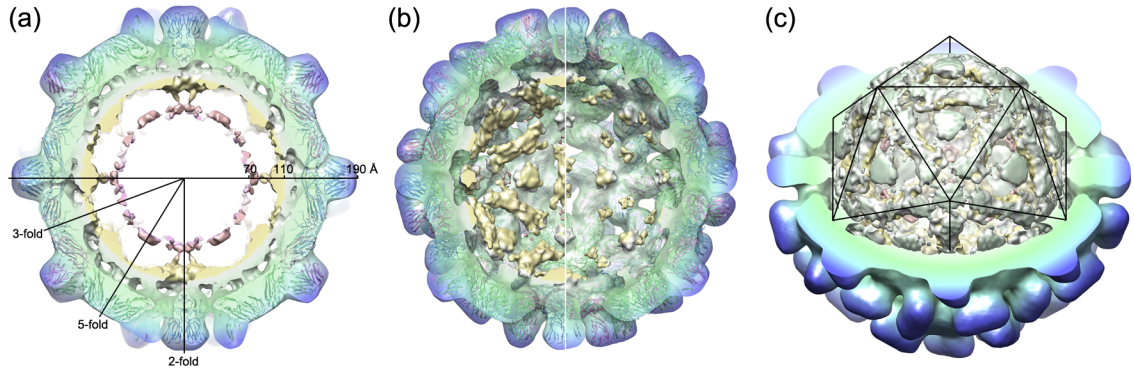


Figure 4.8 Internal structure of the expanded TCV virion
(a) A central section through the cryo-EM structure of the expanded TCV virion, contoured at 0.5σ. The EM density is transparent in the protein capsid region, revealing the fitted atomic coordinates. The colour scheme is identical to that used in figure 4.2. The density is coloured using a radial colour scheme, ranging from pink at a radius of 50 Å to blue at 190 Å (see horizontal scale bar). The icosahedral symmetry axes are also indicated. (b) Internal density at different contour levels. Left, a contour level of 1.1σ illustrates the changes in structure of the internal density from the native virion. A contour level of 1.7σ results in the same volume for the coat protein component of the virus as in the native reconstruction, and illustrates the loss of order in the internal density compared to the native virion. (c) A cutaway view of the virus showing the changed internal structure at a contour level of 1.1σ. An icosahedron is shown in black to indicate the symmetry of the particle. Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

The internal structure of expanded TCV is different to that seen in the native virus. When contoured to show the same volume for the coat protein, very little internal structure can be seen. At lower contour levels, however, a two-shelled internal structure similar in form to that in the native particle can still be seen (Figure 4.8),

although the innermost shell is extremely weak. The cage structure of the middle shell has changed, now forming an almost icosahedral structure consisting of bars of density between the 5-fold vertices (Figure 4.8c). The loss of internal structure in the expanded virus compared to native TCV could be due to two possibilities: a lower occupancy, i.e. loss of internal features through the pores of the virus, or a lower order in the RNA genome. As the radial expansion is 5%, the internal volume of the capsid will have grown by ~17%. This would mean that an identical amount of internal material would be less tightly packed, and this could lead to a lower degree of order. Increased breathing motions of the more dynamic expanded capsid can also have contributed to a loss of order.

4.3.3 Resolution analysis of expanded TCV

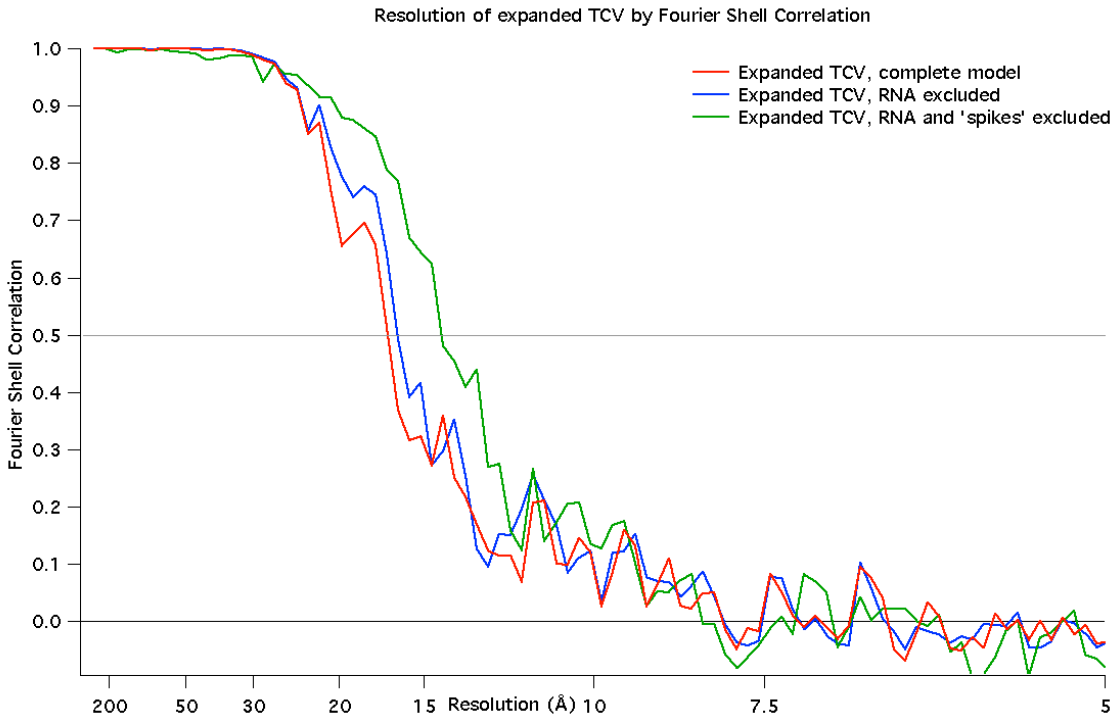


Figure 4.9 Resolution of expanded TCV by FSC
Showing the Fourier shell correlation (FSC) curves for the full reconstruction in red. This indicates the model has a resolution of approximately 17 Å. If the RNA is not taken into account, the resolution improves marginally to about 16 Å (blue curve). When, however, both RNA and coat protein P-domains or 'spikes' are ignored and only the coat protein shell is included in the calculation, the resolution of the model improves to just below 15 Å.

The expanded TCV reconstruction was made from a dataset containing ~5000 images. This and the inherently larger flexibility that is to be expected of expanded

TCV means the resolution was never expected to be as high as for the native dataset. The same resolution analysis was carried out as for the native virion and a similar pattern is seen: excluding the internal density from the resolution calculation results in a slight improvement, but excluding the P-domain 'spikes' causes a marked improvement in resolution (Figure 4.9).

4.4 Expansion mechanism of TCV

In Figure 4.10, the reconstructions of the native and expanded virions are shown next to each other. The most noticeable differences on the exterior are the appearance of the pores in the expanded capsid and the fact that the spikes are no longer connected. The loss of the internal structure is immediately apparent.

The five-fold and three-fold inter-subunit contacts stay intact, but the contacts between the asymmetric subunits are disrupted. In the native particle, six calcium ions, each coordinated by two glutamate or aspartate side chains, hold together the A, B and C-subunit that form one asymmetric unit of the icosahedral protein shell.

A closer look at the fitted atomic coordinates highlights the likely expansion mechanism of TCV. The EDTA in the expansion buffer chelates the calcium ions that mediate the coat protein subunit contacts in the native virus. This means the negatively charged side chains, no longer shielded from each other by the positively charged calcium ions, will repel each other, leading to the expansion of the virus particle (Figure 4.11). As expected, this mechanism is the same as that of TBSV expansion (Robinson and Harrison, 1982).

Although the expanded particles for this study were produced *in vitro*, there are indications that these conditions mimic those encountered on entering a plant cell. Because of its role in cellular signalling, the concentration of free calcium ions in plant cells is tightly regulated (Pittman and Hirschi, 2003). It is therefore likely that in the low-calcium conditions of the plant cell, the viral calcium ions may leave the capsid, causing the virion to expand (Evans, 1988).

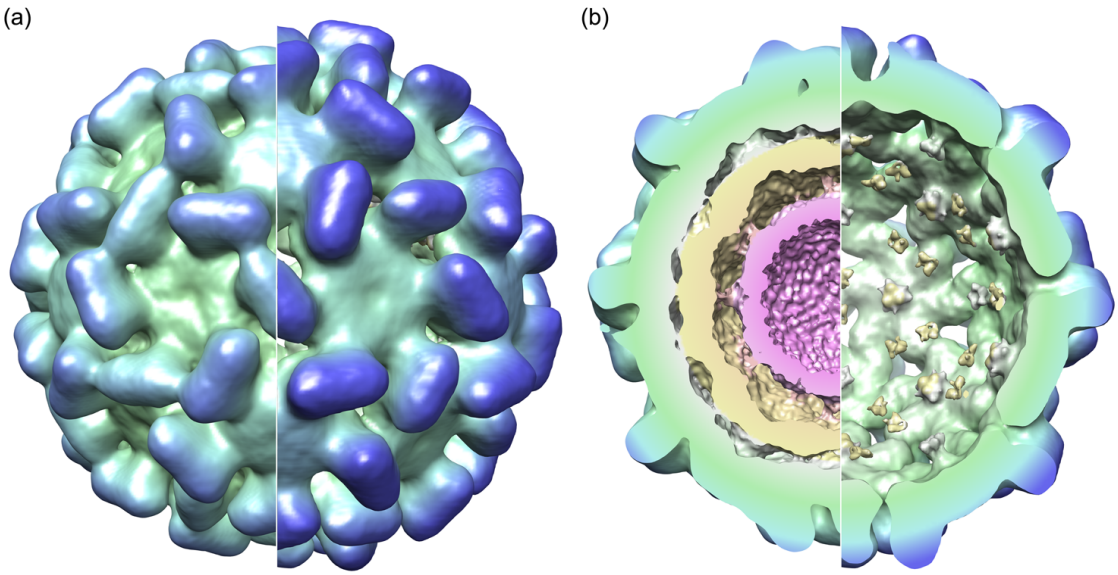


Figure 4.10 Comparison of native and expanded forms of TCV
(a) Surface representation of the exterior of both capsids, with the native virion on the left and the expanded form on the right. The capsids are coloured using the same radial colour scheme so the larger radius of the expanded particle is highlighted by the deeper blue colouration of the tips of the surface spikes. Both models are contoured to incorporate approximately the same volume of density for the coat protein region of the virus. (b) The back half of the capsids, also contoured to show the same volume of density for the coat protein region of the virus, showing the reduction in internal density in the expanded form of the virus.

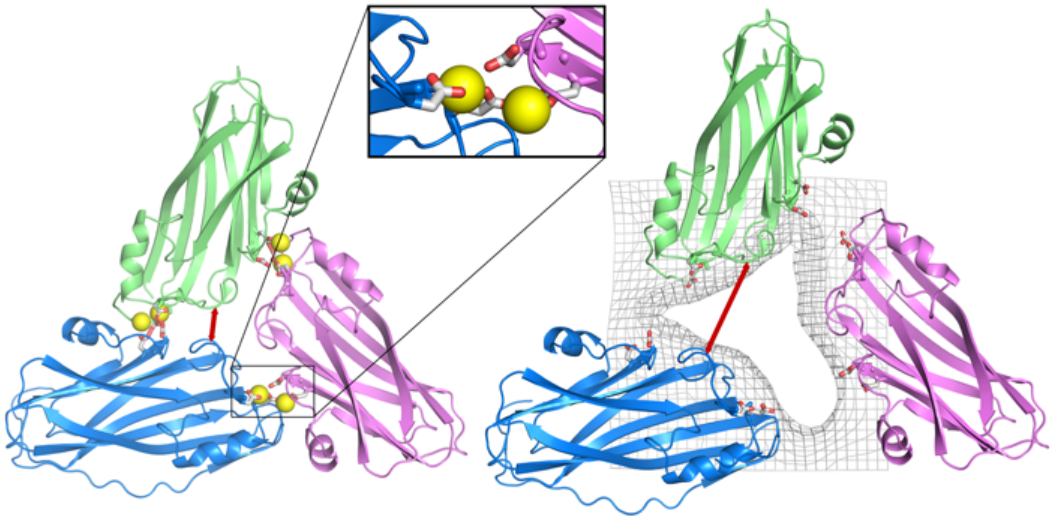


Figure 4.11 Expansion mechanism of TCV
Shows the expansion mechanism of the virus. Only the asymmetric trimer of TCV is shown for the native virus on the left and the expanded particle on the right. The calcium ions, shown as yellow spheres, are shown in the locations they occupy in the TBSV structure (PDB 2TBV); these positions are conserved between TCV and TBSV. In the native virus, the calcium ions are coordinated by acidic side chains on adjacent coat protein monomers (see inset). The virus expands when EDTA removes the calcium ions and the negatively charged residues repel each other. For the expanded virus, the edges of the pores created by the expansion as shown in the cryo-EM reconstruction are indicated as grey mesh. The distance indicated by the red arrows is 6.5 Å in the native particle and 14.5 Å in the expanded virus.

4.5 Biochemical studies of expanded TCV

In order to assess the functional significance of the expanded state of TCV, biochemical studies were carried out.

4.5.1 Formation of virus-ribosome complexes *in vitro*

For various plant viruses, such as southern bean mosaic virus (SBMV), alfalfa mosaic virus (AMV), brome mosaic virus (BMV) and cowpea chlorotic mottle virus (CCMV), it has been shown that addition of the expanded virus to cell-free translation mixtures leads to protein production (Brisco *et al.*, 1986) and, in the case of CCMV and SBMV, formation of virus-ribosome complexes termed striposomes (Roenhorst *et al.*, 1989; Brisco *et al.*, 1985).

When expanded TCV is incubated with wheat germ extract, complexes of ribosomes and virus particles are also formed (Figure 4.12). These complexes are sufficiently stable to survive centrifugation. These results suggest that the genome exits the expanded capsid long enough for the ribosome to bind the RNA, and that this may be the uncoating mechanism of the virus. Co-translational uncoating would allow the viral genome to emerge from the mostly intact capsid as single-stranded RNA, and avoid exposing the double-stranded regions of the genome to the RNA silencing mechanism. TCV, like many other viruses, was thought to release its genome by complete dissociation of the virus particle. However, this strategy would be a dangerous one, as it would expose the folded genome to the host cell defence mechanisms. The folded genome contains substantial regions of dsRNA, which is a potent trigger for initiation of RNA silencing (Baulcombe, 2004).

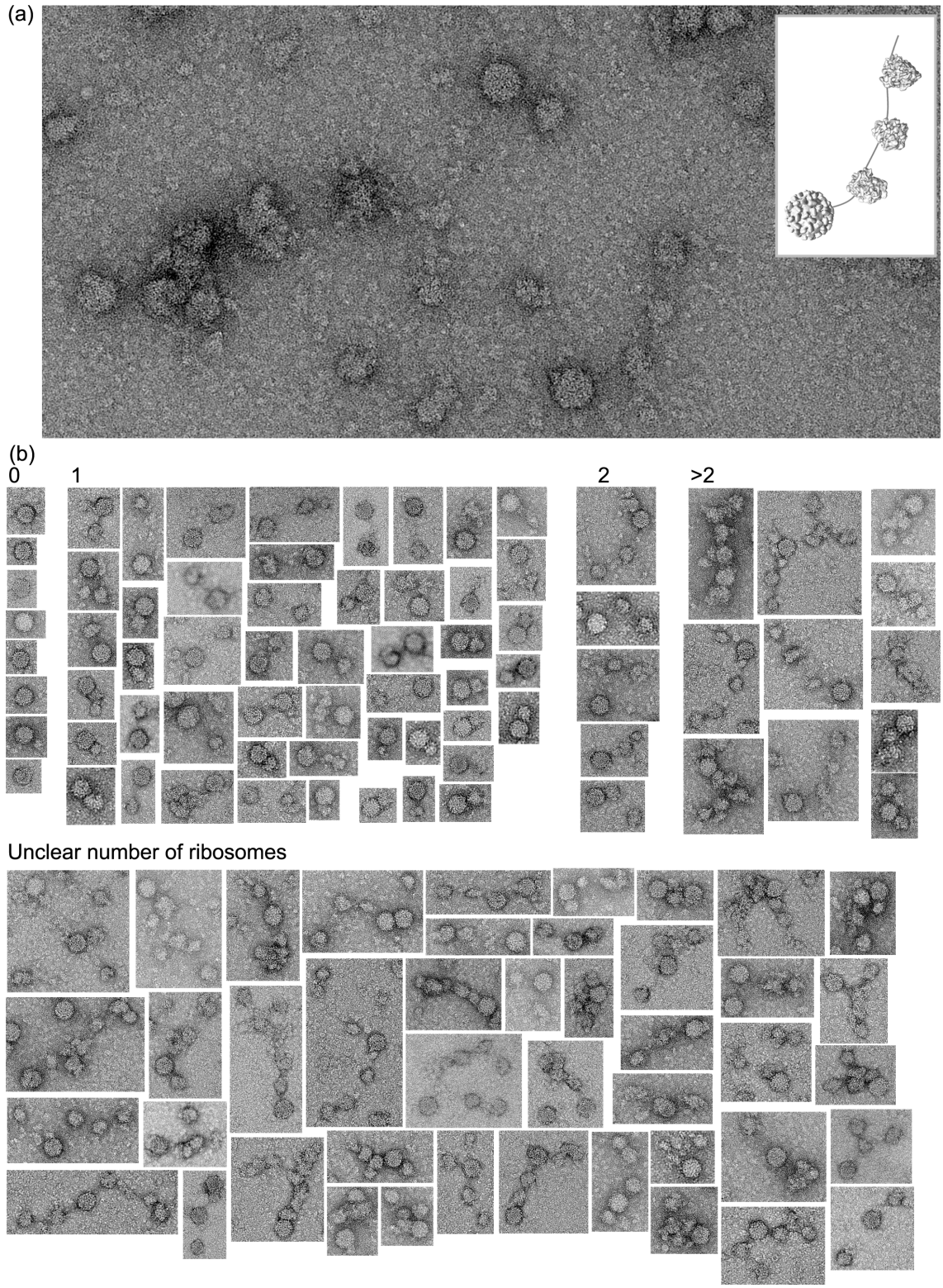


Figure 4.12 'Striposome' complexes are formed between TCV and plant ribosomes
(a) 'Striposome' complexes from a single negative stain electron micrograph. Inset: model of a 'striposome' complex showing that the ribosomes, in various orientations, resemble the particles in the EM images. (b) 'Striposome' complexes from 50 micrographs, arranged by number of ribosomes per virus particle. Examples of various stoichiometries are shown.

From the negative stain EM, it is not clear whether these complexes are polysomes, i.e. multiple ribosomes translating the RNA simultaneously, or whether the ribosomes are bound without producing protein. Repeated binding of ribosomes to the RNA outside the virus, together with the charge repulsion inside the particle, would prevent the movement of the genome into the virus particle and assist uncoating. No translation products have been detected in these reaction mixtures to prove that this process leads to protein translation. The formation of the striposome complexes also appears quite inefficient, although this is difficult to quantify.

4.5.2 Proteolytic cleavage of the coat protein

To assess whether capsid expansion alters the proteolytic sensitivity of TCV, the expanded virus was incubated with chymotrypsin for 60 minutes and samples were taken at various time points (Figure 4.13). The N-terminal region of the coat protein is normally located inside the virus capsid, where the positively charged amino acids neutralise the negative charge of the RNA. Therefore, in the native capsid, the coat protein should be protected from attack by the protease. Because the N-terminal domain of the TCV coat protein is disordered, it should be accessible to the protease if the virus is disassembled, while the two ordered domains remain intact.

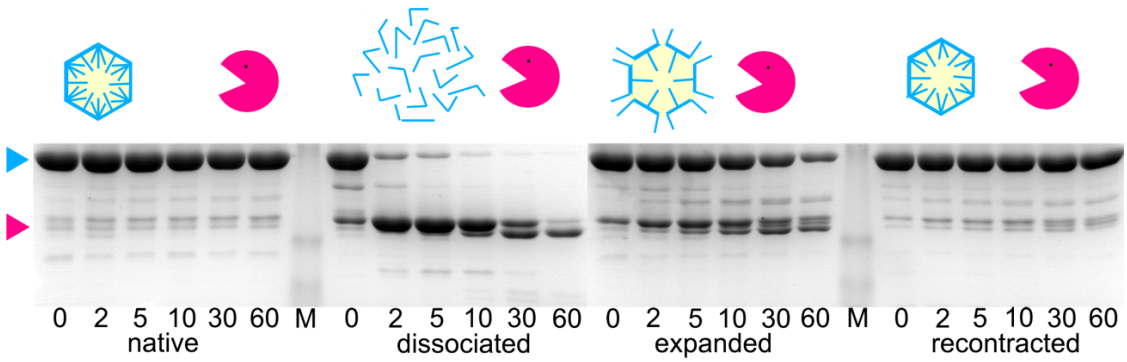


Figure 4.13 Proteolytic cleavage of coat protein subunits
SDS-PAGE gels showing protection against chymotrypsin cleavage in the native virus, while in dissociated virus all coat protein subunits are cleaved to 30 kDa cleavage products. In expanded virus, about two-thirds of the coat protein subunits gets cleaved within 60 minutes. In recontracted virus, very few if any coat protein subunits are cleaved.

As expected, SDS-PAGE analysis showed that in the native virus, the proteolytically sensitive sites of the coat protein are protected from chymotrypsin cleavage, whereas in the dissociated virus, all coat protein subunits are quickly cleaved to the two major cleavage products. In the expanded capsid, cleavage of the coat protein still occurs, but is significantly slower than for dissociated coat protein. Initially about two-thirds of the coat protein subunits are cleaved. When the recontracted virus is subjected to chymotrypsin cleavage, the result is much like that for the native virus, indicating the expansion of the TCV virion is apparently reversible to a native-like state.

Table 4.1 Coat protein cleavage products

N-terminus	Mass spectrometry	Predicted mass	Difference
Blocked (acetylated)	38,182.10	38,139.06 (38,182.11)	43.04 (0.01)
REVTTQ	30,802.50	30,802.44	0.06
TARDGITR	29,648.21	29,648.16	0.05

To determine the chymotrypsin cleavage sites, the cleavage products of the coat protein were analysed by mass spectrometry and N-terminal sequencing. The coat protein gene was also sequenced. Mass spectrometry (performed by Dr James Ault, University of Leeds) indicated the full-length CP has a molecular mass of 38,182.10 Da. Two different cleavage products were identified with molecular weights of 30,802.50 Da and 29,648.21 Da. N-terminal sequencing by Edman degradation (performed by Dr Jeff Keen, University of Leeds) shows that the full-length CP is blocked at the N-terminus and the 43.04 Da difference between the predicted mass of 38,139.06 Da and the experimental mass is consistent with N-terminal acetylation of the CP, which is common for plant virus coat proteins (Tsunasawa & Sakiyama, 1984). Cleavage products with the N-terminal sequences REVTTQ and TARDGITR were confirmed; predicted masses of the cleavage products starting with these sequences were 30,802.44 Da and 29,648.16 Da, respectively (summary in Table 4.1).

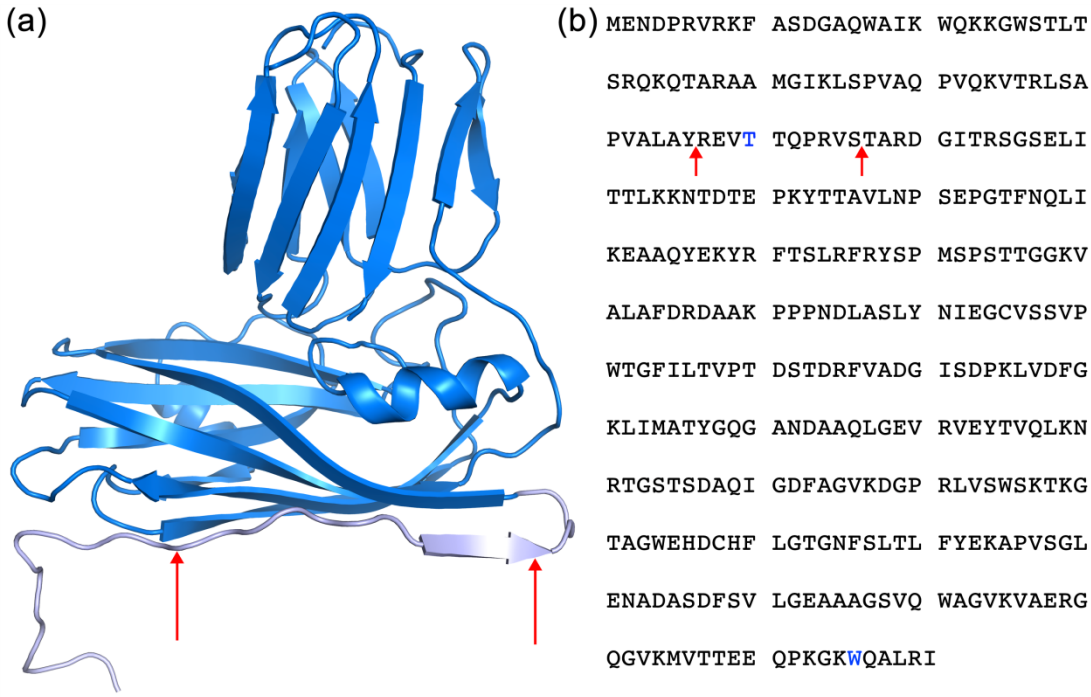


Figure 4.14 Cleavage sites on the coat protein and coat protein sequence
(a) The cleavage sites are indicated with red arrows. The C-subunit is shown in the figure; the part of the protein that is disordered in the A- and B-subunit is coloured in lighter blue. (b) The coat protein sequence obtained from sequencing the genome. Residues coloured in blue are different compared to the previously published structure (Carrington *et al.*, 1987). Coordinates for the TCV crystal structure (Hogle *et al.*, 1986) were kindly provided by Prof. Harrison.

The cleavage sites are situated in the N-terminal arm of the coat protein (see Figure 4.14), which is ordered only in the C-subunits and was not resolved in the A- and B-subunits of the crystal structure. This explains why, in the expanded particle, two-thirds of the coat protein subunits are cleaved but one-third appears protected; the cleavage sites in the more ordered C-subunits lie on the three-fold symmetry axes of the virion and are most likely inaccessible to chymotrypsin even in the expanded particle.

4.5.3 Ribonuclease protection assay

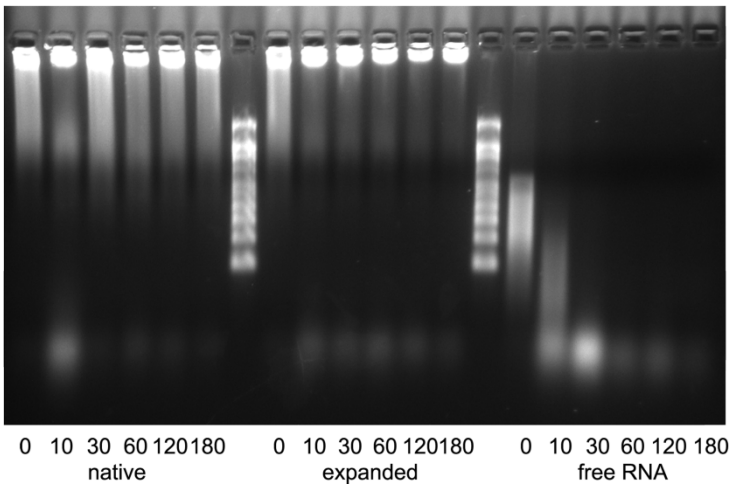


Figure 4.15 Ribonuclease protection assay
Denaturing agarose gel showing the RNA genome of native and expanded TCV after incubation with ribonuclease A. Protein-free genomic RNA (extracted from virus) dissolved in TCV expansion buffer is included as control. The lack of degradation in expanded virus indicates the genome continues to be protected from ribonuclease degradation.

As a part of the coat protein that is normally inside the protein capsid becomes accessible to chymotrypsin in the expanded particle, the genome may also be exposed. To test this, native and expanded TCV were incubated with ribonuclease and the nucleic acid was analysed on a denaturing agarose gel. Protein-free TCV genomic RNA in expansion buffer was used as a control. The ribonuclease rapidly degraded the control RNA (Figure 4.15). In the native capsid, as expected, the RNA genome is protected from nuclease and there is no degradation of the genome. Although the CP is accessible to chymotrypsin, the expanded virus still achieves protection of the genome from nuclease degradation for at least three hours (Figure 4.15). This suggests that the genome does not immediately exit the capsid through the pores, as the nuclease is too large to enter the capsid. This is indeed expected, as exposure of the genome to RNA silencing mechanisms in the plant cytosol would render the virus ineffective. This is confirmed by a reconstruction of the recontracted virus, which shows the same internal structure as the native, untreated virion, albeit at lower resolution (Figure 4.16).

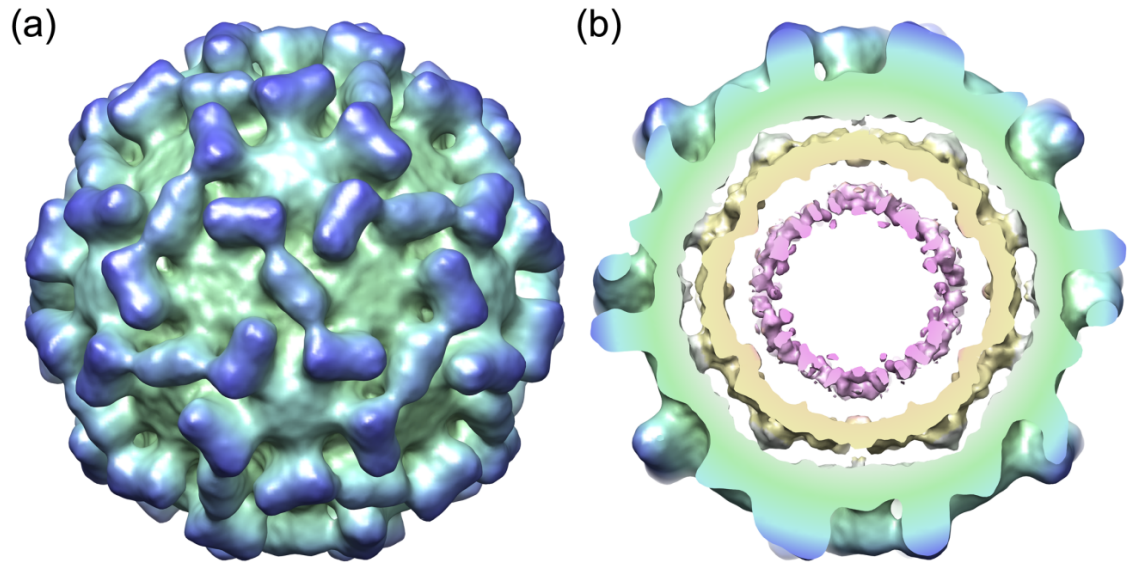


Figure 4.16 Reconstruction of the recontracted TCV virion
(a) Exterior view of recontracted virus, coloured as in figure 4.2 (b) Central slice of the reconstruction of recontracted virus, coloured as in figure 4.2. Virus was incubated in expansion buffer for ten minutes before recontraction by addition of a surplus of calcium and lowering the pH, and preparation of grids.

4.5.4 Probing TCV uncoating

The results presented above indicate that the expanded capsid continues to protect the viral genome. Analytical ultracentrifugation (AUC) was carried out to assess the changes in TCV during expansion. AUC separates molecules based on their sedimentation velocity, which gives information about the shape and size of the molecules. A higher sedimentation velocity (S) indicates a heavier particle. AUC is often used to analyse large molecules such as viruses, and identify different states (Schuck, 2000). As expected, the native particle (Figure 4.17, black trace) migrates as one species with a sedimentation coefficient of about 120S. Dissociated TCV shows peaks for the coat protein dimer and the salt-stable complex (Figure 4.17, yellow trace). The expanded particle (Figure 4.17, blue trace) did not give reproducible results with AUC, which suggests the expanded virion is less stable than the native and breaks apart under the large forces that are applied in the AUC. However, the results for expanded virus do not match those of dissociated virus.

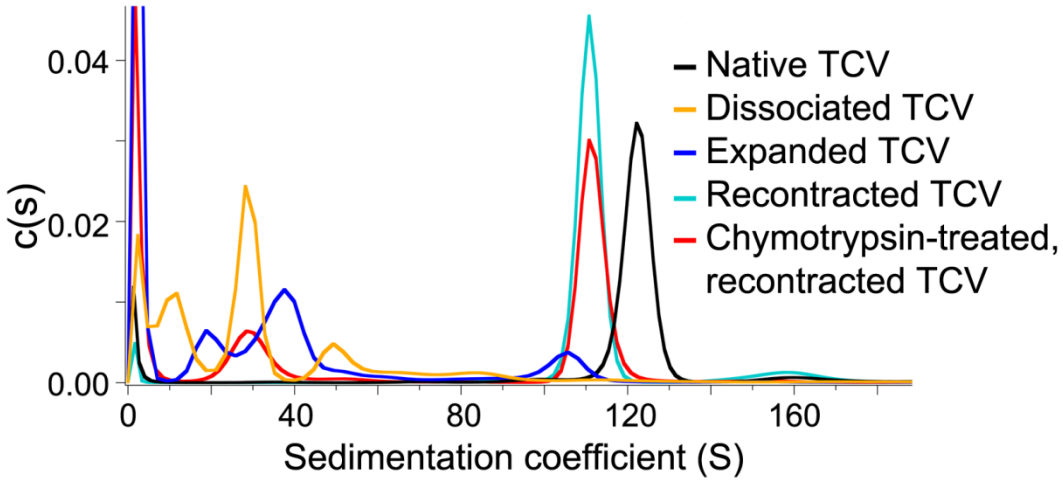


Figure 4.17 Analytical ultracentrifugation of TCV
Analytical ultracentrifugation of TCV: native in black, dissociated in yellow, expanded in blue, recontracted in cyan and chymotrypsin-treated recontracted TCV in red. AUC was performed by Amy Barker.

The most surprising result from the AUC analysis is that the recontracted virus (Figure 4.17, cyan trace) sediments more slowly than native virus (Figure 4.17, black trace), despite appearing native-like both in the chymotrypsin experiment and the cryo-EM reconstruction (Figure 4.16). The cryo-EM reconstruction, however, relies on icosahedral averaging. The most plausible explanation for this behaviour is that the expansion was not fully reversible and the recontracted particle has an asymmetric feature acting as a hydrodynamic drag on the particles in AUC. This can be explained in two ways: either the N-terminal regions of the coat protein that were extruded have not fully returned to their original place, or the covalent coat protein dimer, p80, of which each virion contains one copy, has been extruded and not returned.

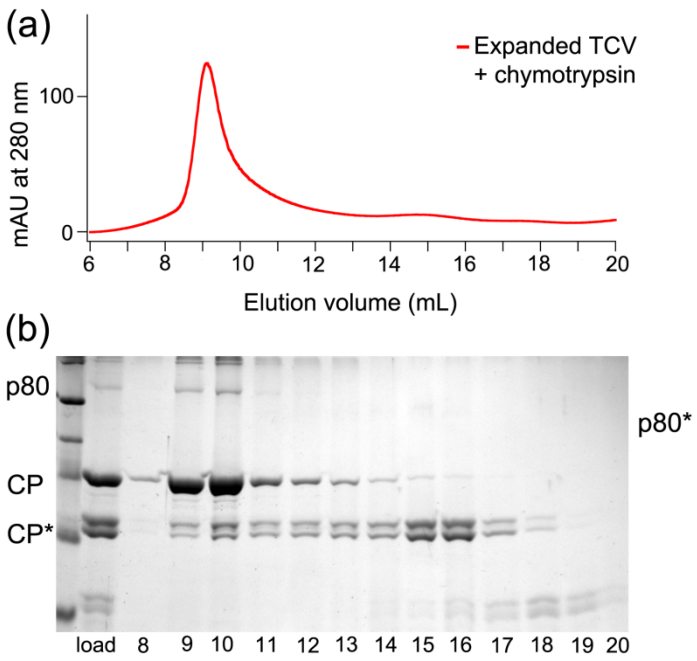


Figure 4.18 Gel filtration of proteolysed TCV
(a) Gel filtration trace of expanded, chymotrypsin treated TCV. (b) SDS-PAGE gel of the fractions after concentrating using a spin-concentrating device. Gel filtration and SDS-PAGE for this figure were performed by Robert Ford.

To probe whether any coat protein is released from the virus particles during proteolysis, expanded chymotrypsin-treated TCV was analysed on a gel-filtration column. In the sample that was loaded onto the column, about one-third of the coat protein is proteolysed. When expanded and chymotrypsin-treated TCV is separated by gel filtration, not all the protein migrates with the virus particles. The majority of intact coat protein is in the particles, whereas the proteolysed coat protein appears to be released and migrates separately as low molecular weight protein (Figure 4.18). The ratio of proteolysed to intact coat protein in the fractions is different from that in the sample that was loaded on the column: in the virus peak the majority of protein is intact, whereas in the later fractions the majority of the protein has been cleaved.

4.5.5 Negative-stain EM of chymotrypsin-treated expanded TCV

Purified chymotrypsin-treated expanded TCV was further analysed by negative stain EM. The results are striking: many of the particles exhibit an asymmetric feature extruding from the capsid (Figure 4.19a). As described in section 4.3.1, no such asymmetrical features were observed in untreated expanded virus. From 20

micrographs, all virus particles were picked, centered and classified by MSA. Rotational alignment to a representative class was then performed and particles were classified again. Representative class averages are shown in Figure 4.19b. Several of these exhibit the asymmetric feature, which may be extruding RNA or coat protein. The particles that do not show the asymmetric protrusion may have the protrusion in line with the rest of the particle, i.e. above or below the particle in the field of view, and do not necessarily lack it.

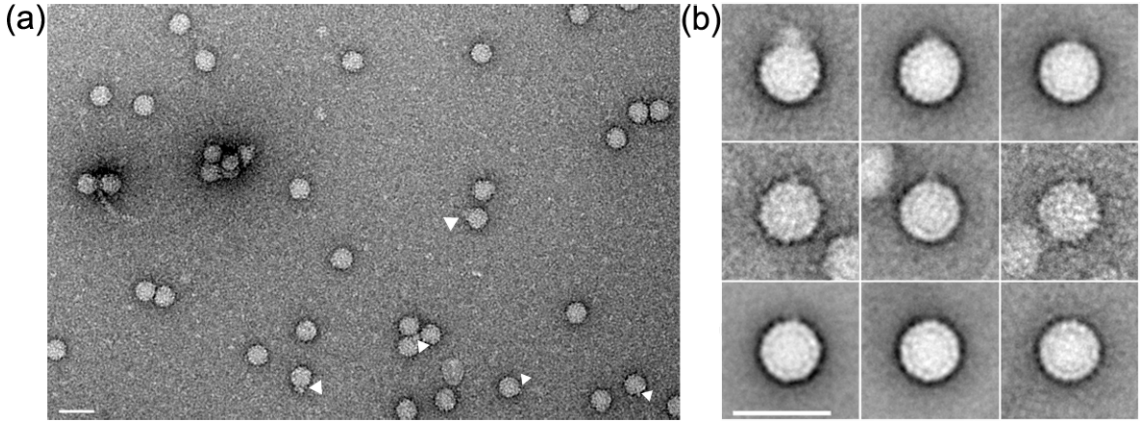


Figure 4.19 Negative stain of protease-treated expanded TCV
(a) A representative field of negatively stained protease-treated expanded TCV. Many of the particles show an asymmetric feature (indicated with white arrowheads). (b) Class averages of protease-treated expanded TCV. 9 images were selected from 50 class averages obtained by MSA classification. The scale bar in both panels represents 50 nm.

The most likely candidate for an asymmetric feature in TCV is the p80 coat protein dimer. From the gel filtration assay it is clear that after proteolysis the p80 dimer still migrates with the virus particle. If proteolysis is allowed to proceed for longer, however, all p80 is released from the virion, and migrates as a ~60 kDa product indicating the N-terminal regions have been cleaved by the chymotrypsin.

4.6 Discussion

The structures of the native and expanded TCV clearly reveal the expansion mechanism of the virus, while the *in vitro* formation of striposomes suggests an uncoating mechanism. Although both mechanisms had already been known to occur in plant viruses, this is the first virus for which both native and expanded structures, as well as biochemical characterisation, have been carried out for the same virus.

Additionally, it has been shown that after proteolysis, the virus loses some coat protein without the particles falling apart. Such protease-treated particles exhibit a clear asymmetric feature, which may be extruding RNA and coat protein, probably an intermediate on the way to uncoating. This suggests that proteolysis is an essential step in the uncoating process of TCV. Put together, these two mechanisms give a clearer view of the initial stages of infection by TCV, and possibly other plant viruses. A new model for TCV uncoating, based on the results discussed in this chapter, is summarised in Figure 4.20. This model helps explain how TCV manages to uncoat its genome without incurring responses from the host cell, such as the RNA silencing machinery. Plant viruses are known to encode proteins that suppress the RNA silencing system, and for TCV the coat protein is instrumental in this respect (Cao *et al.*, 2010). Releasing some coat protein through proteolysis before the genome is uncoated does in this respect seem particularly favourable. However, many other plant viruses, including TBSV, encode separate silencing suppressor proteins (Cao *et al.*, 2010) and for these it would be especially important to ensure the genome can be unpackaged from its protective protein coat and translated without eliciting host responses. RNA silencing is less of a concern for progeny genomes about to be packaged, as viruses commonly remodel the host cells or package their genome during replication (Laliberté and Sanfaçon, 2010; Nugent *et al.*, 1999).

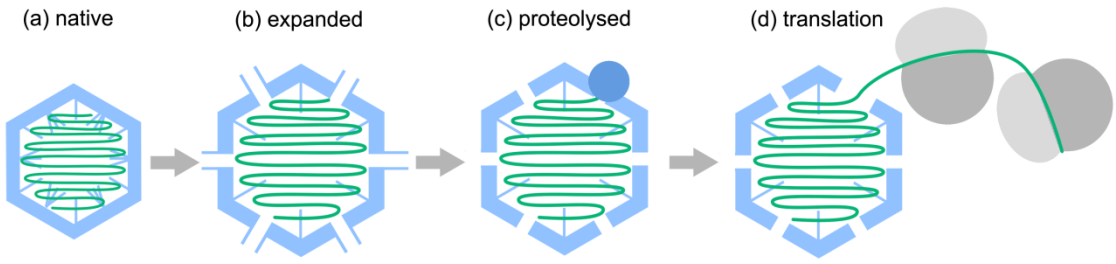


Figure 4.20 Schematic representation of the putative uncoating mechanism of TCV
(a) Native TCV as it is injected into the cell by an aphid feeding on the plant. (b) The low-calcium environment of the plant cell causes the virus to expand, and gives plant proteases access to the N-terminal regions of coat protein. (c) The proteolysis causes some of the cleaved coat protein to dissociate from the capsid and RNA extrudes, possibly bound to more coat protein. (d) The extruded RNA is now accessible to ribosomes and can be co-translationally uncoated.

After the results presented here have been taken into account, there are still some questions remaining about the uncoating process of TCV. Firstly, it is unclear whether the 5' or the 3' end of the genome exits the capsid first, and therefore whether the uncoating is co-translational. This could be tested by cross-linking the RNA to the capsid during uncoating and using labelled nucleotide fragments complementary to the 5' and 3' regions of the genome to determine which end extrudes. A second enigma is the p80 protein. Because of the stoichiometry of 1 covalent dimer to 89 non-covalent coat protein dimers, the p80 protein is hard to study. However, it is enriched in the rp-complex and, with sufficient quantities of virus and improvement in analysis techniques, it may be possible to determine the nature of the cross-link by mass spectrometry, although efforts to date have proved unsuccessful. Its function is also unknown, as is the mechanism of cross-linking. Although possible, it is unlikely that a host protein is responsible for the cross-link, as it would likely result in a more irregular stoichiometry. As the RNA genome is the only asymmetric feature of the virus, and RNA is well known for being able to catalyse a wide variety of reactions (Shechner and Bartel, 2011), it is possible that the cross-linking reaction is catalysed by the RNA genome, either directly or through coordination of a metal ion. This could be investigated by testing reassembly products of reactions using coat protein monomers and protein-free RNA for the presence of covalent dimers. The genome structure could be further investigated by making asymmetric 3D cryoEM reconstructions, either using reassembly products with labelled genome, or sub-tomogram averaging.

5 Native and empty particles of equine rhinitis A virus

5.1 Introducing equine rhinitis A virus (ERAV)

Having proposed a new model for TCV uncoating, we wondered if a similar model would be applicable to other viruses. To investigate, we set out to determine the structure of ERAV.

Equine rhinitis A virus (ERAV) is a member of the *Picornaviridae*, a family of small, non-enveloped viruses with a positive-sense, single-stranded RNA genome. The viruses in this family are classified in several genera. Poliovirus and the major and minor group rhinoviruses make up the *Enterovirus* genus, the *Hepatovirus* genus includes human and simian hepatitis A virus, the genus *Aphthovirus* consists of ERAV and foot-and-mouth disease virus (FMDV), and the *Cardiovirus* genus includes mengovirus and Theiler's murine encephalomyelitis virus (Tuthill *et al.*, 2009). The picornaviruses measure about 30 nm in diameter and have a *T*=1 icosahedral architecture comprised of 60 copies of each of four coat proteins (Figure 5.2). Many of these viruses cause devastating diseases in humans or animals. Rhinoviruses are the causative agent for 30-50% of cases of the common cold (Eccles, 2005), while poliovirus causes poliomyelitis, which mainly affects children and can result in irreversible paralysis (Nathanson, 2011).

FMDV routinely causes major economic damage to livestock, manifesting as high fever and formation of blisters around the mouth and feet. Because of its fast replication and short incubation time, animals infected with FMDV are usually culled, as vaccination is not used in the UK. The economic loss from the outbreak in the United Kingdom in 2001 was estimated to be 3 billion pounds in the agricultural sector. A further 3 billion pounds was estimated to be lost in tourism to the affected area (Thompson *et al.*, 2002). The disease is endemic in many African and Asian countries (Grubman and Baxt, 2004). Because of the high risk involved in working with FMDV, a

good model system for the study of FMDV is highly desirable. Like FMDV, ERAV is a member of the *Aphtovirus* genus of the *Picornaviridae*, and as such is closely related to and suitable as a model system for this virulent pathogen. The natural host for ERAV is the horse and infection in the upper respiratory tract manifests as cold-like symptoms (Li *et al.*, 1996).

5.1.1 Picornavirus infection events

The genomes of the picornaviruses are 7000 to 8500 nucleotides long, encompassing a single open reading frame and untranslated regions (UTR) at the 5' and 3' ends (Figure 5.1a). The 5' UTR contains an internal ribosome entry site (IRES) and other control elements. The 5' end is bound to the small peptide VPg. The 3' UTR is shorter and ends with a poly(A) tract (Tuthill *et al.*, 2010). The nomenclature of the *Picornaviridae* proteins has been standardised and the functions of the proteins are well conserved within the family (Rueckert and Wimmer, 1984). Once the genome has entered the cytoplasm, the single open reading frame is translated as a polyprotein. Assembly of poliovirus and FMDV has been most extensively studied and the mechanism of assembly for these two viruses is very similar (Hogle, 2002; Mason *et al.,* 2003; Grubman and Baxt, 2004), and it is assumed other picornaviruses assemble in the same way.

In the aphthoviruses and cardioviruses, the first part of the polyprotein is the L-protease (L^{pro}), which autocatalytically cleaves itself from the remaining polyprotein. L^{pro} targets cellular eIF4G, a translation initiation factor required for recruiting ribosomes to capped cellular mRNA. Capping, which occurs for the vast majority of eukaryotic mRNAs, is the modification of mRNA by addition of 7-methylguanosine (m^7G) via a triphosphate linkage. For the translation of viral proteins, only the C-terminal cleavage product of eIF4G is required, as the viral genome uses the IRES to initiate translation rather than m^7G capping (Mason *et al.*, 2003; Grubman and Baxt, 2004). The

enteroviruses lack the L-protease (L^{pro}), and eIF4G cleavage is performed by the 2A protease instead (Bovee *et al.*, 1998).

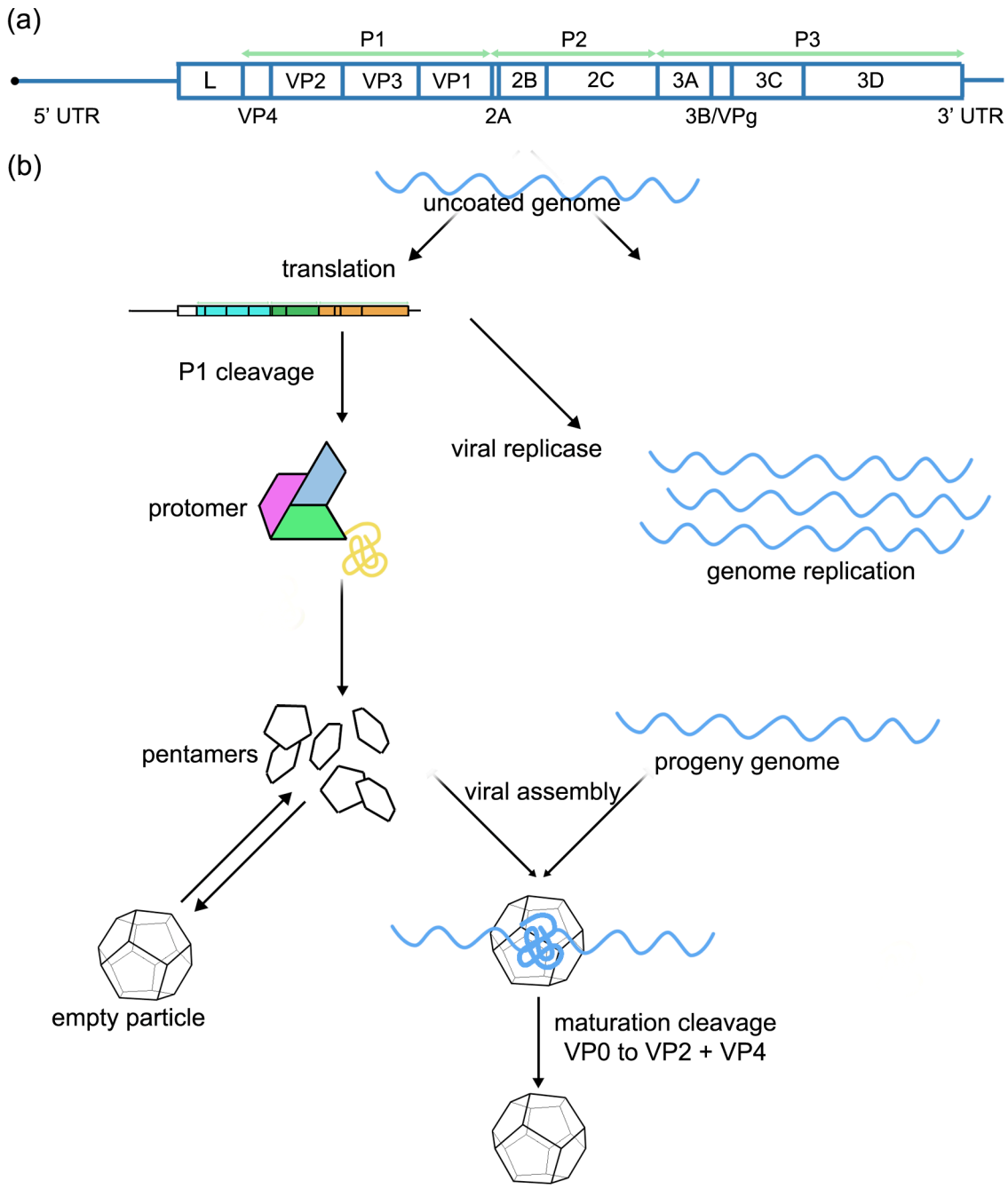


Figure 5.1 Picornavirus assembly
(a) The structure of the picornavirus genome. The genome shown is that of FMDV. The figure has been adapted from Grubman and Baxt, 2004 (b) Picornavirus assembly. The P1 region of the polyprotein is cleaved into VP0, VP1 and VP3, forming the protomer. Five protomers assemble into a pentamer and 12 pentamers assemble around the progeny genome into particles. The mature particle is formed by autocatalytic cleavage of VP0 to VP2 and VP4. The figure has been adapted from Hogle, 2002.

The remainder of the polyprotein has three distinct regions based on the functions of the protein. In FMDV, the 18-amino acid 2A protease cleaves C-terminal to itself

resulting in the P1/2A cleavage product, containing the capsid proteins and 2A, and the P2/P3 fragment containing all other viral proteins. Although further cleavage of the P1 region is required for assembly, the resulting VP0, VP1 and VP3 protomers never dissociate. After cleavage, the protomers assemble into coat protein pentamers (Hogle, 2002; Mason *et al.,* 2003). These assemble around the RNA genome during genome replication, rather than forming empty capsids that are subsequently filled (Nugent *et al.*, 1999). Maturation by autocatalytic cleavage of the VP0 coat protein into the VP2 and VP4 proteins is required for infectivity and is thought to prime the particle for RNA release upon receptor binding. The P2 and P3 regions of the polyprotein contain all other viral proteins. As a consequence of the single translation initiation site on the picornavirus genome, one would expect a full set of the non-structural viral proteins to be produced for each coat protein precursor molecule, resulting in a molar excess of non-structural proteins compared to capsids. However, the action of the 2A protease appears to cause a 2- to 5-fold molar excess of the N-terminal portion of the polyprotein, i.e. the coat proteins, compared to the C-terminal portion, redressing the balance somewhat (Donnelly *et al.,* 2001).

Cells infected with poliovirus or FMDV exhibit morphological changes, including the formation of membranous vesicles, obtained from the cellular pathway for membrane protein glycosylation (Egger *et al.*, 2000). Viral proteins 2B, 2BC, 2C, 3A and 3AB all interact and co-localise with the membranous vesicles and expression of 2BC or 2C caused formation of similar vesicles. 2B has also been shown to increase membrane permeability and disrupt secretory pathways, implying these proteins may induce vesicle formation, while protein 2C is required for the synthesis of negative-strand RNA (Mason *et al.*, 2003). 3CD and its cleavage product 3C are proteases, performing all polyprotein cleavages except between P1 and P2/3 and the capsid maturation cleavage of VP0 into VP2 and VP4. 3D is the polymerase, which creates both negative-sense template RNA and positive-sense progeny genomes. It is currently unknown what determines the switch between producing template or progeny. 3B or

VPg is a small uridinylated peptide that acts as a primer for the polymerase during synthesis of negative-sense template RNA as well as progeny genomes. RNA synthesis takes place at the surface of the vesicles mentioned above (Grubman and Baxt, 2004; Mason *et al.*, 2003).

5.1.2 The structure of the virion

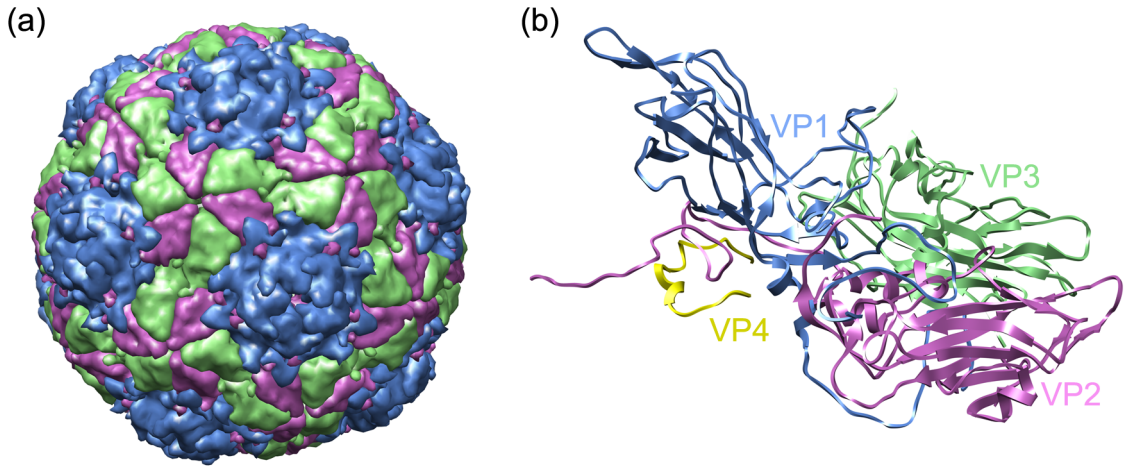


Figure 5.2 Structure of ERAV
(a) The capsid structure of ERAV. Surface representation of PDB entry 2WFF (Tuthill *et al.*, 2009). (b) The four structural proteins of ERAV, with VP1 in blue, VP2 in pink, VP3 in green and VP4 in yellow.

The capsid of all picornaviruses, including ERAV, consists of 60 copies each of four coat proteins, named VP1, VP2, VP3 and VP4 in descending order of size (Figure 5.2a). Although VP1, VP2 and VP3 differ in size, they share the jelly-roll fold that also forms the basis of other virus capsids including TCV (Figure 5.1b; Stuart, 1993; Hogle *et al.,* 1986; Hogle *et al.*, 1985; Tuthill *et al.*, 2009; Acharya *et al.*, 1989). As such, picornaviruses can be considered to have a pseudo-*T*=3 architecture, if the three different structural proteins are regarded as quasi-equivalent subunits (Tuthill *et al.*, 2010). VP4 has no apparent structural role in the mature virion, but is located on the inner surface of the capsid surface (Tuthill *et al*., 2009; Hogle *et al.*, 1985; Acharya *et al.*, 1989). VP4 is myristoylated *in vivo* and has been suggested to play a role in virus uncoating.

The ERAV particle contains a single copy of the 7734-nucleotide positive-sense single-stranded RNA genome (Figure 5.2c). In some picornaviruses, VP0 is not cleaved into VP2 and VP4, and VP4 is not myristoylated in hepatitis A virus. Enteroviruses have a hydrophobic pocket within VP1, which is occupied by a pocket factor. The pocket appears to play a role in uncoating (see section 5.1.2).

As expected, the structure of ERAV is very similar to that of FMDV. The capsid is about 30 nm in diameter. Unlike other picornaviruses, which exhibit a deep canyon on the virus surface where the receptor binding site is hidden, the aphthoviruses have a smooth surface, and lack the hydrophobic pocket (Tuthill *et al.*, 2009). Of the three major coat proteins, VP1 is the most different from FMDV and has two extended loops compared to FMDV, giving it a slight protrusion around the five-fold axis (Tuthill *et al.*, 2009). In addition to a crystal structure from crystals grown at pH 7.0, a second crystal structure was published from crystals grown at pH 4.6. Although ERAV, like FMDV, is known to dissociate into pentamers at low pH, the differences between the two structures are small. The N-terminus of VP2 is ordered in the structure at pH 7.0, and stabilises the pentamer interface. While the N-terminus of VP3 is ordered in both structures, it forms a loop on the pentamer interface at low pH, but stabilises the VP2-VP3 interface at pH 7.0. Of VP4, only residues 16 to 36 could be resolved as a helix located across the N-terminus of VP2. No density for the RNA genome was seen in either of the structures (Tuthill *et al.*, 2009). A crystal structure at 4 Å resolution of ERAV with sialyllactose bound, as a mimic of the receptor, has also been published. Although sialic acid is present in about 50% of the binding sites, it does not appear to have changed the virus structure in the vicinity of the binding site (Fry *et al.*, 2010). Sialyllactose does inhibit infection *in vitro* for ERAV, but not poliovirus, and removing sialic acid from the cell surface enzymatically also inhibits infection (Fry *et al.*, 2010).

5.1.3 Uncoating of picornaviruses

The details of the uncoating process are not very well understood, and a variety of pathways are used by the different picornaviruses. They enter the cell by endocytosis, which can be mediated by clathrin, caveolin or other pathways (Tuthill *et al.*, 2010). The receptor used for cell attachment and entry determines the tissue specificity of the virus. Some picornaviruses, including FMDV, use an arginine-glycine-aspartate (RGD) motif on the virion surface to target integrins, while poliovirus uses the poliovirus receptor (PvR) and rhinoviruses target ICAM1 or the LDL-receptor (Tuthill *et al.*, 2010). ERAV binds to sialic acid on an unknown receptor and enters cells by clathrin-mediated endocytosis (Fry *et al.*, 2010).

For poliovirus and other enteroviruses, receptor binding or acidification of the endosome causes the formation of A-particles, which have lost a significant proportion of their VP4 proteins (Hogle, 2002). Additionally, A-particles have an altered, more hydrophobic surface and have become sensitive to proteases, indicating that some previously buried parts of the protein have become surface-exposed. Although the A-particles have lost their receptor-binding capacity, they remain infectious in a receptor-independent manner (Hogle *et al.*, 2002). Poliovirus RNA release can be initiated *in vitro* by heat treatment, soluble receptor or low ionic strength conditions (Wetz and Kucinski, 1991).

In addition to the A particle, two distinct forms of empty poliovirus capsids have been identified. The 80S particle is a second uncoating intermediate for poliovirus, and represents a later step in the uncoating process when the RNA has been released. It contains no RNA, and like the A particle has lost the majority of its VP4. The 75S particle, on the other hand, is an assembly intermediate with uncleaved VP0 that can be dissociated into assembly-competent pentamers (Levy *et al.,* 2010). Recent work on the poliovirus uncoating process has produced tomographic evidence that the genome is released via a pore near the two-fold icosahedral symmetry axes, between

pentamers, rather than via the five-fold pore as had been previously thought (Bostina *et al.*, 2011).

For the *Aphthoviruses* and *Cardioviruses* no stable uncoating intermediates have been observed, leading to the belief that these viruses disassemble directly into pentamers (Berryman *et al.*, 2005). However, considering the hostile conditions in the endosome, this seems an unlikely strategy for a virus to pursue. Indeed, it has been shown for ERAV that acidification in the early endosome causes structural changes that enable genome release without exposing the genome to the endosomal conditions (Groppelli *et al.*, 2010). Inhibition of late endosome formation has no effect on infection rate, but inhibition of early endosome formation completely disabled the virus, indicating the early endosome is critical for infection (Figure 5.3). Infection in the presence of ribonuclease results in the co-localisation of virus particles and enzyme in endocytic vesicles, but does not affect infectivity rates, indicating that the genome leaves the endosome without ever being exposed to the endosomal contents or co-transfected ribonuclease (Groppelli *et al.*, 2010).

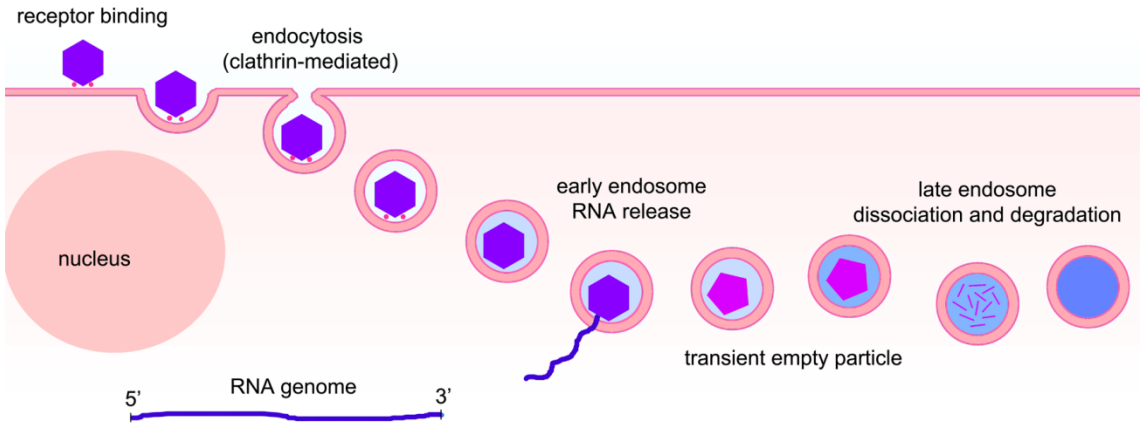


Figure 5.3 ERAV cell entry
ERAV enters the cell by clathrin-mediated endocytosis and releases its genome from the early endosome without exposure to the endosomal contents.

5.2 Cryo-EM reconstructions of ERAV

ERAV, purified by sucrose gradient centrifugation, was applied to grids and imaged in the electron microscope. A small dataset of 822 particles was obtained and used to

make a three-dimensional reconstruction. As the dataset was small and the particles do not have a distinctive morphology, a starting model was obtained from a crystal structure. Although an ERAV crystal structure is available, the poliovirus structure (PDB entry 1asj) was used to mitigate the effects of model bias (Wien *et al.*, 1997).

5.2.1 Reconstruction of native ERAV

Of the 822 particles in the dataset, 260 (~31%) were used to obtain the final reconstruction, which has a resolution of about 25 Å. A side-by-side comparison of all the averages and corresponding reprojections is given in Appendix C. The crystal structure fits well into the map without requiring any modifications (Figure 5.4). Compared to poliovirus, which was used as a starting model, the projections at the five-fold and canyons on the virus surface are less pronounced. This is also seen when the crystal structures of poliovirus and ERAV are compared. The deep canyon around the five-fold axis of the poliovirus capsid houses the PvR binding site, protecting it from antibody binding. The receptor of ERAV is sialic acid, which binds to a site partially shielded by the EF-loop of VP1 (Fry *et al.*, 2010).

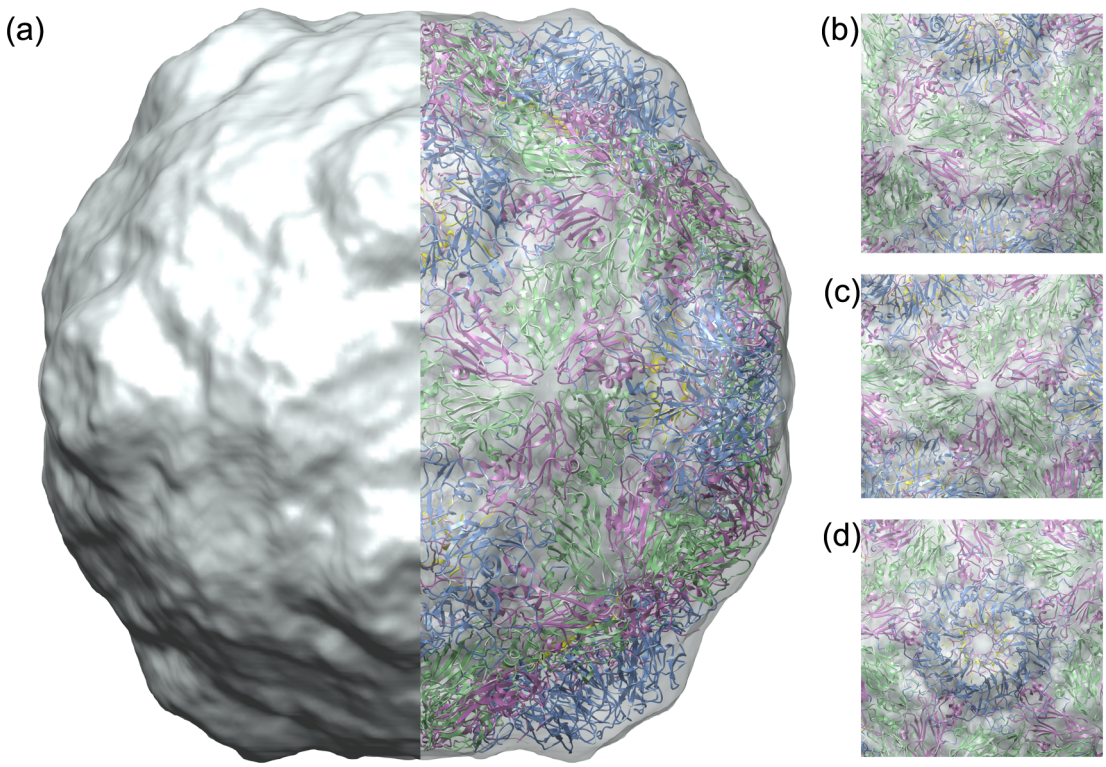


Figure 5.4 Cryo-EM reconstruction of native ERAV
(a) Surface view and atomic structure fitting of the native ERAV virion with on the left, a solid grey surface representation of the cryo-EM density. On the right, the surface has been made transparent in order to show the fitted coordinates of the ERAV crystal structure (PDB entry 2WFF, Tuthill *et al.*, 2009) shown in ribbon representation (coloured as in Figure 5.2a). The atomic coordinates are coloured as follows: VP1 in blue, VP2 in green, VP3 in magenta and VP4 in yellow. (b)-(d) Close-up views of the fitted coordinates, perpendicular to the (b) 2-fold, (c) 3-fold and (d) 5-fold symmetry axes of the particle.

As for TCV, there is considerable density inside the viral capsid that, in this case, can almost entirely be attributed to the RNA genome, although it must also contain those protein residues that were not observed in the crystal structure: the N- and C-terminal regions of VP4 and the N-terminal 11 residues of VP2.

The strongest density forms a dodecahedral cage just under the coat protein shell, apparently complementing the positive charges on the coat protein, which form a long thin patch approximately perpendicular to the two-fold symmetry axes, along the inside of the capsid. Weaker density reaches from the dodecahedral shell into the centre of the capsid. Although in this reconstruction a central feature is present, this is likely to be an averaging artefact. Additionally, there are small patches of density on the five-fold symmetry axes directly under the coat protein (Figure 5.5).

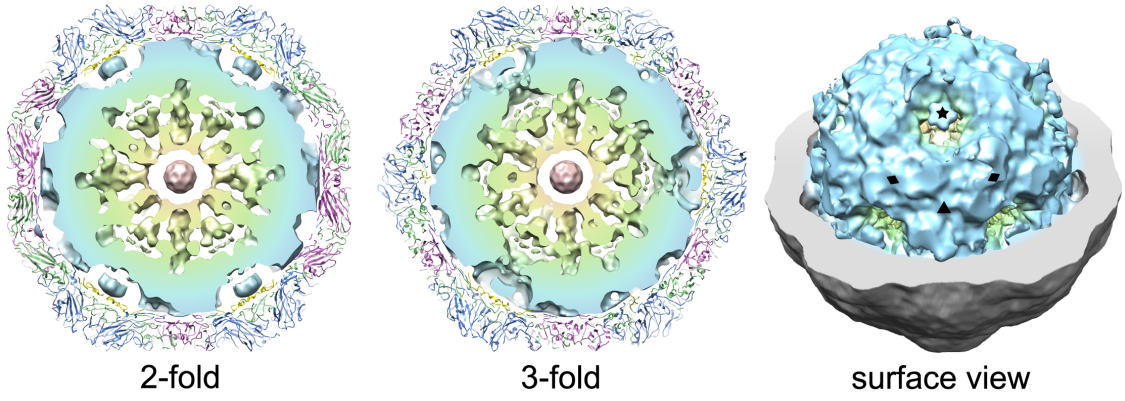


Figure 5.5 Internal density in ERAV
Central slice of the ERAV density after subtraction of the capsid density shown from the two-fold and three-fold axis, as well as a cutaway view of the capsid showing the dodecahedral RNA cage of ERAV. Capsid coordinates (PDB entry 2WFF, Tuthill *et al.* 2009) are shown as ribbons and coloured as in Figure 5.2. In the cutaway view the two-fold symmetry axis is indicated with a diamond, the three-fold with a triangle and the five-fold axis with a star.

5.2.2 Expanded ERAV particles

Although the sample had been extensively purified at pH 7.0, two distinct species were observed in the micrographs of ERAV unexpectedly. As well as the native particles, expanded particles were seen. From the raw micrographs, it is immediately obvious that the expanded particles contain very little, if any, RNA (Figure 5.6b). This was possibly caused by prolonged storage of the sample at -20°C and -80°C and freeze-thawing of the sample. A dataset consisting of 663 expanded particles was selected from the same micrographs as the native particles. 227 or 34% of these images were used to generate the final 3D reconstruction. A side-by-side comparison of all the averages and corresponding reprojections is given in Appendix D. The particle surface exhibits large holes (Figure 5.6a) and there is no evidence of any internal density for the RNA genome (Figure 5.6c). The reconstruction shows a marked difference between the expanded and the native particle. Pentamers, the putative assembly and disassembly unit of ERAV and other *Picornaviridae*, were fitted into the density and fit well. The changes observed in the low-pH crystal structure relative to native ERAV result in a better fit in the cryo-EM density, although in both cases the N-terminal region of VP1 appears to be outside the cryo-EM density, suggesting this part

may be disordered in the empty particle. No density, however, was observed for the VP4 position of the pentamers, indicating VP4 may have been released from the particle (Figure 5.6b).

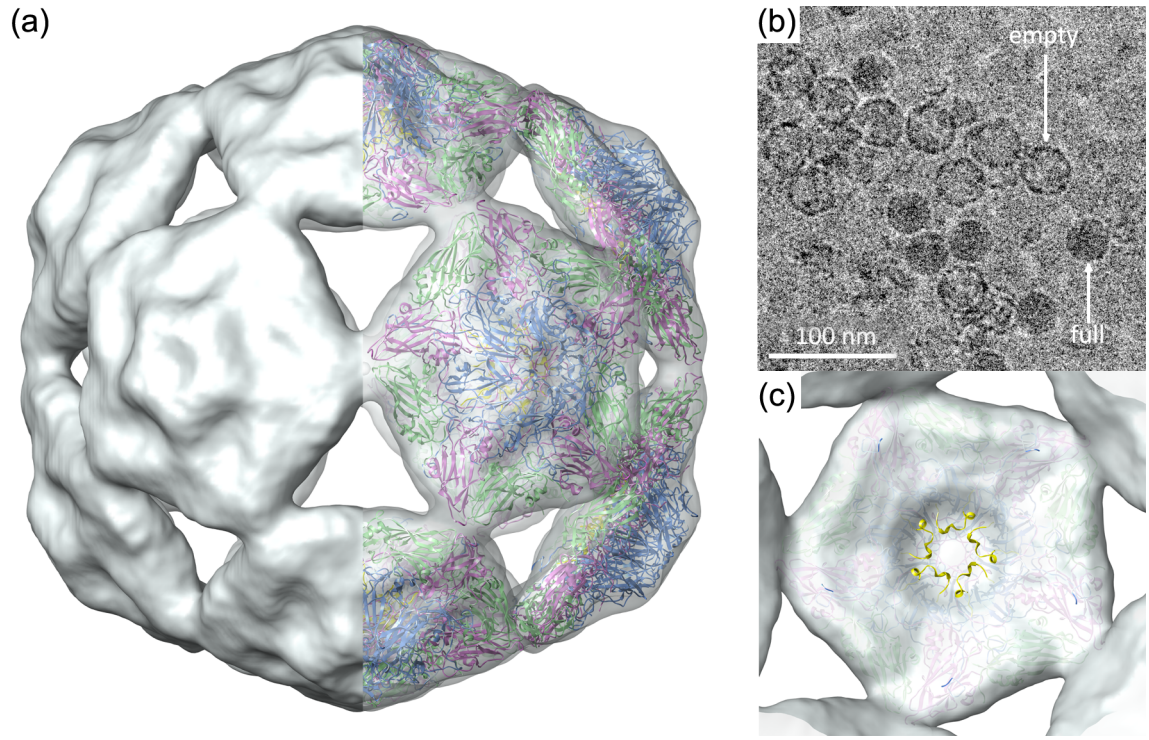


Figure 5.6 3D reconstruction of empty ERAV
(a) Surface view and atomic structure fitting of the ERAV empty particle with on the left a solid grey surface representation of the cryo-EM density. On the right, the surface has been made transparent in order to show the fitted coordinates of the ERAV low-pH crystal structure (PDB entry 2WS9, Tuthill *et al.*, 2009) shown in ribbon representation (coloured as in Figure 5.2a). The atomic coordinates are coloured as follows: VP1 in blue, VP2 in green, VP3 in magenta and VP4 in yellow. (b) Raw micrograph of the ERAV sample showing the difference between full and empty ERAV. (c) View of one pentamer from the inside of the particle. No density is seen for the VP4 protein.

5.2.2.1 Fourier-space vs. real-space reconstruction methods

During the refinement of the ERAV dataset, a problem was encountered: the reconstructions that were made using the same protocol as the TCV reconstructions (chapter 4.2.1), could not be fitted with the crystal structure. Although the crystal structure fitted well within the outer surface of the map, when looking at a slice through the map the coat protein shell was not thick enough to encompass the entire crystal structure: some of the crystal structure was in low-density regions of the map. Initially the problem was attributed to the size or quality of the dataset. Various modifications to

the protocol were attempted but not found to improve the map in this respect. Finally the reconstruction protocol was changed to use real-space refinement (BP RP) instead of Fourier space refinement (BP 3F) in all cycles of the refinement, which resulted in a final map that corresponds well to the crystal structure.

5.2.2.2 Resolution of the native and expanded ERAV reconstructions

The resolution of the ERAV reconstructions has been determined as ~25 Å for the native particle, while the resolution for the empty particle is ~17 Å (Figure 5.7a). Although the native dataset consisted of a slightly larger number of particles, the very smooth appearance of the virus may have resulted in less accurate alignment which results in a lower resolution. By contrast, the distinctive features of the empty particle result in more accurate determination of the particle orientation.

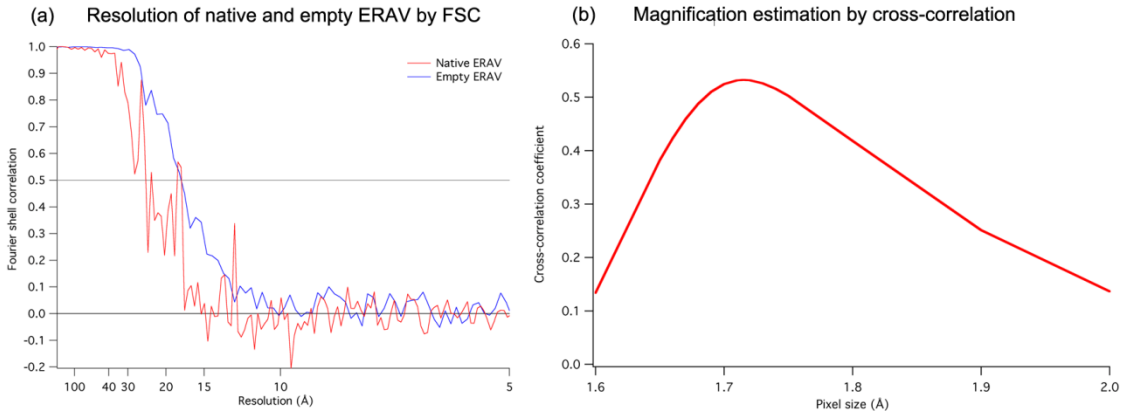


Figure 5.7 Resolution and magnification of ERAV
(a) Fourier shell correlation curves of native ERAV (in red) and empty particles of ERAV (in blue). The resultions are ~25 Å for the native reconstruction and ~17 Å for the empty particles. (b) The magnification was estimated to be 1.72 Å/pixel, by cross-correlation of the map with density obtained from the crystal structure

5.2.2.3 Magnification calibration for native ERAV

As there is a crystal structure available for native ERAV, the magnification of the microscope can be calibrated by cross-correlation of the cryo-EM reconstruction to the crystal structure at different magnifications. The result of this calculation is shown in Figure 5.7b. The curve is asymmetric because the internal density in the cryo-EM reconstruction means the cross-correlation of smaller models is higher, as there is still density present. The lack of RNA in the crystal structure also accounts for the low

(<0.6) overall correlation coefficient between the X-ray structure and the cryo-EM reconstruction.

5.2.3 Structural changes in ERAV

From even a cursory examination of the native and expanded structures, it is immediately obvious a large structural change must take place during the transition between the two. In the native virion, the interface between adjacent pentamers has a large buried surface area containing positively and negatively charged areas placed opposite each other. These oppositely charged areas should form favourable interactions and assist assembly. The pentamer vertices are slightly offset relative to each other (Figure 5.8a). By contrast, the interaction interface of the expanded particle of ERAV is small and does not contain strongly charged surface areas, although some weakly charged areas are present and complement each other (Figure 5.8b).

The binding affinity of the pentamer interactions can be calculated *in silico*, by taking the difference between a numerical estimation of the Gibbs free energy for the complex as a whole and separately for the components of the complex (Schymkowitz *et al.*, 2005; Guerois *et al.*, 2002). This calculation has been performed for physiological pH 7.0 and pH 5.5, the endosomal pH. The complex is subjected to energy minimisation using the CHARMM19 force field and EEF1 implicit solvent and the interaction energy is calculated for the complex every 2 femtoseconds over a period of 0.4 picoseconds using FoldX (Schymkowitz *et al.*, 2005; Guerois *et al.*, 2002).

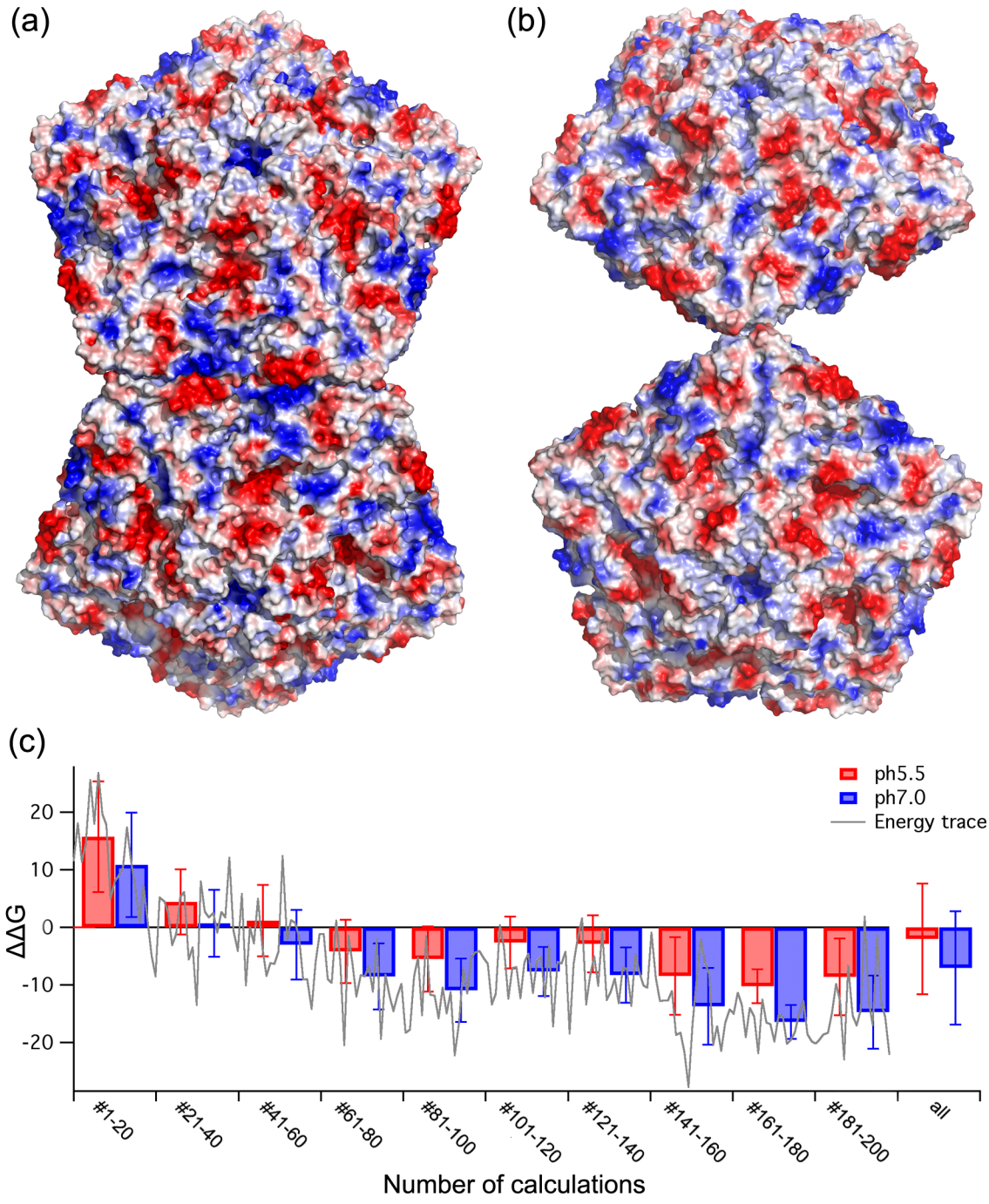


Figure 5.8 Contact interfaces of native and empty ERAV
(a) Contact interface of two pentamers in native ERAV, taken from the crystal structure 2WFF. The pentamers are shown in a surface representation and coloured by charge, with positively charged areas coloured blue and negative areas red. The charge map was obtained by vacuum interactions in PyMOL. (b) Contact interface of two pentamers in empty ERAV, after rigid-body fitting into the cryo-EM reconstruction. (c) Average ΔΔG (interaction energy) in kcal/mol for twenty consecutive calculations (indicated with numbers) and average for all calculations (far right). Lower energies indicate a more favourable interaction. The values for individual calculations have been included for comparison (grey trace). Interaction energy calculations were carried out by James Ross.

The results of these calculations indicate the native pentamer interaction is more favourable at pH 7.0 than at the lower pH (Figure 5.8c), as the complex energy is

consistently lower for the complex at pH 7.0 (calculations were carried out by James Ross). A longer simulation may have been more informative, but on the other hand it could result in a deviation from the crystal structure.

Although this does not explain the formation of the empty particle, it suggests the acidification of the endosome weakens the pentamer interaction as expected if a pore is to be formed between pentamers.

5.2.4 Discussion

Very few expanded particles (3 empty to ~2000 native) are seen when freshly purified native virus is frozen. The high number of expanded particles in the sample used here (~600/~800) indicates that prolonged storage or repeated freeze-thawing may have caused the transition to expanded particles. Freeze-thawing can be used to obtain empty particles of turnip yellow mosaic virus (TYMV; Adrian *et al.*, 1992), and this has been used to study the genome structure of this virus (Bottcher *et al.*, 1996). For picornaviruses, changes in temperature, pH or ionic strength are all known to cause genome release and produce empty particles *in vitro* (Wetz and Kucinski, 1991). However, the empty particles identified for other picornaviruses are far more native-like than the empty particle presented here. Although poliovirus empty capsids can form *in vivo* and accumulate when genome reproduction is inhibited, they have not undergone the maturation cleavage and can disassemble into assembly-competent pentamers. This suggests empty capsids and pentamers exist in equilibrium and the genome is required to stabilise the capsid (Hogle, 2002). During assembly, VP1, VP3 and VP0 assemble into pentamers before VP0 is further processed to VP2 and VP4 to form the mature particle. In poliovirus, the VP0 cleavage site is located in a position occupied by VP1 in the mature particle, preventing the ordering and formation of the network of protein that stabilises the mature particle (Hogle, 2002). A similar network is present in native ERAV, but appears to be weakened at the pentamer interfaces in the low-pH crystal structure (Tuthill *et al.*, 2009).

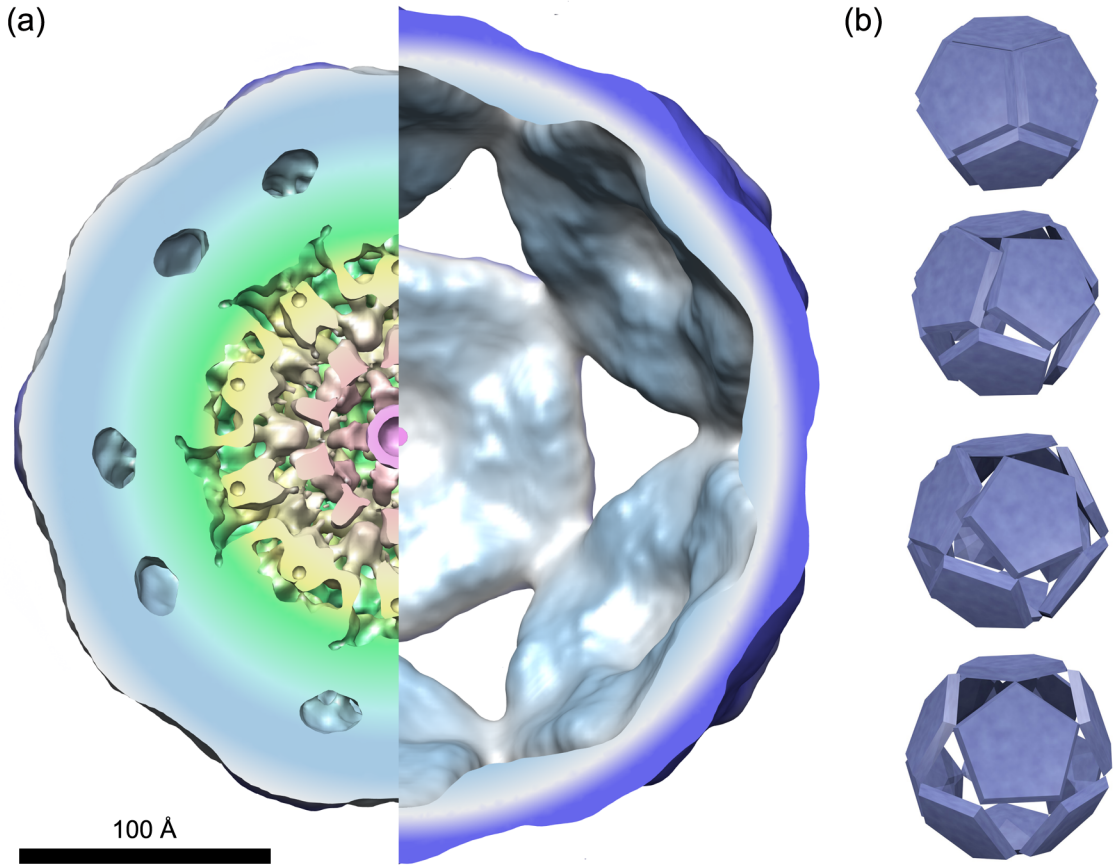


Figure 5.9 Transition from full to empty ERAV
(a) Showing native and expanded ERAV next to each other on the same scale. Both models are coloured according to the same radial colour scheme, from pink at 0 Å radius to purple at the outer edge. The scale bar is 100 Å. (b) Four frames from a movie illustrating a possible transition mechanism that does not required the release of any pentamers. The entire movie can be found on the Appendix CD.

In the empty ERAV particle presented here, there is no density for the genome or for VP4 (Figure 5.10). This suggests it is a disassembly rather than an assembly intermediate. It is unlikely that particles that have disassembled *in vitro* during storage have spontaneously adopted the less favourable conformation required for assembly. In poliovirus, the cleavage of VP0 into VP2 and VP4 releases the unfavourable disorder in the N-terminal region of VP1 and allows the ordered conformation of the mature particle.

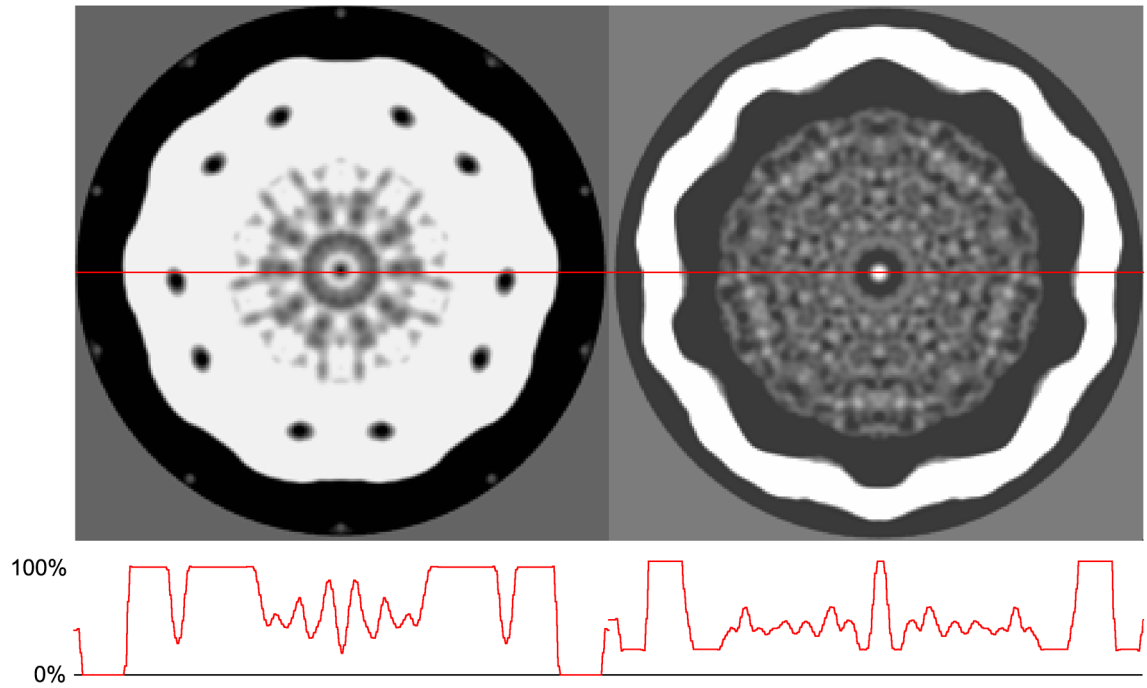


Figure 5.10 Density profiles for native (left) and expanded (right) ERAV
The density profile is shown for the centre of the particle. Shown is a central slice through the map and the line for which the density profile is shown is indicated in red in each image. Both images are shown in greyscale and normalised, with highest density shown in white and lowest in black. Although there is some density within the expanded particle, it is markedly lower than the coat protein density, unlike the native particle where the majority of the internal density is as high as the coat protein density.

The expanded particle presented here is very different from the disassembly intermediates seen for other picornaviruses. The A particles that are formed by poliovirus and enteroviruses, and even the empty 80S particles created after poliovirus genome release, do no exhibit any obvious holes and appear to look more like the native particle. However, the position of the holes in the ERAV expanded particle corresponds to where the genome is released in poliovirus tomograms (Bostina *et al.*, 2011). In picornaviruses, the strongest interactions are those within the protomers, closely followed by the intra-pentamer interactions. A pore within either of these units would be unlikely unless dramatic changes occur in the interaction interface. The inter-pentamer interaction is weakest and forms last during the assembly process. This is therefore the most likely interaction to be disrupted by pore formation as seen in poliovirus previously and the ERAV empty particle presented here (Bostina *et al.*, 2011). A more recent paper has shown that in poliovirus, the N-terminal portion of VP1

also localises around the two-fold symmetry axis in the empty 80S particles, but not in the A particles where it localises at the propeller tip (Lin *et al.*, 2011). This is consistent with VP1 being involved in pore formation and genome release at the two-fold axis during uncoating.

ERAV uncoating *in vivo* takes place in the slightly acidic conditions of the early endosome, and an ERAV structure was published from crystals grown at pH 4.6. This low-pH crystal structure shows comparatively few differences from the native particle: the N-terminus of VP2 has become disordered and been replaced by the N-terminal loop of VP1. In the native particle, these two loops stabilise the pentamer interface (Tuthill *et al.*, 2009). Calcium ions that were present at the three-fold icosahedral symmetry axes in the native particle were also seen in the low-pH structure and no pore formation or expansion were seen. Although no evidence for RNA was seen, this was not expected considering the limited resolution range of the X-ray data (15-3.5 Å), which was collected at room temperature from a limited number of small crystals (Tuthill *et al.*, 2009). The lack of pores in the structure and the continued presence of VP4 suggest that the genome is also still present in the capsid and another intermediate may exist during uncoating. This suggests the low-pH crystal structure may represent a species primed for RNA release, like the A particle of other picornaviruses, stabilised by the rapidly formed crystal contacts or the high ionic strength of the crystallisation buffer, while the 3D cryo-EM reconstruction of empty particles presented here may be a post-release intermediate, similar to the 80S particle in poliovirus (Figure 5.11). At this resolution it is not possible to say whether the native cryo-EM reconstruction presented here most resembles the low-pH crystal structure or the native crystal structure.

Further investigation is required to provide more evidence for this model. Evidence of the existence of either the low-pH particle or the empty particle in infected cultured cells would provide support for this model of uncoating. It is possible that the empty particles form in experiments *in vitro*, but are detected as pentamers because of their

instability. A single-particle cryo-EM reconstruction of ERAV in the low-pH crystallisation buffer would prove whether these particles retain the RNA genome, although the high ionic strength of the crystallisation buffer may prove problematic in cryo-EM reconstructions.

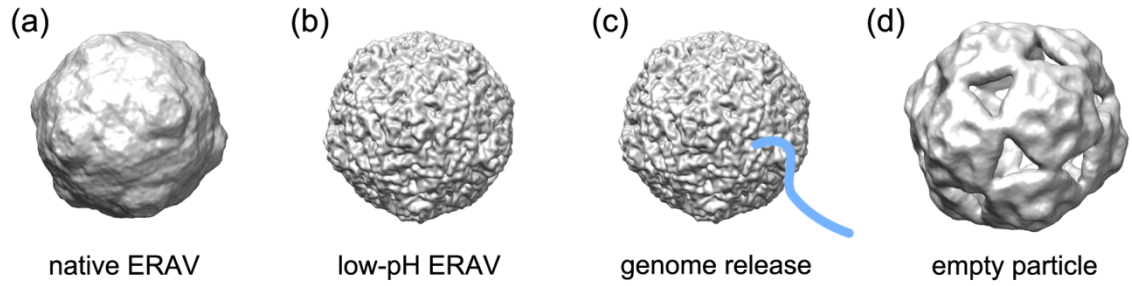


Figure 5.11 Proposed model for ERAV uncoating
Native particle binds surface receptor and triggers endocytosis. (b) Acidification of the endosome causes conformational changes resulting in the low-pH particle (c) A pore is formed through externalisation of VP4 and the N-terminus of VP1 and the RNA is released through the pore at the weakened interface between pentamers (d) Empty particle or dissociation to pentamers

It would also be interesting to study the time course of the transition from full to empty particles. The inter-pentamer interaction in the expanded particle is probably weak, and it is likely the particle will dissociate into pentamers in different ionic strength or during centrifugation, making it hard to study. The presence of the empty particles indicates current storage conditions are not optimal and should also be optimised to provide a better starting material for the study of ERAV. Potentially, the formation of empty particles may give an artificially low estimate for the infectivity of the virus.

Ideally, however, one would image the virus in the process of RNA release. This would show whether the site of RNA release corresponds to the holes seen in the reconstruction of the empty particles. In order to facilitate isolation of virus particles during RNA release, cells could be supplemented with 4-thiouridine during virus replication. The modified nucleotide is incorporated into the viral genome and can subsequently be cross-linked to the capsid using UV light without damaging the virus (Dr. Tobias Tuthill, personal communication). Furthermore, studies with receptor-containing membranes, like those that have been done for influenzavirus (Lee, 2010), could shed light on the mechanism of export of the RNA genome from the early

endosome without exposure to the endosomal contents. As influenzavirus also uses sialic acid as a receptor, the viruses may uncoat under similar conditions *in vitro.*

6 Crystallographic characterisation of STNV virus-like particles

6.1 Introduction to STNV

The aim of this chapter is to elucidate the interactions involved in the sequence specificity of the protein-RNA binding in satellite tobacco necrosis virus (STNV) by X-ray crystallography. Although there have been studies of the STNV genome at lower resolutions, I wanted to resolve the protein-RNA interaction at higher resolution afforded by crystallography. STNV is one of the smallest viruses known. It is a *T*=1 icosahedral virus with a diameter of 17 nm. It has a 1039-nucleotide, single-stranded positive-sense RNA genome, which encodes only the coat protein. The coat protein consists of 195 amino acids and 60 copies are required to make up the capsid (Jones and Liljas, 1984).

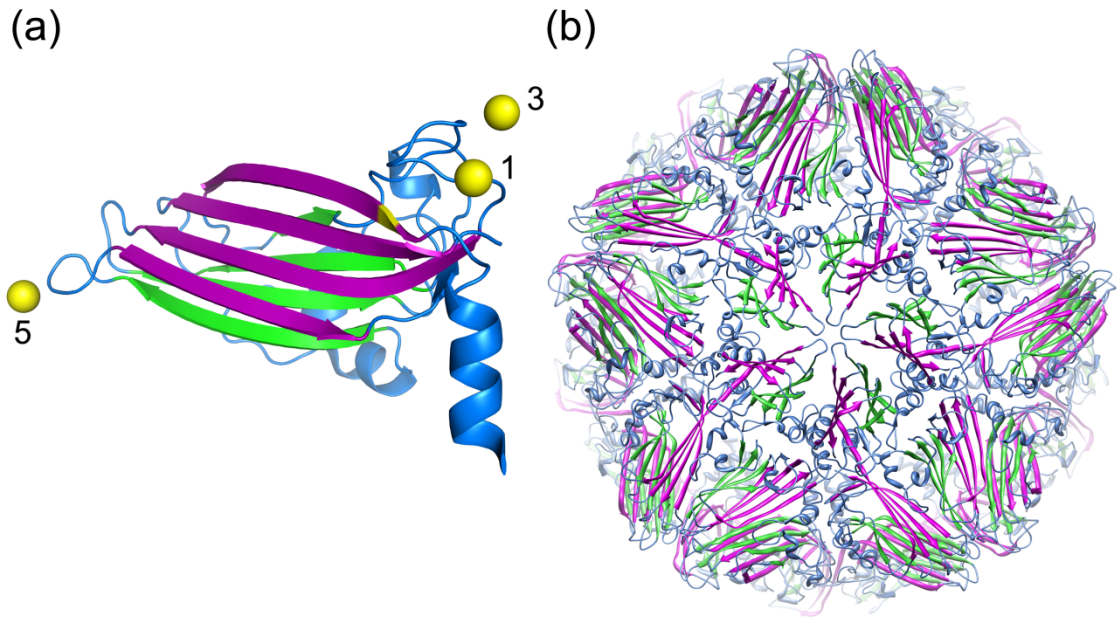


Figure 6.1 Coat protein structure of STNV
(a) Coat protein structure with the calcium ions shown as yellow spheres. The calcium ions are numbered to indicate their position: 1 is positioned within the coat protein subunit, 3 sits on the icosahedral thee-fold axis, and 5 sits on the icosahedral 5-fold. Also indicated in yellow is the chymotrypsin cleavage site Arg28. (b) Capsid structure of STNV. The coat protein is coloured as in (a).

As the genome only encodes the coat protein, STNV cannot infect a plant and multiply on its own – it requires the replication machinery of its 'helper' virus, tobacco necrosis virus (TNV). TNV is a *T*=3 icosahedral plant virus that is not antigenically related to STNV. Both viruses are transmitted by the fungus *Olpidium brassicae* (Montelius *et al.*, 1988; Oda *et al.*, 2000).

Several crystal structures are available for STNV. The structure of wild-type virus was refined to 2.5 Å in the 1980s (Jones and Liljas, 1984). The coat protein has a jelly-roll fold, like many other viruses including TCV and ERAV, that consists of two four-stranded beta-sheets (Figure 6.1a). The coat protein is wedge-shaped, with the narrow end of the wedge at the five-fold symmetry axes and the wide end spanning between three-fold symmetry axes. The N-terminal 12 residues were not resolved, and residues 12 to 24 form a helix. The N-terminal region, including the 12-residue helix, contains many basic residues and is thought to interact with the RNA genome.

The capsid structure contains 92 calcium ions. There is one calcium ion per coat protein subunit, coordinated by the side chains of Asp194, the carbonyl oxygens of residues 61 and 64, and a symmetry-related Glu25 side-chain. Additionally, there is one calcium ion on each five-fold axis, coordinated by the symmetry-related carbonyl oxygens of residue 138, and one on each three-fold axis coordinated by the side-chains of symmetry-related Asp55 (Jones and Liljas, 1984).

Treatment with EDTA causes a small expansion of the STNV virion in a pH-dependent manner: at pH 5 to 6 the radial expansion is 2%, at pH 6.5 the expansion is 4.5% and at pH 7 the expansion is 7% (Unge *et al.*, 1986). As for TBSV and TCV, EDTA removes the calcium atoms bound to the capsid and this results in the negative charges on the protein repelling each other. The pH dependency is consistent with protonation of the negatively charged side chains, and consequently less charge repulsion, at lower pH. The expanded virions can be recontracted by addition of calcium, magnesium or barium and this results in particles that sediment like the native virion (Unge *et al.*, 1986). The three types of calcium binding site have been shown to

have different calcium binding affinities. Calcium is most easily removed from the three-fold axis binding site, followed by the calcium ion coordinated within the coat protein subunit. The calcium binding site at the five-fold has the highest affinity. The EDTA-treated virus contained 25±10 calcium ions, compared to 92 in the native virus (Unge *et al.*, 1986). This is consistent with the loss of the calcium ions from the three-fold axis binding site and most of the calcium ions bound within the subunit, but retention of the 12 calcium ions bound at the five-fold icosahedral symmetry axes. Unlike the native virus, EDTA-treated STNV is sensitive to trypsin, which cleaves the N-terminal helix at Arg28 (Figure 6.1a; Unge *et al.*, 1986).

Four different crystal forms have been obtained from EDTA-treated STNV (Montelius *et al.*, 1988; Unge *et al.*, 1986; Montelius *et al.*, 1990). Type I crystals were crystallised at pH 6.5, after pre-treatment with EDTA and dialysis into water. The radial expansion compared to native STNV appears to be 4.5% and this was confirmed by AUC. The resolution of this structure was 7.5 Å and it was analysed by rigid-body fitting of the high-resolution native structure. The subunits appear to have moved outwards, with the movement most pronounced at the three-fold symmetry axes, where most calcium ions have been lost and charge repulsion is greatest. The two-fold and five-fold inter-subunit contacts are mostly retained. As expected, no density is seen in this structure for the calcium atoms (Montelius *et al.*, 1988).

Type III crystals grew in the same crystallisation conditions after shorter pre-treatment with EDTA. These crystals have the same unit cell dimensions as the native virus and diffracted to 2.5 Å. Although all calcium ions appear to have been removed from this particle, it shows no evidence of expansion. It appears the Asp55 side chains have been protonated and moved inwards, and the calcium binding site on the three-fold axis is occupied by a water molecule instead. The crystallisation solution for these crystals contains magnesium, which may have enabled the recontraction of the particle at pH 6.5. The RNA in these viruses is degraded, although the degradation appears to be limited to the production of a unique product. (Unge *et al.*, 1986; Montelius *et al.*,

1990). It is possible the intact RNA is required for the expansion to occur by charge repulsion of non-neutralised phosphates.

No crystal structure has been reported for type II crystals, but the fourth crystal form has been analysed. In these crystals, the calcium ion on the three-fold axis has been removed but otherwise these viruses seem unchanged from the untreated virus (Montelius *et al.*, 1990).

Like TCV, STNV retains icosahedral symmetry after expansion, but it is unknown whether this is also the case after proteolysis and it is also unclear whether proteolysis is required for genome release (Unge *et al.*, 1986).

More recently, STNV coat protein has been expressed recombinantly in *E. coli*, where it forms virus-like particles that can be easily purified. The VLPs are produced from a synthetic gene and therefore cannot package their genome but, interestingly, they appear to preferentially package the coat protein mRNA present in the *E.coli* cell. A crystal structure has been solved of the recombinantly expressed STNV particles and is almost indistinguishable from the native structure (Lane *et al.*, 2011). A low-resolution crystal structure was also determined for these VLPs and density was seen that was attributed to 3-basepair RNA helices and a single unpaired base (Lane *et al.*, 2011).

6.1.1 Assembly of STNV VLPs from coat protein

Recombinantly expressed STNV CP can be disassembled and reassembled into VLPs in the presence, but not the absence of RNA. Like the expanded particle, the disassembled coat protein is sensitive to protease, and the N-terminal helix of the protein is required for VLP reassembly (Bunka *et al.*, 2011).

A short RNA fragment (aptamer), named B3, that tightly binds the STNV coat protein and was identified by SELEX, is also able to trigger reassembly. It matches 16 out of 25 bases between bases 57 to 81 of the STNV-1 genome, including a ten-base continuous stretch (Figure 6.2). This section is part of the coat protein gene and is

therefore also present on the coat protein mRNA that is packaged by recombinant coat protein. The predicted secondary structure of the B3 aptamer consists of a four-base double stranded RNA stem with a four-base loop. The high affinity of B3 for STNV CP strongly suggests a specific RNA-coat protein interaction is linked to capsid assembly, although the matching sequence within the STNV genome does not appear to be the sole initiation site. Rather, it has been suggested that multiple, similar but not identical, stem-loops on the genome contribute to the assembly and stability of the particle (Bunka *et al.*, 2011). The consensus binding motif identified is AxxA and 30 predicted stem-loops displaying this pattern are present in the STNV genome. Cooperativity between coat protein-RNA binding and protein-protein interactions has been suggested to be required for the packaging specificity of STNV for its own genome or mRNA compared to other RNA (Bunka *et al.*, 2011).

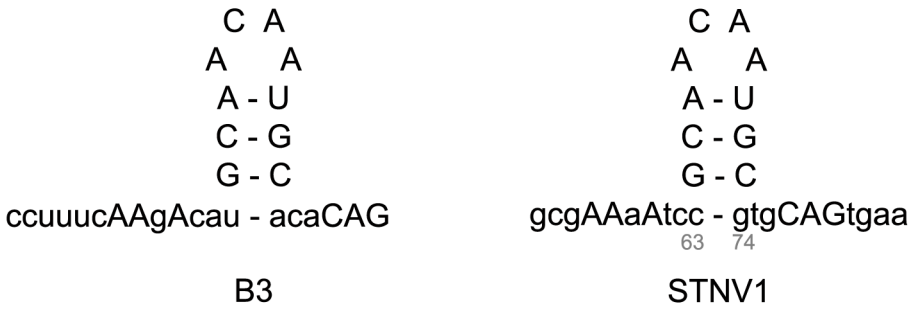


Figure 6.2 Predicted structure of the B3 aptamer
The B3 aptamer is shown on the left with the corresponding region on the STNV genome on the right. Identical residues are shown in capital letters, differing bases in lowercase.

In order to determine whether B3 makes similar interactions with the STNV coat protein as the coat protein mRNA, we initiated crystallographic studies of STNV VLPs reassembled with B3.

6.2 Crystallographic studies of STNV-B3

6.2.1 Experimental setup

Purified reassembled STNV-B3 was provided by Robert Ford. Purified B3-containing VLPs were concentrated and crystals were obtained in the same conditions

as the published structure of recombinant STNV (Lane *et al.*, 2011). Cryo-protection was achieved as for the low-resolution structure of recombinant STNV in a three-step protocol with mother liquor containing increasing glycerol concentrations (Lane *et al.*, 2011). The best crystal diffracted to a resolution of 2.3 Å. Three datasets were collected from this crystal, using different detector distances and beam-stop positions to obtain the largest possible resolution range (see Table 6.1 and Figure 6.3).

Table 6.1 Data collection and processing statistics

Figures in brackets are for the highest resolution shell

Space group	C2
Unit cell (Å)	a=315.0, b=300.6, c=183.5, β=94.4°
Wavelength (Å)	0.9795
Resolution range	217.15-2.29 (2.35-2.29)
Number of unique reflections	756410 (55678)
Mean <I>/σ<I>	9.3 (3.3)
R_{merge}	0.156 (0.655)
Completeness (%)	99.6 (99.2)
Redundancy	5.5 (3.8)

The three datasets were integrated using the automated data reduction system Xia2 (Winter, 2010). Because of its iterative approach, Xia2 is able to find the optimal unit cell parameters to integrate the three datasets together. The integrated, combined dataset has a resolution range from 217 to 2.3 A.

As the unit cell parameters for this crystal are nearly identical to those of the previously published structure of STNV, initial phases were obtained by a rigid-body refinement using the published coordinates in Refmac5. All waters were subsequently removed and a map was calculated. Several rounds of real-space refinement in Coot

followed by maximum-likelihood refinement in Refmac5 were then performed to improve the fit of side chains and waters in the electron density. Strict 60-fold non-crystallographic symmetry restraints were used throughout. The occupancy of the calcium atoms and waters on the five-fold symmetry axes was set to 0.2 to allow for their special position and prevent the refinement pushing them apart.

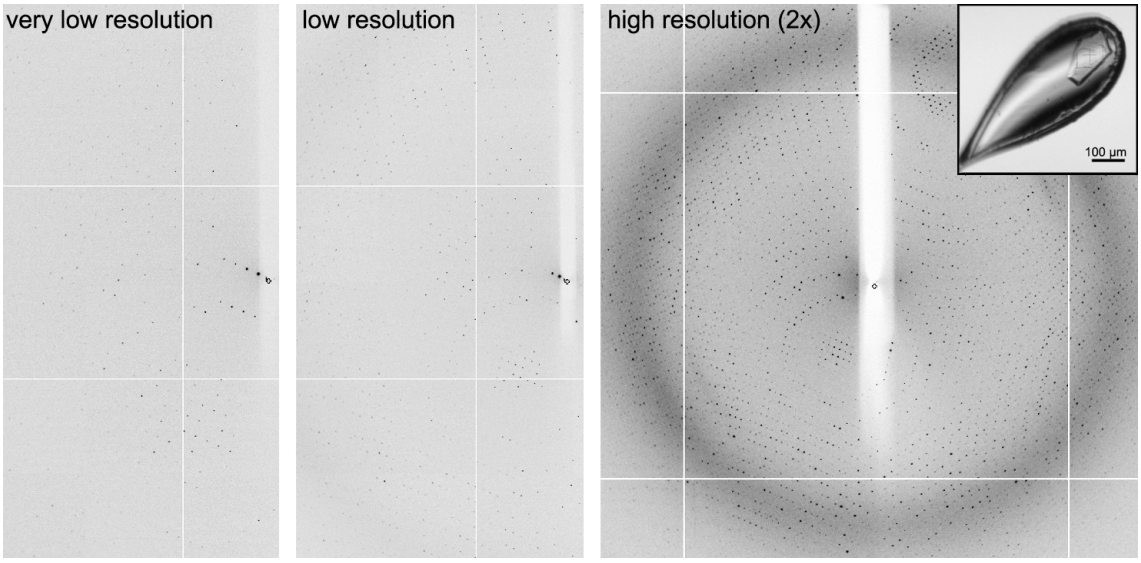


Figure 6.3 Diffraction patterns of STNV-B3
From left to right diffraction patterns from the very low resolution (217-30 Å), low resolution (116-3.8 Å) and high-resolution (30-2.29 Å) datasets are shown. The high-resolution image is shown magnified compared to the other two images for clarity. The inset shows the crystal these datasets were collected from.

6.2.2 Protein structure

The structure of the coat protein is very similar to that of previously published structures. The structure presented here has a backbone r.m.s.d. of 0.138 Å compared to PDB entry 3S4G and 0.271 Å compared to PDB entry 2BUK (Lane *et al.,* 2011; Jones and Liljas, 1984). The most variable region is the N-terminal helix of the coat protein, which has previously been described as the most flexible (Lane *et al.*, 2011).

There is density for the calcium ions in the five-fold positions and the coat protein positions, but no density is present for the three-fold calcium binding sites (Figure 6.5). The three-fold site has previously been identified as the one with lowest calcium affinity

and is the first site to be depleted when the virus is exposed to EDTA (Montelius *et al*, 1990).

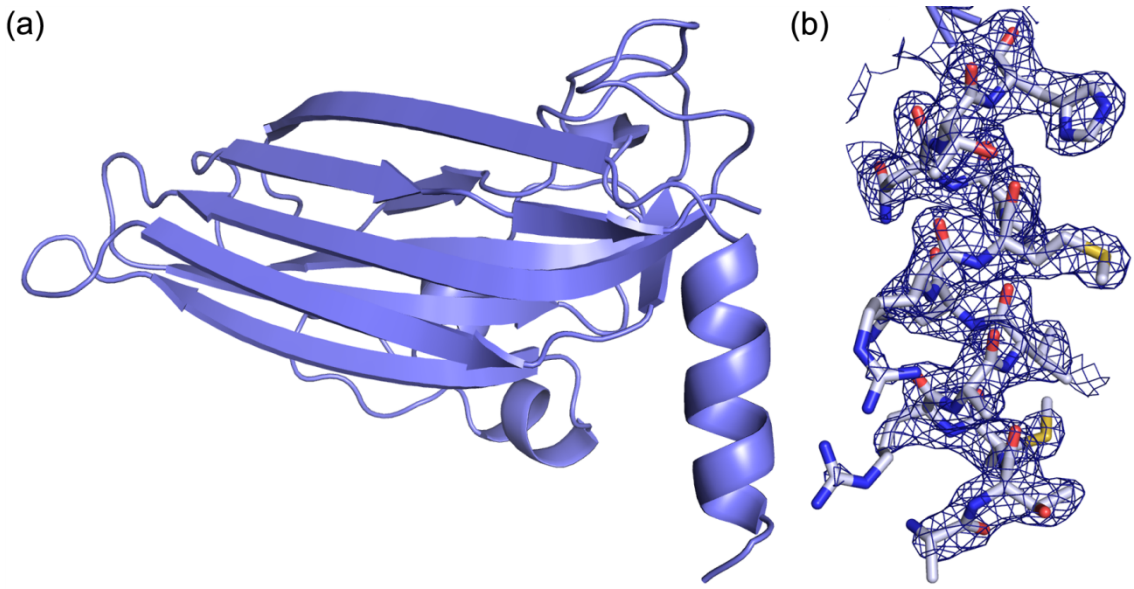


Figure 6.4 STNV coat protein structure
(a) Showing the STNV coat protein in cartoon representation. It is very similar to previous structures, c.f. figure 1.1. (b) Close-up view of the N-terminal helix of the protein with density. The backbone is well defined, but there is no clear density for the side-chains of Arg14 and Arg18, indicating these are flexible. The map maps shown is a $2F_o$-F_c map, 30-fold averaged in Coot and shown at a contour level of 1σ.

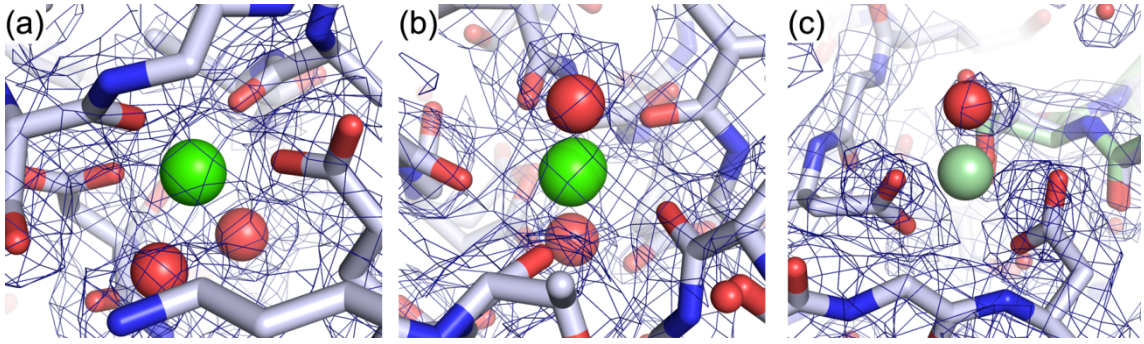


Figure 6.5 Calcium binding sites in STNV
(a) Octahedral calcium binding site within the coat protein. The density clearly shows the presence of the calcium. The structure is coloured by atom type, with C in pale blue, N in blue, O in red and calcium coloured green. (b) Calcium binding site on the five-fold axis. The two bound water molecules complete this pentagonal bipyramidal binding geometry. Colouring as in (a). (c) The calcium binding site on the three-fold symmetry axis. The calcium is in the position is occupies in the previous structure 2BUK and coloured in a lighter green. The model has been refined without this calcium atom and the absence of density indicates the calcium is not bound. All maps shown are 2Fo-Fc maps, 30-fold averaged in Coot and shown at a contour level of 1σ.

6.2.3 Internal structure

The aim of this experiment was to resolve the binding of the B3 aptamer in the STNV virus-like particles by collecting data over a wide resolution range. Features that do not closely follow the icosahedral symmetry, such as the genomic RNA, are averaged out in high-resolution maps, but can be seen when looking at maps in lower resolution.

Table 6.2 Refinement statistics

Values in parentheses are for the highest resolution shell.

Data range (Å)	217.15-2.29 (2.35-2.29)
Number of reflections	718474 (52951)
R (%)	16.3 (22.6)
R_{free} (%)	19.0 (26.0)
Total number of atoms	100140
Protein atoms	86580
Solvent atoms	13380
r.m.s.d. bond lengths (Å)	0.023
r.m.s.d. angles (°)	1.683
Average B factor (Å^2)	16.85
protein	15.28
ions	12.88
solvent	27.11

In Figure 6.6, a feature in the electron density observed at the five-fold symmetry axes is shown with different resolution cut-offs. The 30-fold averaged, $2F_o$-F_c map calculated with all data included (Figure 6.6a) does not show the additional density at

the five-fold axis, while this feature appears in maps calculated with data beyond a certain value excluded (Figure 6.6b-d).

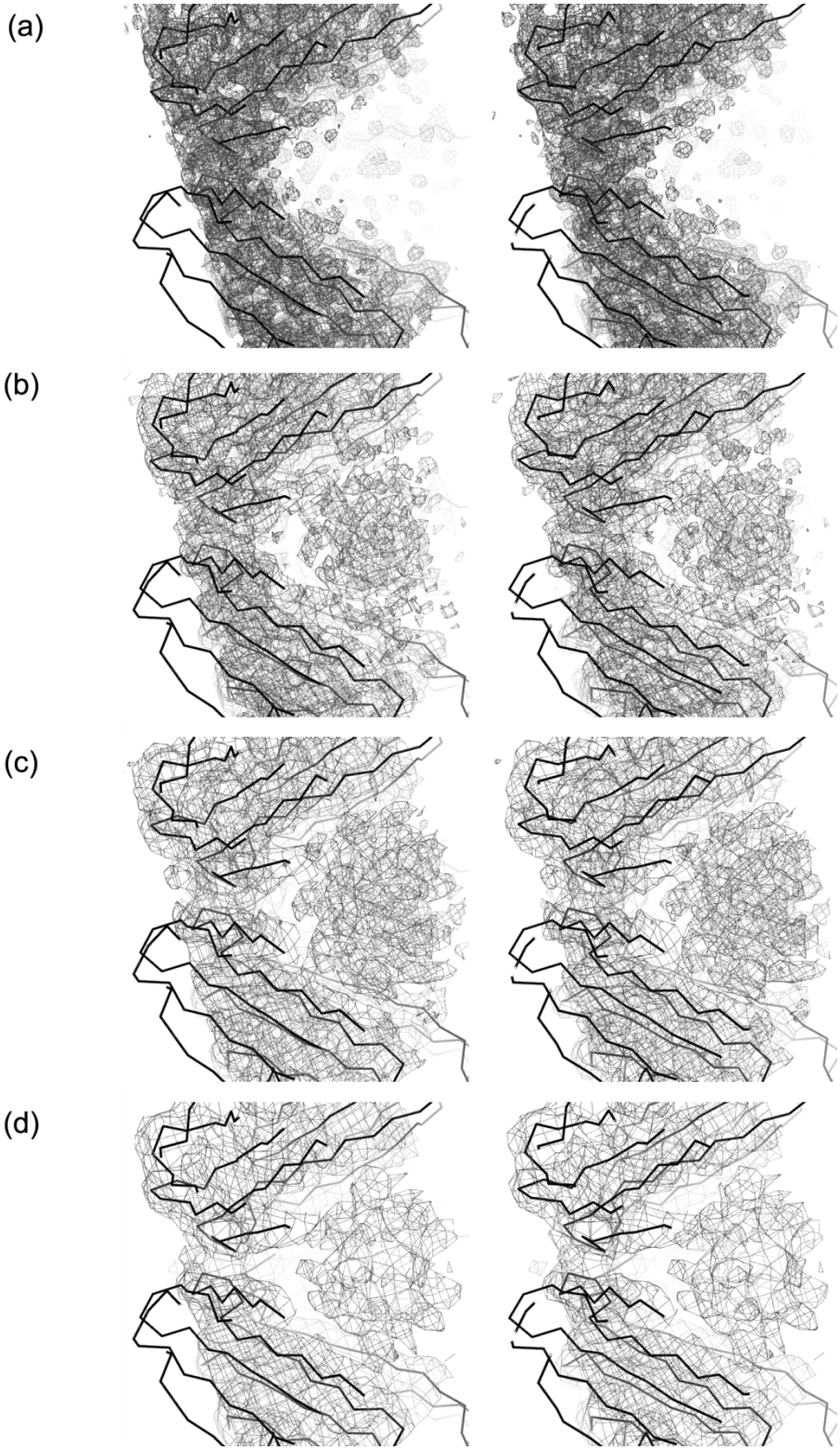


Figure 6.6 Internal density in STNV-B3 at different resolution
Stereo views are shown for: (a) $2F_o$-F_c map including all data, resolution range 217 - 2.3 Å. (b) $2F_o$-F_c map resolution cut to 4 Å (c) $2F_o$-F_c map resolution cut to 5 Å (d) $2F_o$-F_c map resolution

cut to 6 Å. All maps are 30-fold averaged in Coot (Emsley and Cowtan, 2004) and contoured at 1σ.

The density observed in the averaged maps with resolution cut-off of 6.0 Å consists of a globular feature on the five-fold icosahedral symmetry axes with cylinders of density running perpendicular to the two-fold symmetry axes (Figure 6.7). This arrangement corresponds closely with density distributions obtained from neutron scattering experiments of wild-type STNV (Bentley *et al.*, 1987) and a low-resolution crystal structure of wild-type STNV (Montelius *et al.*, 1988). However, it does not correspond to the more recent low-resolution crystal structure of recombinant STNV, which shows density for bound RNA, interpreted as a 3 base-pair double helix with a single unpaired base (Lane *et al.*, 2011).

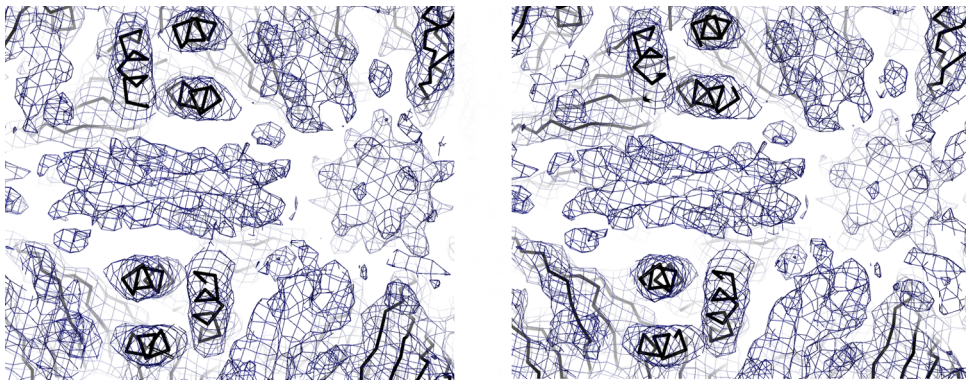


Figure 6.7 Internal density of STNV-B3
The $2F_o$-F_c map with resolution cut-off of 6.0 Å was icosahedrally averaged in Coot and is shown in stereo view at a contour level of 1σ.

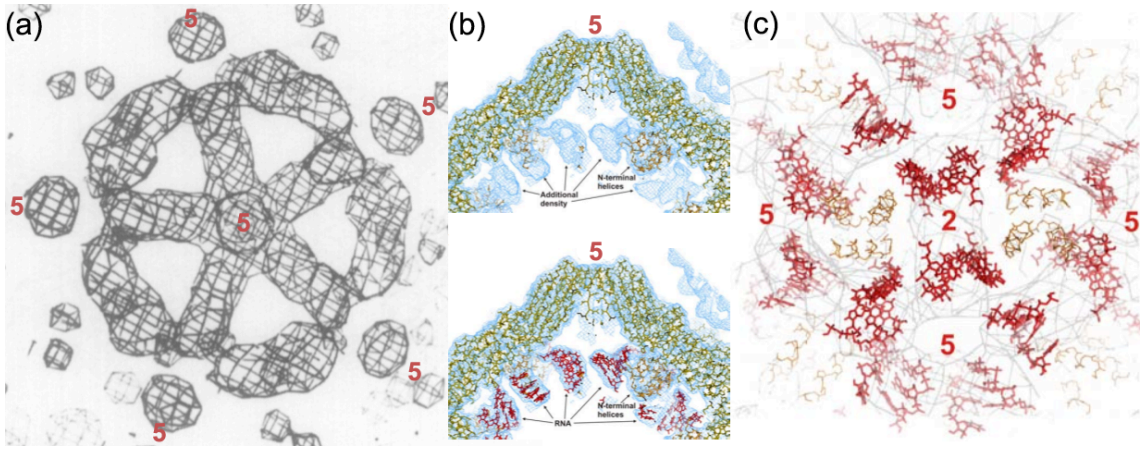


Figure 6.8 Previously published RNA density for STNV
(a) The density attributed to RNA in wild-type STNV by neutron scattering experiments with contrast matching. The map for 38% D_2O is shown. Figure taken from Bentley *et al.*, 1987. (b) Density for RNA in the low-resolution structure of recombinant STNV containing the STNV coat

protein mRNA. Figure taken from Lane *et al.*, 2011. (c) The RNA model from the recombinant STNV structure superposed on the RNA structure from neutron scattering. Figure taken from Lane *et al.*, 2011.

In line with earlier experiments on bacteriophage MS2 with small RNA fragments, and based on the high binding affinity of the STNV coat protein for B3, it was expected the aptamer would bind to the coat protein in a specific manner, resulting in icosahedrally ordered density, allowing elucidation of the aptamer structure and specific protein-RNA interactions. It appears, however, that although it is clear the RNA is bound at the CP dimer interface, the aptamer is not consistently bound in the same orientation. This means that when averaging is applied, no high-resolution electron density is retained for the aptamer, but low-resolution density can be seen. If 30 copies of the 30-base B3 aptamer were bound, the capsid would contain 900 nucleotides. This is sufficient to explain the observed internal density in the B3-containing VLPs. This binding at the dimer interface is consistent with the coat protein existing as dimers in solution before assembly.

It is also possible that the averaging procedures used in this work were not optimal, and better results may be obtained by using different procedures. As averaging in Coot does not allow for the selection of a larger mask that encloses an area within the capsid, it is possible that density in that region is removed by the averaging. However, averaging using a model in which the additional region has been marked using dummy atoms does not lead to significant differences in the averaged $2F_o$-F_c map. Similarly, an icosahedrally averaged F_o-F_c (difference) map, contoured to show positive density only and with a resolution cut-off of 6 Å, indicates there is RNA bound, but again, no unique aptamer orientation could be fitted into the density.

6.2.4 A new crystal form of STNV

A different crystal form for STNV-B3 was found to grow from the same crystallisation conditions. Rather than flat, diamond-shaped plates, the crystals were smaller and cube-shaped. Although diffraction of these crystals was promising (> 2.5 Å), one of the

cell edges was >600 Å, which meant data could not be collected without overlapping spots because of the preferential orientation of the crystals in the loops.

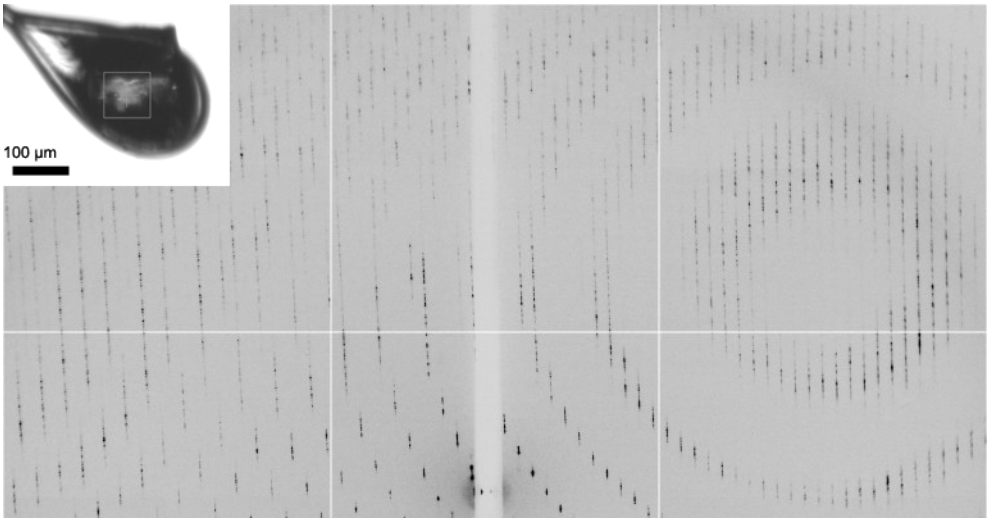


Figure 6.9 Diffraction pattern for new crystal form of STNV-B3
A diffraction pattern for a new crystal from of STNV-B3, that grew from the same conditions as the other crystals. Although the diffraction is clean, the spots cannot be separated without using a kappa-goniometer, which at time of writing was not available.

6.3 Discussion

The protein structure of STNV reassembled with the B3 aptamer is very similar to that of previously published structures (2STV/2BUK, Jones and Liljas, 1984; 3S4G, Lane *et al.*, 2011). Despite the extensive resolution range of the data recorded, no density interpretable as a unique binding conformation of the aptamer was present inside the capsid, although it is clear that something is bound. It is possible that the aptamer binding is insufficiently specific to restrain the RNA to a single conformation. There is obviously sufficient binding to initiate assembly, as the VLPs form, but the RNA density is averaged by the crystal packing as in the native structure. The density that is observed indicates regions where the RNA is more likely to bind, while other regions that show less density are less favourable for RNA binding. The pattern corresponds with the distribution of positive charges on the interior surface of the capsid, suggesting the affinity of the aptamer may be partly based on charge.

The aptamers identified for the STNV coat protein and the 30 predicted stem-loops in the STNV genome show a wide variety of structures. The AxxA motif is common

among the aptamers, although it occurs in 4, 5 or 6-base loops. The stems vary in length from a 2 base-pair stem to a 5 base-pair stem, commonly with bulges of one or two bases. The stem-loops in the STNV1 genome were found based on the motifs implied by the aptamers and therefore have a similar variety. In the stems of the genomic stem-loops there is non-canonical base pairing as well as Watson-Crick base pairing. This implies the capsid binding site must be flexible enough to accommodate all these structures. As the B3 aptamer is relatively small with a 4 base-pair stem and a 4-base loop, it can probably bind in several orientations, resulting in the apparent disorder observed. It may be possible to obtain better results with VLPs formed in the presence of one of the larger stem-loops. Non-specific binding of the unstructured regions of the aptamer that flank the stem-loop may also contribute to the observed disordered binding.

Interestingly, although the size of the STNV capsid limits the maximum size of the STNV genome, forcing it to rely on co-infection with TNV for replication, the genome contains ~600 bases that are untranslated and have no currently known function. At 1239 nucleotides, the STNV genome is almost twice the length of the CP mRNA, although the CP mRNA is sufficient for encoding the protein and initiating assembly. As such, it is unclear what additional functions this additional genomic RNA performs during infection. A possible role is in viral expansion and uncoating.

When recombinant STNV containing only CP mRNA are exposed to EDTA, they dissociate instead of expanding (Robert Ford, personal communication), suggesting the additional length of the genomic RNA may be required to hold the virus particle together in the expansion conditions, which mimic the conditions in the plant cytosol. This putative role in capsid expansion is supported by the observation that, from crystallographic studies, EDTA treated STNV particles containing a shorter degraded RNA are not expanded, whereas particles containing full-length RNA, in identical crystallisation conditions, are expanded by about 4.5% (Montelius *et al.*, 1988; Montelius *et al.*, 1990). This implies the almost two-fold increase in charge repulsion

provided by the genomic RNA compared to shorter degraded RNA or CP mRNA may be required for expansion under mild conditions.

7 Conclusions

The assembly and uncoating processes of unenveloped, single-stranded RNA viruses has long been a mystery. In contrast, tailed double-stranded DNA bacteriophages have been much more extensively studied.

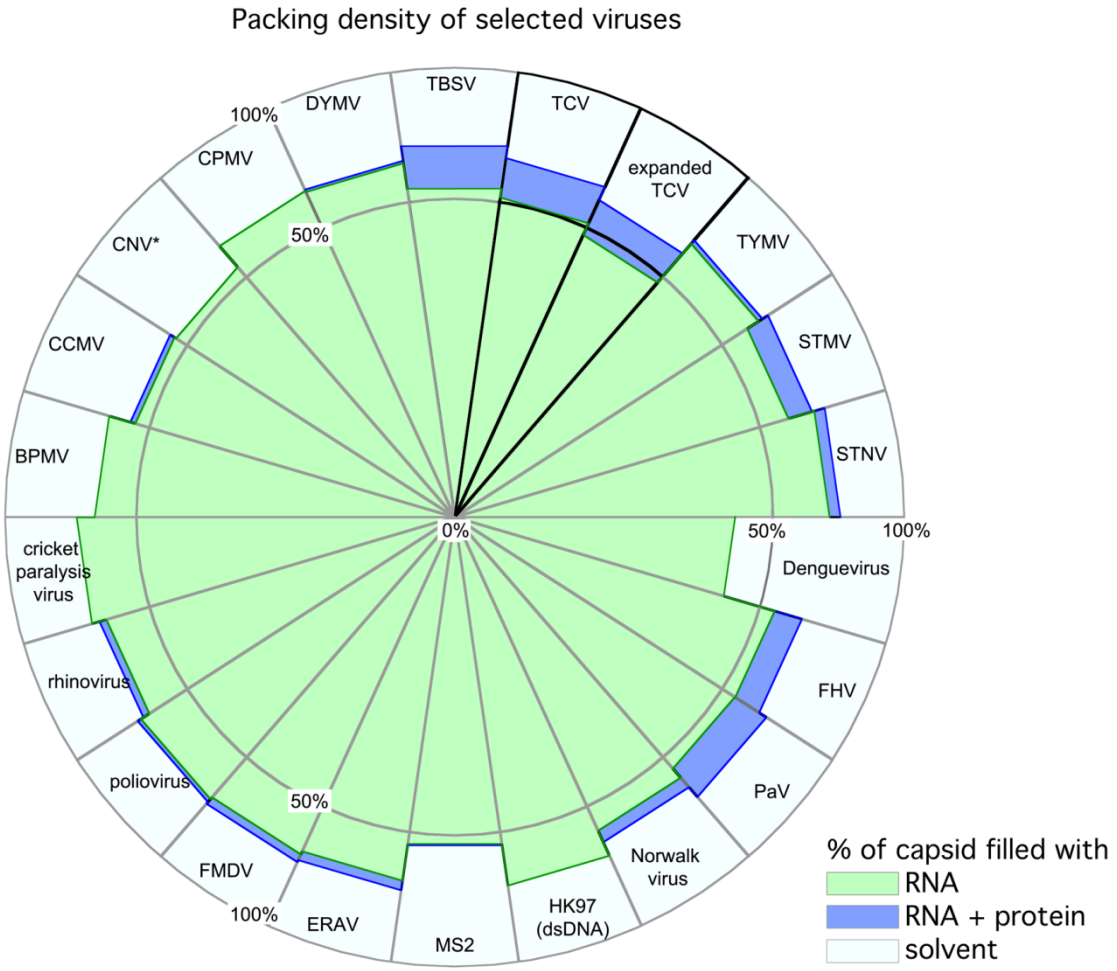


Figure 7.1 Packing density of viruses
Showing the percentage of the capsid volume occupied by RNA in green, protein (not observed in crystal structure) in blue, and solvent in pale blue. The percentage occupied by various components has been represented as the percentage of the area in the graph, rather than the percentage of radius. *For CNV no crystal structure is available, so no value for the protein component could be calculated.

The DNA phages assemble as empty pro-heads, that are filled with the DNA genome by a packaging motor. In this way, considerable force can be exerted during

packing, resulting in a pressurised capsid. Upon receptor binding, the 'plug' holding the genome is released and the pressure causes the DNA to be injected into the host cell (Ivanovska *et al.*, 2007). The pressure in dsDNA phages is not caused by compression of the nucleotides, but results from charge repulsion and the greater stiffness of the double-stranded DNA compared to single-stranded RNA (Evilevitch *et al.*, 2008). In fact, the packing density of 'pressurised' dsDNA phages is similar to the packing density of ssRNA viruses (Figure 7.1).

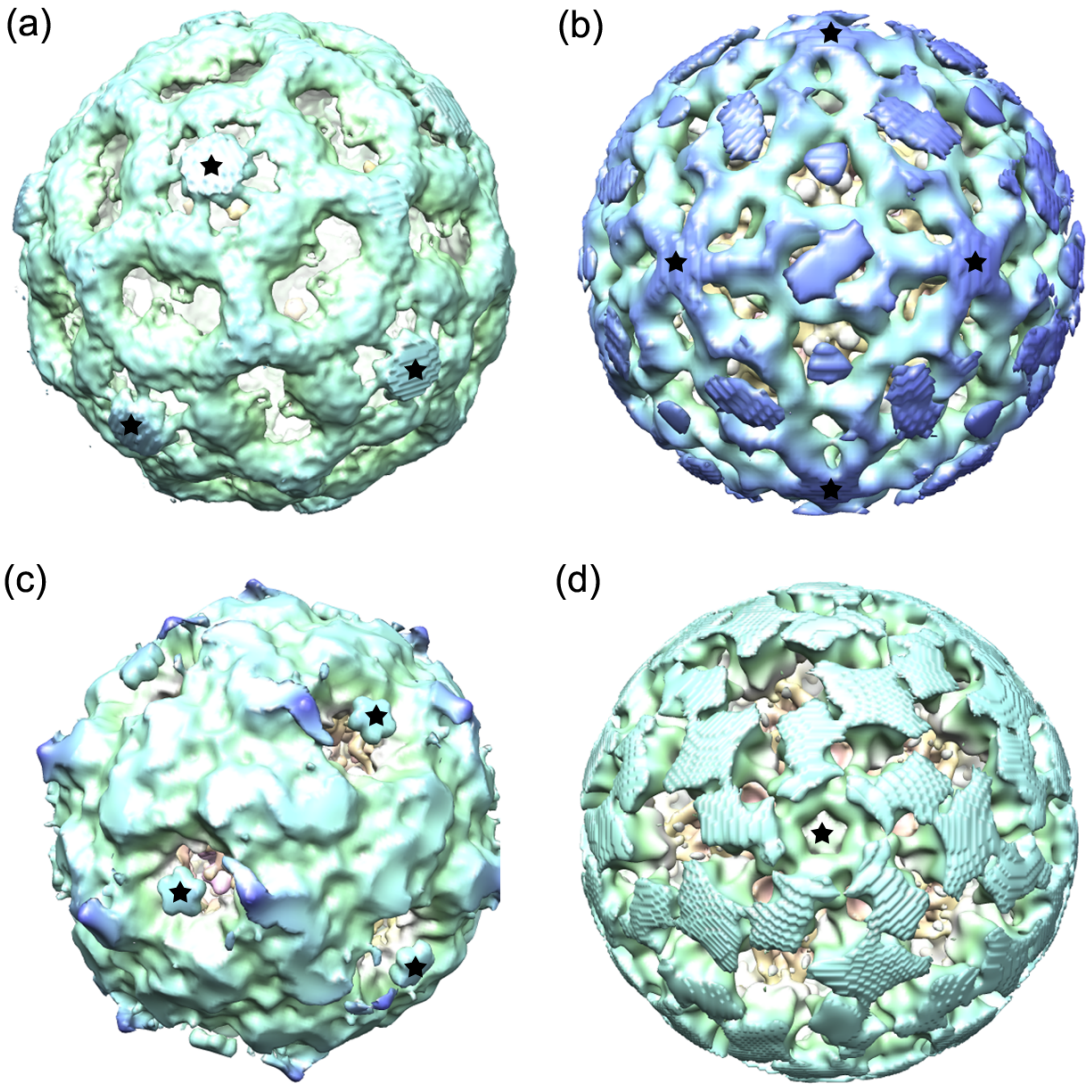


Figure 7.2 Examples of different RNA genome structures
(a) RNA structure of TCV as described in Chapter 4 of this work. (b) The RNA structure of CNV as described in Katpally *et al.*, 2007. Thanks to Dr Tom Smith, (Danforth Plant Science Center, St. Louis, MO, USA) for access to the map. (c) RNA structure of ERAV as described in Chapter 5 of this work. (d) RNA structure of MS2 as described in Toropova *et al.*, 2008. Map from the EMDB accession number 1431. For all maps, the density corresponding to the coat protein was removed using the Volume Eraser functionality in Chimera. Maps indicate the structure of the RNA only and are not shown to scale. Five-fold symmetry axes are indicated with a star.

Many of the viruses studied in this work and previously show an inner structure composed of multiple shells of density. TBSV, TCV, CNV and MS2 all have a multi-shell inner structure with regions of lower density separating the shells (Toropova *et al.*, 2008; Aramayo *et al.*, 2005; Katpally *et al.*, 2007). A longer genome seems to result in more shells in these viruses, while shorter genomes such as that of TCV compared to TBSV, like the subgenomic fragments of the MS2 genome, preferentially occupy the outer shells (Aramayo *et al.*, 2005; Rolfsson *et al.*, 2010). A notable exception to this pattern are the picornaviruses ERAV and poliovirus, which have a similar capsid size but genomes up to twice as long as TCV, and do not exhibit an inner structure with multiple well-defined shells (Bubeck *et al.,* 2005a). When looking at the shells in detail, a single structure or pattern is not apparent. Although the outer shell (or in the case of ERAV the outer parts of the RNA structure) in all ssRNA viruses studied to date has been featured, i.e. has exhibited a distinct pattern of lower and higher density following the icosahedral symmetry imposed during 3D reconstruction, the features differ even for closely related viruses such as TCV, TBSV and CNV, all *Tombusviridae*. These patterns reflect the distribution of positive and negative charges on the inner surface of the capsid, as well as the presence of more specific binding sites on the coat protein. An extremely striking example of highly ordered RNA structure in a virion is seen in Pariacoto virus, which shows a dodecahedral cage of double-stranded RNA positioned just inside the capsid (Tsuruta *et al.*, 1998). A dodecahedral cage is also seen in the outer layer of the ERAV genome (Figure 5.5), although for this virus it is less obvious what secondary structure elements make up the cage structure. The TCV RNA, on the other hand, shows a pattern of density (Figure 4.5) that is almost a truncated icosahedron, interacting with the capsid at the three-fold symmetry axes, while the CNV RNA interacts with the capsid predominantly at the five-fold symmetry axes (Katpally *et al.*, 2007). The different viral RNA structures found in this work strengthen the view that there probably is not one single genome structure for RNA viruses. Instead the differences suggest that the evolutionary pressure on the genome to

encode functional proteins and recognition sites for the coat protein, as well as form a structure that can be packaged within the capsid, has resulted in different solutions for the different viruses studied to date.

Additionally, it is important for viruses to release their genome in a controlled way, as the genome may trigger host anti-viral responses if detected by the cell. DNA bacteriophages can inject their dsDNA genome directly into the cytoplasm as bacteria have no nucleus, although the viral DNA may still be degraded by restriction enzymes. If DNA or dsRNA is detected in the cytoplasm of eukaryotic hosts, however, it elicits responses such as RNA silencing or components of the innate immune system. Therefore, DNA viruses that infect eukaryotes only expose their genome in the nucleus, and also package progeny genomes there. For RNA viruses, it is important to keep the secondary structure of the genome hidden from the host cell.

In Chapter 4 of this work, a mechanism has been suggested for the uncoating of TCV, proposing a model of how this virus can uncoat its genome while avoiding exposure to the host cell. A similar mechanism is likely to be used by other icosahedral plant viruses. TCV and STNV, and many other plant viruses, such as CCMV and SBMV (Speir *et al.*, 1995, Abad-Zapatero *et al.*, 1980), contain divalent metal ions, usually calcium, within the capsid. These can be removed from the completed particle by adding EDTA, resulting in charge repulsion, which causes the formation of pores in the viral capsid, often accompanied by expansion of the particle. In the case of red clover necrotic mottle virus (RCNMV), however, pores are formed after removal of calcium ions without a notable particle expansion (Sherman *et al.*, 2006). The addition of EDTA *in vitro* mimics the low-calcium environment in the plant cell, as intracellular calcium is tightly regulated. Plant cells, like animal muscle, use calcium as an intracellular messenger. The free calcium concentration in the plant cytosol is thought to be as low as 10^{-7} M, compared to 10^{-3} M for magnesium (Laliberté and Sanfaçon, 2010; Evans, 1988; Montelius *et al.*, 1990). It has long been thought that calcium release is the first step in plant virus uncoating (Durham *et al.*, 1977). The removal of

calcium causes the virus to expand and become sensitive to the activity of proteases in the plant cytosol or *in vitro*. Proteolysis subsequently causes the release of a subset of coat protein subunits, and in the case of TCV allows a protrusion to be formed that breaks the icosahedral symmetry of the particle as described in Chapter 4. The formation of polysome-like virus-ribosome complexes for TCV and other viruses (Roenhorst *et al.*, 1989) suggests that ribosomes may be instrumental in the uncoating of the genome from the capsid, like the Trojan horse was unwittingly pulled into Troy. It is at this time unknown whether other virus capsids display protrusions after proteolysis, but many plant viruses have protease-sensitive N-terminal regions, including STNV, CCMV and SBMV (Chapter 4, Chapter 6, Lane *et al.*, 2011; Chidlow and Tremaine, 1971; Tremaine and Ronald, 1978), suggesting they may employ a similar mechanism. The result of the proteolysis of most plant viruses appears to be a single or limited product, indicating the protein remains folded and other protease-sensitive residues are not exposed in the conditions tested. The results presented in Chapter 4 indicate that the formation of expanded states and associated changes in protease sensitivity seen in TCV, STNV and other viruses is likely to be essential for the uncoating process of these viruses.

Similarly to these plant viruses, the uncoating intermediates of some *Picornaviridae*, called A particles, which are produced after receptor binding, have an altered surface and increased susceptibility to proteases (Hogle, 2002; Lin *et al.*, 2011). These particles have externalised the VP4 protein as well as the N-terminus of the VP1 coat protein. After externalisation, the N-terminus of VP1 is localised on the propeller tip and is sensitive to proteases (Bubeck *et al.*, 2005a). After genome release, the N-terminal portion of VP1 can be found in two positions: on the propeller tip and near the two-fold symmetry axis, and the particles no longer have icosahedral symmetry (Lin *et al.*, 2011). The two-fold symmetry axis, where the N-terminal portion of VP1 is located on some of the empty poliovirus particles, is also where the genome exits the capsid (Bostina *et al.*, 2011). This is consistent with the formation of a pore, consisting of the

N-terminal region of VP1, together with the VP4 protein, at or near the two-fold symmetry axis of the poliovirus particle.

Although no A particles have been found for the aphthoviruses ERAV or FMDV, it is clear they must uncoat their genome in a controlled way, as the genome is not exposed to the contents of the endosome (Groppelli *et al.*, 2011). A previously published low-pH ERAV structure (Tuthill *et al.*, 2009) and the empty particle presented in Chapter 5 of this work provide an insight into the formation of uncoating intermediates in ERAV, which was previously assumed to dissociate into pentamers upon lowering acidification of the endosome after endocytosis (Tuthill *et al.*, 2010). The results for the ERAV empty particle, discussed in Chapter 5, suggest that in ERAV, like poliovirus, the two-fold symmetry contact, i.e. the interface between two pentamers, is the weakest interaction in the virus particle. This is therefore the most likely interface to allow for pore formation during uncoating, unlike the interactions between protomers at the five-fold axis, which are formed early during assembly and do not appear to change during further assembly or maturation.

Although the genome uncoating processes of TCV and ERAV have been the focus of this work, Chapter 6 looks at the assembly process and genome packaging. The role of the RNA during ssRNA virus assembly has long been underestimated, because the genome structure was not usually observed in the icosahedrally averaged crystal structures. Additionally, some plant viruses with segmented genomes form empty particles *in vivo*, suggesting the coat protein was more important. It has been shown, however, that the genome is essential for the assembly of many ssRNA viruses with non-segmented genomes. For MS2, STNV and TCV it has been shown assembly cannot take place *in vitro* in the absence of RNA, although assembly can take place with RNA other than the genomic RNA, such as aptamers, parts of the genomic RNA, messenger RNA, or the 18S ribosomal RNA (Rolfsson *et al.*, 2010; Qu and Morris, 1997; Sorger *et al.*, 1986; Bunka *et al.*, 2011). Despite this apparent promiscuity, viruses usually achieve selective packaging of their own genome in the background of

cellular RNA, and genomic RNA outcompetes heterologous RNA in *in vitro* reassembly experiments (Sorger *et al.*, 1986).

STNV is a small virus with an equivalent, *T*=1 capsid, and therefore the RNA doesn't have to perform the conformation switching of quasi-equivalent subunits, as it does in MS2 and TBSV (Harrison, 1980; Stockley *et al.*, 2007), Despite that, STNV still requires RNA for assembly and RNA was observed in the structure of recombinant STNV, which packages its coat protein mRNA. Although RNA binding was seen in the structure for STNV VLPs reassembled with the B3 aptamer, no specific interactions could be resolved. This means it remains unclear how the RNA binds to the STNV coat protein and what interactions are required for assembly to occur.

As mentioned above, many plant viruses contain calcium ions in their capsids, which they acquire from the cell during assembly. As calcium is tightly regulated in healthy plant cells and used as a second messenger for cellular signalling, this requirement for calcium ions means profound changes to the cell must be caused by these viruses, including those viruses that have only a small genome. Cellular remodelling takes place after infection (Laliberté and Sanfaçon, 2010) presumably disrupting calcium regulation and potentially creating compartments where all the components required for assembly are present in high concentrations. These would have the additional benefit to the virus of shutting out the host defence mechanisms and allowing genome multiplication to proceed undisturbed, as well as providing an environment where the majority of the RNA is the viral genome to promote selective packaging. The formation of vesicles that provide an anchor for genome replication and virus assembly is observed in cells infected with poliovirus or FMDV. Cellular remodelling, together with interference in normal host cell processes and activation of host anti-viral responses are probably the main causes of the symptoms of viral infection.

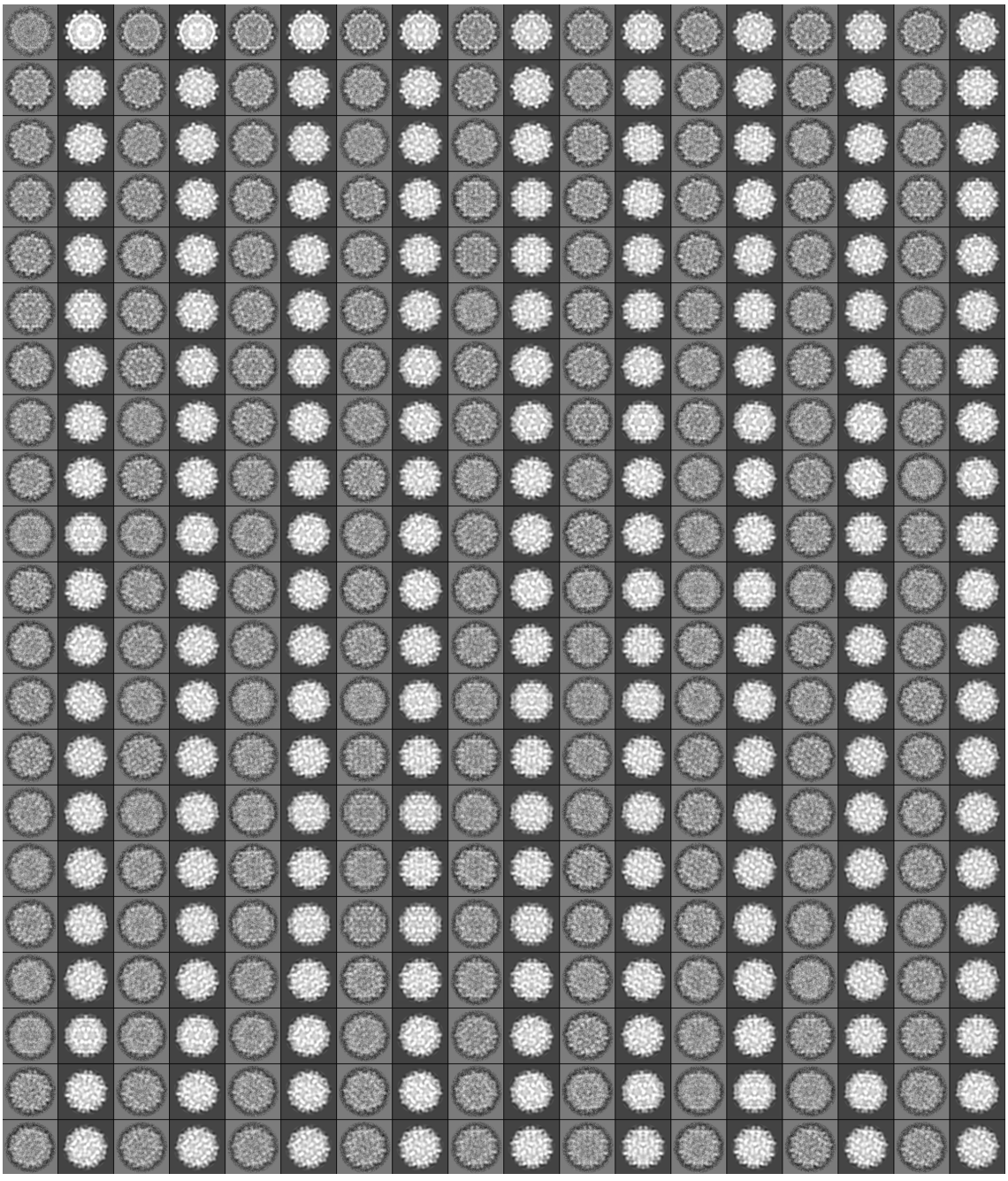
In this work, cryo-EM reconstructions have been described for native and expanded TCV, and it has been shown that expansion and subsequent proteolysis are required

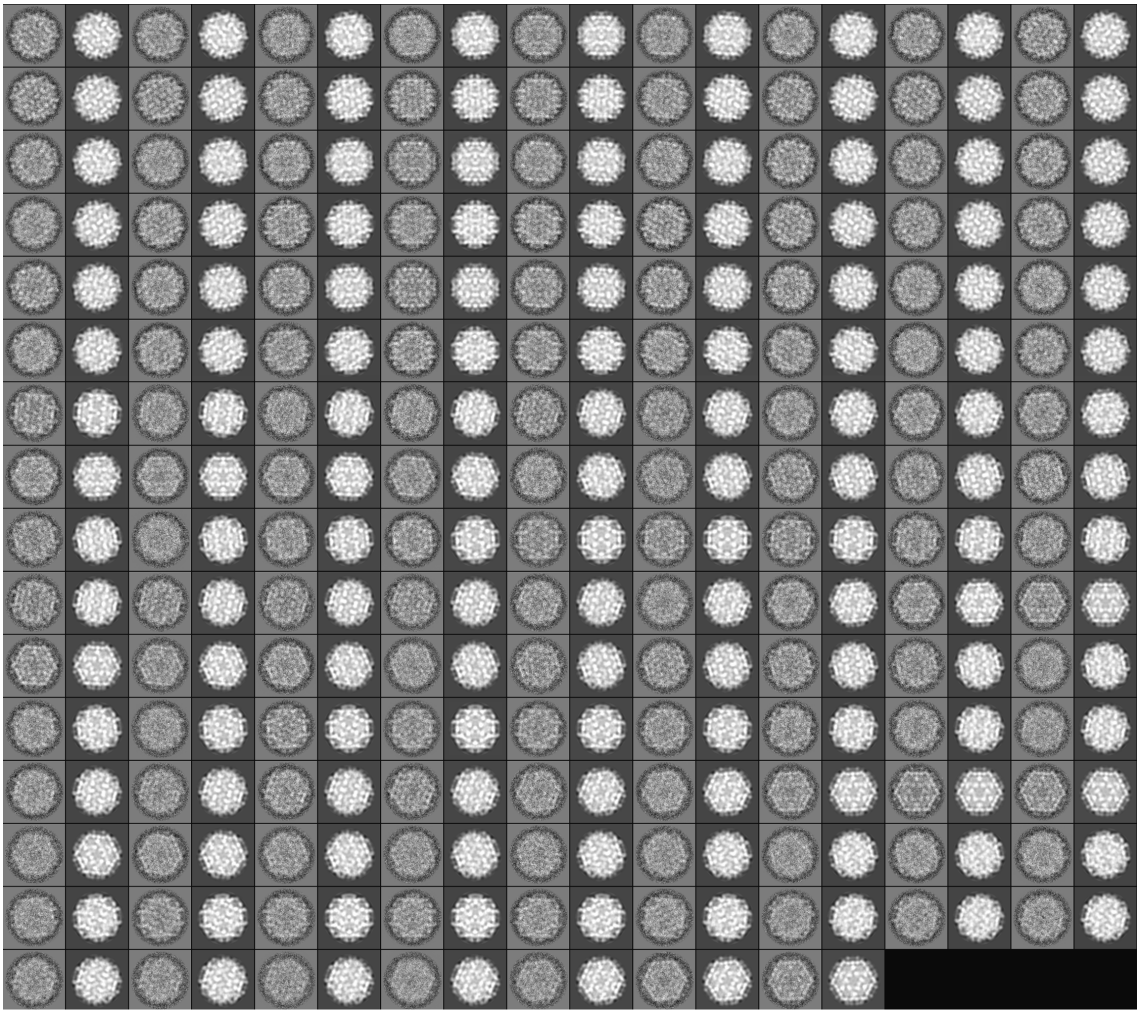
steps in the uncoating of the virus. Expansion is a common mechanism of uncoating for the two plant viruses discussed in this work, as well as several other plant viruses. A cryo-EM reconstruction of an expanded, empty form of ERAV has also been presented, which is dramatically different from that of the native virus. This empty particle, like the particle represented by low-pH crystal structure of ERAV, may be a disassembly intermediate in parallel with the poliovirus A particle, or post-disassembly particle in line with the 80S particle of poliovirus. There is a possibility that the low-pH particle of ERAV was expanded or showed a pore, but the crystallisation conditions caused it to recontract. A similar phenomenon has been shown in crystallisation of STNV, where expanded particles containing a degraded genome recontracted during crystallisation (Unge *et al.*, 1986).

In this work, models are proposed for the genome uncoating of TCV and ERAV, that may be applicable to other plant viruses and aphthoviruses, respectively, providing a hypothesis for how simple single-stranded RNA viruses uncoat their genomes without provoking the host cell to set off defence mechanisms such as RNA silencing.

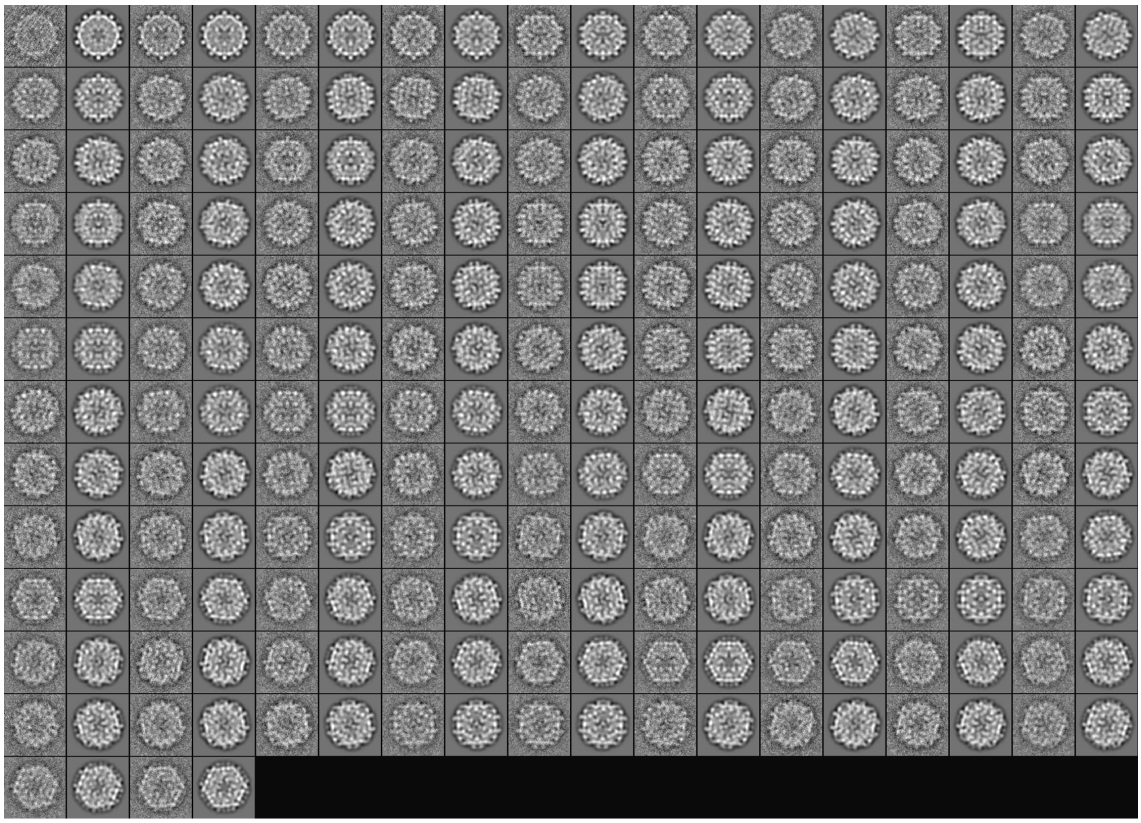
8 Appendices

Appendix A: TCV averages and reprojections (native)

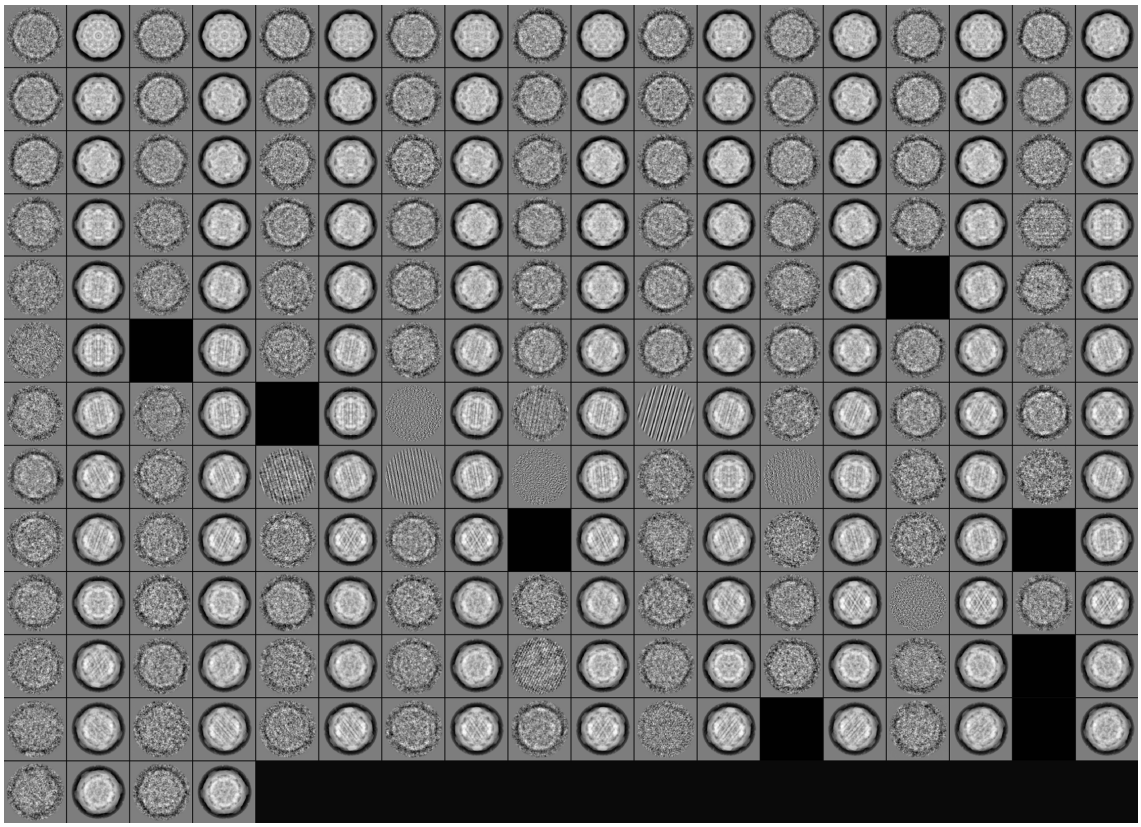




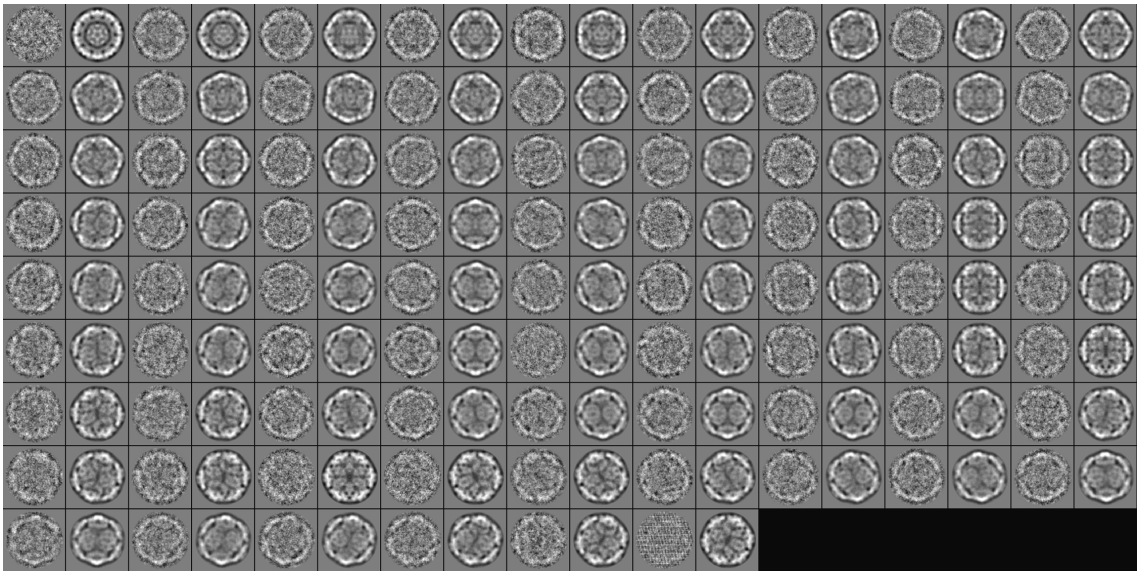
Appendix B: expanded TCV averages and reprojections



Appendix C: native ERAV averages and reprojections



Appendix D: expanded ERAV averages and reprojections



On accompanying CD:

Filename	Instructions
icos_centre.txt	Script for centering images. Change extension to your Spider procedure extension and run in Spider.
icos_oneiteration.txt	Script to run one iteration of 3D model refinement. Assumes presence of certain files and directories. Change extension to your Spider procedure extension and run in Spider.
EMfittingCNS.pdf	Explains procedure for the use of CNS to fit coordinates in EM maps.
TCV_native.spi TCV_expanded.spi	Cryo-EM maps in Spider format, to be viewed in Chimera. Pixel size 1.89 Å/pixel for expanded and 1.323 Å/pixel for native.
TCV_native_coords.pdb TCV_expanded_coords.pdb	PDB files of fitted coordinates for native and expanded TCV. View in Chimera and use Multiscale Models to obtain the complete virus particle.
ERAV_native.spi ERAV_empty.spi	Cryo-EM maps in Spider format, to be viewed in Chimera. Pixel size for both models is 1.72 Å/pixel.
ERAV_native_coords.pdb ERAV_empty_coords.pdb	PDB files of fitted coordinates for native and empty ERAV. Contains all subunits.
ERAV_transition.mpg	Movie of the ERAV transition mechanism from native to the empty particle.
SEBakker_thesis.pdf	This work in pdf format.

9 References

Abad-Zapatero, C., Abdel-Meguid, S., Johnson, J., Leslie, A., Rayment, I., Rossmann, M., Suck, D., Tsukihara, T. (1980). Structure of Southern Bean Mosaic-Virus at 2.8-Å Resolution. *Nature*, *286* (5768), 33-39.

Abrescia, N., Cockburn, J., Grimes, J., Sutton, G., Diprose, J., Butcher, S., Fuller, S., Martin, C., Burnett, R., Stuart, D., Bamford, D., Bamford, J. (2004). Insights Into Assembly From Structural Analysis of Bacteriophage PRD1. *Nature*, *432* (7013), 68-74.

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., & Brown, F. (1989). The 3-Dimensional Structure of Foot-and-Mouth-Disease Virus at 2.9-Å Resolution. *Nature*, *337* (6209), 709-716.

Adrian, M., Dubochet, J., Lepault, J., & McDowall, A. (1984). Cryo-Electron Microscopy of Viruses. *Nature*, *308* (5954), 32-36.

Adrian, M., Timmins, P., & Witz, J. (1992). *In Vitro* Decapsidation of Turnip Yellow Mosaic Virus Investigated by Cryo-Electron Microscopy: a Model for the Decapsidation of a Small Isometric Virus. *The Journal of General Virology*, *73 (Pt 8)*, 2079-83.

Aramayo, R., Merigoux, C., Larquet, E., Bron, P., Perez, J., Dumas, C., Vachette, P., Boisset, N. (2005). Divalent Ion-Dependent Swelling of Tomato Bushy Stunt Virus: a Multi-Approach Study. *Biochimica Et Biophysica Acta-General Subjects*, *1724* (3), 345-354.

Azevedo, J., Garcia, D., Pontier, D., Ohnesorge, S., Yu, A., Garcia, S., Braun, L., Begrdoll, M., Hakimi, M., Lagrange, T., Voinnet, O. (2010). Argonaute Quenching and Global Changes in Dicer Homeostasis Caused by a Pathogen-Encoded GW Repeat Protein. *Genes & Development*, *24* (9), 904-915.

Baker, M., Jiang, W., Rixon, F., & Chiu, W. (2005). Common Ancestry of Herpesviruses and Tailed DNA Bacteriophages. *Journal of Virology*, *79* (23), 14967-14970.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews, 35*(3), 235-241.

Bamford, D., Grimes, J., & Stuart, D. (2005). What Does Structure Tell Us About Virus Evolution? *Current Opinion in Structural Biology*, *15* (6), 655-63.

Ban, N., Larson, S., & McPherson, A. (1995). Structural Comparison of the Plant Satellite Viruses. *Virology*, *214* (2), 571-583.

Basnak, G., Morton, V., Rolfsson, O., Stonehouse, N., Ashcroft, A., & Stockley, P. (2010). Viral Genomic Single-Stranded RNA Directs the Pathway Toward a T=3 Capsid. *Journal of Molecular Biology*, *395* (5), 924-936.

Baulcombe, D. (2004). RNA Silencing in Plants. *Nature*, *431* (7006), 356-363.

Bentley, G., 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 H_2O/H_2O(H-1, H-2) Solvent Contrast Variation. *Journal of Molecular Biology*, *194* (1), 129-141.

Bernal, J., & Carlisle, C. (1948). Unit Cell Measurements of Wet and Dry Crystalline Turnip Yellow Mosaic Virus. *Nature*, *162* (4108), 139-140.

Bernal, J., Fankuchen, I., & Riley, D. (1938). Structure of the Crystals of Tomato Bushy Stunt Virus Preparations. *Nature*, *142*, 1075-1075.

Berryman, S., Clark, S., Monaghan, P., & Jackson, T. (2005). Early Events in Integrin $\alpha v\beta 6$-Mediated Cell Entry of Foot-and-Mouth Disease Virus. *Journal of Virology*, *79* (13), 8519-8534.

Bhella, D., Ralph, A., & Yeo, R. (2004). Conformational Flexibility in Recombinant Measles Virus Nucleocapsids Visualised by Cryo-Negative Stain Electron Microscopy and Real-Space Helical Reconstruction. *Journal of Molecular Biology*, *340* (2), 319-31.

Bostina, M., Levy, H., Filman, D., & Hogle, J. (2011). Poliovirus RNA is Released From the Capsid Near a Two-fold Symmetry Axis. *Journal of Virology*, *85* (2), 776-83.

Bottcher, B., & Crowther, R. (1996). Difference Imaging Reveals Ordered Regions of RNA in Turnip Yellow Mosaic Virus. *Structure*, *4* (4), 387-394.

Bottcher, B., Wynne, S., & Crowther, R. (1997). Determination of the Fold of the Core Protein of Hepatitis B Virus by Electron Cryomicroscopy. *Nature*, *386* (6620), 88-91.

Bovee, M., Lamphear, B., Rhoads, R., & Lloyd, R. (1998). Direct Cleavage of eIF4G by Poliovirus 2A Protease is Inefficient *in Vitro*. *Virology*, *245* (2), 241-249.

Bradley, D. (1954). Evaporated Carbon Films for Use in Electron Microscopy. *British Journal of Applied Physics*, *5* (Feb), 65-69.

Bragg, W. (1922). Nobel Lecture: The Diffraction of X-Rays by Crystals. *Nobel Lectures, Physics 1901-1921*. Elsevier Publishing Company, Amsterdam, 1967.

Brandenburg, B., Lee, L., Lakadamyali, M., Rust, M., Zhuang, X., & Hogle, J. (2007). Imaging Poliovirus Entry in Live Cells. *PLoS Biology*, *5* (7), E183.

Brenner, S., & Horne, R. (1959). A Negative Staining Method for High Resolution Electron Microscopy of Viruses. *Biochimica Et Biophysica Acta*, *34* (1), 103-110.

Brisco, M., Hull, R., & Wilson, T. (1985). Southern Bean Mosaic Virus-Specific Proteins are Synthesized in an *in vitro* System Supplemented with Intact, Treated Virions. *Virology*, *143* (2), 392-398.

Brisco, M., Hull, R., & Wilson, T. (1986). Swelling of Isometric and of Bacilliform Plant-Virus Nucleocapsids is Required for Virus-Specific Protein-Synthesis *in vitro*. *Virology*, *148* (1), 210-217.

Brunger, A. (1997). Free R Value: Cross-Validation in Crystallography. *Macromolecular Crystallography, Pt A*, *277*, 366-396.

Brunger, A., Adams, P., Clore, G., Delano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., Warren, G. (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica Section D-Biological Crystallography*, *54*, 905-921.

Brunger, A., Krukowski, A., & Erickson, J. (1990). Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing. *Acta Crystallographica Section A*, *46*, 585-593.

Bubeck, D., Filman, D., & Hogle, J. (2005a). Cryo-Electron Microscopy Reconstruction of a Poliovirus-Receptor-Membrane Complex. *Nature Structural & Molecular Biology*, *12* (7), 615-8.

Bubeck, D., Filman, D., Cheng, N., Steven, A., Hogle, J., & Belnap, D. (2005b). The Structure of the Poliovirus 135S Cell Entry Intermediate at 10-Angstrom Resolution Reveals the Location of an Externalized Polypeptide that Binds to Membranes. *Journal of Virology*, *79* (12), 7745-55.

Bunka, D., Lane, S., Lane, C., Dykeman, E., Ford, R., Barker, A., Twarock, R., Phillips, S., Stockley, P. (2011). Degenerate RNA Packaging Signals in the Genome of Satellite Tobacco Necrosis Virus: Implications for the Assembly of a T=1 Capsid. *Journal of Molecular Biology, 413* (1), 51-65.

Cao, M., Ye, X., Willie, K., Lin, J., Zhang, X., Redinbaugh, M., Simon, A., Morris, T., 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* (15), 7793-802.

Carlisle, C., & Dornberger, K. (1948). Some X-Ray Measurements on Single Crystals of Tomato Bushy-Stunt Virus. *Acta Crystallographica*, *1* (1-6), 194-198.

Carrillo-Tripp, M., Shepherd, C., Borelli, I., Venkataraman, S., Lander, G., Natarajan, P., Johnson, J., Brooks, C., Reddy, V. (2009). VIPERdb(2): An Enhanced and Web API Enabled Relational Database for Structural Virology. *Nucleic Acids Research*, *37*, D436-D442.

Carrington, J., Heaton, L., Zuidema, D., Hillman, B., & Morris, T. (1989). The Genome Structure of Turnip Crinkle Virus. *Virology*, *170* (1), 219-226.

Caspar, D. (1956). Structure of Bushy Stunt Virus. *Nature*, *177* (4506), 476-477.

Caspar, D., & Klug, A. (1962). Physical Principles in Construction of Regular Viruses. *Cold Spring Harbor Symposia on Quantitative Biology*, *27*, 1-24.

Chen, V., Arendall, W., Headd, J., Keedy, D., Immormino, R., Kapral, G., Murray, L., Richardson, J., Richardson, D. (2010). Molprobity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallographica Section D-Biological Crystallography*, *66*, 12-21.

Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., Johnson, J. (1989). Protein-RNA Interactions in an Icosahedral Virus at 3.0 Å Resolution. *Sciencenew Series*, *245* (4914), 154-159.

Chescoe, D., & Goodhew, P. (1990). The Operation of Transmission and Scanning Electron Microscopes. *Oxford University Press*.

Chidlow, J., & Tremaine, J. (1971). Limited Hydrolysis of Cowpea Chlorotic Mottle Virus by Trypsin and Chymotrypsin. *Virology*, *43* (1), 267-278.

Crick, F. (1962). Nobel Lecture: On the Genetic Code. *Nobel Lectures, Physiology Or Medicine 1942-1962*. Elsevier Publishing Company, Amsterdam, 1964.

Crick, F., & Watson, J. (1956). Structure of Small Viruses. *Nature*, *177* (4506), 473-475.

Donnelly, M., Hughes, L., Luke, G., Mendoza, H., Ten Dam, E., Gani, D., Ryan, M. (2001). The 'Cleavage' Activities of Foot-and-Mouth Disease Virus 2A Site-Directed Mutants and Naturally Occurring '2A-Like' Sequences. *The Journal of General Virology*, *82*, 1027-1041.

Durham, A., Hendry, D., & Vonwechmar, M. (1977). Does Calcium-Ion Binding Control Plant-Virus Disassembly? *Virology*, *77* (2), 524-533.

Dykeman, E., Grayson, N., Toropova, K., Ranson, N., Stockley, P., & Twarock, R. (2011). Simple Rules for Efficient Assembly Predict the Layout of a Packaged Viral RNA. *Journal of Molecular Biology*, *408* (3), 399-407.

Eccles, R. (2005). Understanding the Symptoms of the Common Cold and Influenza. *The Lancet Infectious Diseases*, *5* (11), 718-25.

Egger, D., Teterina, N., Ehrenfeld, E., & Bienz, K. (2000). Formation of the Poliovirus Replication Complex Requires Coupled Viral Translation, Vesicle Production, and Viral RNA Synthesis. *Journal of Virology*, *74* (14), 6570-6580.

Emsley, P., & Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D-Biological Crystallography*, *60*, 2126-2132.

Evans, D. (1988). Regulation of Cytoplasmic Free Calcium by Plant-Cell Membranes. *Cell Biology International Reports*, *12* (5), 383-396.

Evilevitch, A., Fang, L., Yoffe, A., Castelnovo, M., Rau, D., Parsegian, V., Gelbart, W., Knobler, C. (2008). Effects of Salt Concentrations and Bending Energy on the Extent of Ejection of Phage Genomes. *Biophysical Journal*, *94* (3), 1110-1120.

Faruqi, A., & Henderson, R. (2007). Electronic Detectors for Electron Microscopy. *Current Opinion in Structural Biology*, *17* (5), 549-55.

FEI Company (2009). FEI's new direct electron detector revolutionizes Electron Microscopy of Biological and other Beam-sensitive Samples. Press release. http://investor.fei.com/releasedetail.cfm?ReleaseID=399045.

Finch, J., & Klug, A. (1959). Structure of Poliomyelitis Virus. *Nature*, *183* (4677), 1709-1714.

Fisher, A., & Johnson, J. (1993). Ordered Duplex RNA Controls Capsid Architecture in an Icosahedral Animal Virus. *Nature*, *361* (6408), 176-9.

Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies.* Oxford University Press, New York, USA. ISBN 978-019-518218-7.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A. (1996). Spider and Web: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Structural Biology*, *116* (1), 190-9.

Franklin, R. (1955). Structure of Tobacco Mosaic Virus. *Nature*, *175* (4452), 379-81.

Friedrich, W., Knipping, P., & Von Laue, M. (1912). Interferenz-Erscheinungen Bei Röntgenstrahlen. *Sitzungsberichte Der Mathematisch-Physikalischen Classe Der Königlich-Bayerischen Akademie Der Wissenschaften Zu Munchen*, *303*.

Fry, E., Tuthill, T., Harlos, K., Walter, T., Rowlands, D., & Stuart, D. (2010). Crystal Structure of Equine Rhinitis A Virus in Complex with its Sialic Acid Receptor. *The Journal of General Virology*, *91*, 1971-1977.

Garman, E. (2010). Radiation Damage in Macromolecular Crystallography: What is it and Why Should We Care? *Acta Crystallographica Section D-Biological Crystallography*, *66*, 339-351.

Golden, J., & Harrison, S. (1982). Proteolytic Dissection of Turnip Crinkle Virus Subunit in Solution. *Biochemistry*, *21* (16), 3862-3866.

Grahn, E., Stonehouse, N., Murray, J., Van Den Worm, S., Valegard, K., Fridborg, K., Stockley, P., Liljas, L. (1999). Crystallographic Studies of RNA Hairpins in Complexes with Recombinant MS2 Capsids: Implications for Binding Requirements. *RNA*, *5* (1), 131-138.

Groppelli, E., Tuthill, T., & Rowlands, D. (2010). Cell Entry of the Aphthovirus Equine Rhinitis A Virus is Dependent on Endosome Acidification. *Journal of Virology*, *84* (12), 6235-6240.

Grubman, M., & Baxt, B. (2004). Foot-and-Mouth Disease. *Clinical Microbiology Reviews*, *17* (2), 465-493.

Guerois R., Nielsen J., Serrano L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of Molecular Biology, 320*, 369-387

Harris, J., & Hindley, J. (1961). The Protein Subunit of Turnip Yellow Mosaic Virus. *Journal of Molecular Biology*, *3*, 117-20.

Harris, J., & Knight, C. (1955). Studies on the Action of Carboxypeptidase on Tobacco Mosaic Virus. *Journal of Biological Chemistry*, *214* (1), 215-30.

Harrison, S. (1980). Protein Interfaces and Intersubunit Bonding. The Case of Tomato Bushy Stunt Virus. *Biophysical Journal*, *32* (1), 139-53.

Harrison, S., Olson, A., Schutt, C., Winkler, F., & Bricogne, G. (1978). Tomato Bushy Stunt Virus at 2.9-Å Resolution. *Nature*, *276* (5686), 368-373.

Harvey, J., Lewsey, M., Patel, K., Westwood, J., Heimstaedt, S., Carr, J., Baulcombe, D. (2011). An Antiviral Defense Role of AGO2 in Plants. *PLoS One*, *6* (1), E14639.

Hearne, P., Knorr, D., Hillman, B., & Morris, T. (1990). The Complete Genome Structure and Synthesis of Infectious RNA From Clones of Tomato Bushy Stunt Virus. *Virology*, *177* (1), 141-151.

Hodgkin, D. (1964). Nobel Lecture: The X-Ray Analysis of Complicated Molecules. *Nobel Lectures, Chemistry 1963-1970.* Elsevier Publishing Company, Amsterdam, 1972.

Hogle, J. (2002). Poliovirus Cell Entry: Common Structural Themes in Viral Cell Entry Pathways. *Annual Review of Microbiology*, *56*, 677-702.

Hogle, J., Chow, M., & Filman, D. (1985). 3-Dimensional Structure of Poliovirus at 2.9 Å Resolution. *Science (New York, Ny)*, *229* (4720), 1358-1365.

Hogle, J., Maeda, A., & Harrison, S. (1986). Structure and Assembly of Turnip Crinkle Virus .1. X-Ray Crystallographic Structure-Analysis at 3.2 Å Resolution. *Journal of Molecular Biology*, *191* (4), 625-638.

Hope, H. (1988). Cryocrystallography of Biological Macromolecules - A Generally Applicable Method. *Acta Crystallographica Section B-Structural Science*, *44*, 22-26.

Hopper, P., Harrison, S., & Sauer, R. (1984). Structure of Tomato Bushy Stunt Virus. V. Coat Protein Sequence Determination and its Structural Implications. *Journal of Molecular Biology*, *177* (4), 701-13.

Horne, R., & Wildy, P. (1961). Symmetry in Virus Architecture. *Virology*, *15* (3), 348-373.

Huber, R. (1988). A Structural Basis of Light Energy and Electron Transfer in Biology. *Nobel Lectures, Chemistry 1981-1990,* World Scientific Publishing Co., Singapore, 1992.

Ivanovska, I., Wuite, G., Jonsson, B., & Evilevitch, A. (2007). Internal DNA Pressure Modifies Stability of WT Phage. *Proceedings of the National Academy of Sciences of the United States of America*, *104* (23), 9603-9608.

Johnson, K., Tang, L., Johnson, J., & Ball, L. (2004). Heterologous RNA Encapsidated in Pariacoto Virus-Like Particles Forms a Dodecahedral Cage Similar to Genomic RNA in Wild-Type Virions. *Journal of Virology*, *78* (20), 11371-8.

Jones, S., Daley, D., Luscombe, N., Berman, H., & Thornton, J. (2001). Protein-RNA Interactions: a Structural Analysis. *Nucleic Acids Research*, *29* (4), 943-54.

Jones, T., & Liljas, L. (1984). Structure of Satellite Tobacco Necrosis Virus After Crystallographic Refinement at 2.5-Å Resolution. *Journal of Molecular Biology*, *177* (4), 735-767.

Kabsch, W. (1988). Automatic-Indexing of Rotation Diffraction Patterns. *Journal of Applied Crystallography*, *21*, 67-71.

Katpally, U., Kakani, K., Reade, R., Dryden, K., Rochon, D., & Smith, T. (2007). Structures of *T*=1 and *T*=3 Particles of Cucumber Necrosis Virus: Evidence of Internal Scaffolding. *Journal of Molecular Biology*, *365* (2), 502-512.

Kausche, G., Pfankuch, E., & Ruska, H. (1939). Die Sichtbarmachung Von Pflanzlichem Virus Im Ubermikroskop. *Naturwissenschaften*, *27* (18), 292-300.

Kawai, T., & Akira, S. (2006). Innate Immune Recognition of Viral Infection. *Nature Immunology*, *7* (2), 131-137.

Kleywegt, G., & Jones, T. (1996). xdlMAPMAN and xdlDATAMAN - Programs for Reformatting, Analysis and Manipulation of Biomacromolecular Electron-Density Maps and Reflection Data Sets. *Acta Crystallographica Section D-Biological Crystallography*, *52*, 826-828.

Klug, A., Finch, J., & Franklin, R. (1957). Structure of Turnip Yellow Mosaic Virus. *Nature*, *179* (4561), 683-684.

Klug, A., Finch, J., & Franklin, R. (1957). The Structure of Turnip Yellow Mosaic Virus - X-Ray Diffraction Studies. *Biochimica Et Biophysica Acta*, *25* (2), 242-252.

Kobayashi, T., Ooms, L., Chappell, J., & Dermody, T. (2009). Identification of Functional Domains in Reovirus Replication Proteins μNS and μ2. *Journal of Virology*, *83* (7), 2892-2906.

Koning, R., Van Den Worm, S., Plaisier, J., Van Duin, J., 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* (2), 415-422.

Laemmli, U. (1970). Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature*, *227* (5259), 680-&.

Laliberté, J.-F., & Sanfaçon, H. (2010). Cellular Remodeling During Plant Virus Infection. *Annual Review of Phytopathology*, *48*, 69-91.

Lane, S., Dennis, C., Lane, C., Trinh, C., Rizkallah, P., Stockley, P., Phillips, S. (2011). Construction and Crystal Structure of Recombinant STNV Capsids. *Journal of Molecular Biology, 413* (1), 41-50.

Larson, S., & McPherson, A. (2001). Satellite Tobacco Mosaic Virus RNA: Structure and Implications for Assembly. *Current Opinion in Structural Biology*, *11* (1), 59-65.

Larson, S., Day, J., Canady, M., Greenwood, A., & McPherson, A. (2000). Refined Structure of Desmodium Yellow Mottle Tymovirus at 2.7 Angstrom Resolution. *Journal of Molecular Biology*, *301* (3), 625-642.

Larson, S., Day, J., Greenwood, A., & McPherson, A. (1998). Refined Structure of Satellite Tobacco Mosaic Virus at 1.8 Å Resolution. *Journal of Molecular Biology*, *277* (1), 37-59.

Larson, S., Koszelak, S., Day, J., Greenwood, A., Dodds, J., & McPherson, A. (1993). Double-Helical RNA in Satellite Tobacco Mosaic Virus. *Nature*, *361* (6408), 179-82.

Larson, S., Lucas, R., Greenwood, A., & McPherson, A. (2005). The RNA of Turnip Yellow Mosaic Virus Exhibits Icosahedral Order. *Virology*, *334* (2), 245-54.

Lee, K. (2010). Architecture of a Nascent Viral Fusion Pore. *The EMBO Journal*, *29* (7), 1299-311.

Levy, H., Bostina, M., Filman, D., & Hogle, J. (2010). Catching a Virus in the Act of RNA Release: a Novel Poliovirus Uncoating Intermediate Characterized by Cryo-Electron Microscopy. *Journal of Virology*, *84* (9), 4426-4441.

Li, F., Browning, G., Studdert, M., & Crabb, B. (1996). Equine Rhinovirus 1 is More Closely Related to Foot-and-Mouth Disease Virus Than to Other Picornaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, *93* (3), 990-995.

Liljas, L., Tate, J., Lin, T., Christian, P., & Johnson, J. (2002). Evolutionary and Taxonomic Implications of Conserved Structural Motifs Between Picornaviruses and Insect Picorna-Like Viruses. *Archives of Virology*, *147* (1), 59-84.

Liljas, L., Unge, T., Jones, T., Fridborg, K., Lövgren, S., Skoglund, U., Strandberg, B. (1982). Structure of Satellite Tobacco Necrosis Virus at 3.0 Å Resolution. *Journal of Molecular Biology*, *159* (1), 93-108.

Lin, J., Cheng, N., Chow, M., Filman, D., Steven, A., Hogle, J., Belnap, D. (2011). An Externalized Polypeptide Partitions Between Two Distinct Sites on Genome-Released Poliovirus Particles. *Journal of Virology, 85* (19), 9974-9983.

Lin, T., Cavarelli, J., & Johnson, J. (2003). Evidence for Assembly-Dependent Folding of Protein and RNA in an Icosahedral Virus. *Virology*, *314* (1), 26-33.

Ludtke, S., Baldwin, P., & Chiu, W. (1999). Eman: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology*, *128* (1), 82-97.

Makinen, M., & Fink, A. (1977). Reactivity and Cryoenzymology of Enzymes in Crystalline State. *Annual Review of Biophysics and Bioengineering*, *6*, 301-343.

Malhotra, A., Penczek, P., Agrawal, R., Gabashvili, I., Grassucci, R., Junemann, R., Burkhardt, N., Nierhaus, K., Frank, J. (1998). *Escherichia Coli* 70 S Ribosome at 15

Angstrom Resolution by Cryo-Electron Microscopy: Localization of fMet-tRNA$_f^{Met}$ and Fitting of L1 Protein. *Journal of Molecular Biology*, *280* (1), 103-116.

Mason, P., Grubman, M., & Baxt, B. (2003). Molecular Basis of Pathogenesis of FMDV. *Virus Research*, *91* (1), 9-32.

Matthews, R., & Witz, J. (1985). Uncoating of Turnip Yellow Mosaic-Virus RNA *in vivo*. *Virology*, *144* (2), 318-327.

Maurer, U., Sodeik, B., & Grünewald, K. (2008). Native 3D Intermediates of Membrane Fusion in Herpes Simplex Virus 1 Entry. *Proceedings of the National Academy of Sciences of the United States of America*, *105* (30), 10559-64.

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

Mettenleiter, T., Klupp, B., & Granzow, H. (2009). Herpesvirus Assembly: an Update. *Virus Research*, *143* (2), 222-234.

Mindell, J., & Grigorieff, N. (2003). Accurate Determination of Local Defocus and Specimen Tilt in Electron Microscopy. *Journal of Structural Biology*, *142* (3), 334-347.

Montelius, I., Liljas, L., & Unge, T. (1990). Sequential Removal of Ca^{2+} From Satellite Tobacco Necrosis Virus - Crystal-Structure of 2 EDTA-Treated Forms. *Journal of Molecular Biology*, *212* (2), 331-343.

Montelius, I., Liljas, L., & Unge, T. (1988). Structure of EDTA-Treated Satellite Tobacco Necrosis Virus at pH-6.5. *Journal of Molecular Biology*, *201* (2), 353-363.

Morton, V., Stockley, P., Stonehouse, N., & Ashcroft, A. (2008). Insights Into Virus Capsid Assembly From Non-Covalent Mass Spectrometry. *Mass Spectrometry Reviews*, *27* (6), 575-595.

Muramoto, Y., Takada, A., Fujii, K., Noda, T., Iwatsuki-Horimoto, K., Watanabe, S., Horimoto, T., Kida, H., Kawaoka, Y. (2006). Hierarchy Among Viral RNA (vRNA) Segments in their Role in vRNA Incorporation Into Influenza A Virions. *Journal of Virology*, *80* (5), 2318-2325.

Murshudov, G., Vagin, A., & Dodson, E. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D-Biological Crystallography*, *53*, 240-255.

Nathanson, N. (2011). Eradication of Poliovirus: Fighting Fire with Fire. *Journal of Infectious Diseases*, *203* (7), 889-890.

Nave, C. (1995). Radiation-Damage in Protein Crystallography. *Radiation Physics and Chemistry*, *45* (3), 483-490.

Nugent, C., Johnson, K., Sarnow, P., & Kirkegaard, K. (1999). Functional Coupling Between Replication and Packaging of Poliovirus Replicon RNA. *Journal of Virology*, *73* (1), 427-435.

Oda, Y., Saeki, K., Takahashi, Y., Maeda, T., Naitow, H., Tsukihara, T., Fukuyama, K. (2000). Crystal Structure of Tobacco Necrosis Virus at 2.25 Angstrom Resolution. *Journal of Molecular Biology*, *300* (1), 153-169.

Odegard, A., Kwan, M., Walukiewicz, H., Banerjee, M., Schneemann, A., & Johnson, J. (2009). Low Endocytic pH and Capsid Protein Autocleavage are Critical Components of Flock House Virus Cell Entry. *Journal of Virology*, *83* (17), 8628-37.

Odhner, C. (1901). Nobel Award Ceremony Speech. *Nobel Lectures, Physics 1901-1921.* Elsevier Publishing Company, Amsterdam, 1967.

Oh, J., Kong, Q., Song, C., Carpenter, C., & Simon, A. (1995). Open Reading Frames of Turnip Crinkle Virus Involved in Satellite Symptom Expression and Incompatibility with *Arabidopsis thaliana* Ecotype Dijon. *Molecular Plant-Microbe Interactions*, *8* (6), 979-87.

Opalka, N., Tihova, M., Brugidou, C., Kumar, A., Beachy, R., Fauquet, C., Yeager, M. (2000). Structure of Native and Expanded Sobemoviruses by Electron Cryo-Microscopy and Image Reconstruction. *Journal of Molecular Biology*, *303* (2), 197-211.

Orlova, E., Dube, P., Harris, J., Beckman, E., Zemlin, F., Markl, J., Van Heel, M. (1997). Structure of Keyhole Limpet Hemocyanin Type 1 (KLH1) at 15 Angstrom

Resolution by Electron Cryomicroscopy and Angular Reconstitution. *Journal of Molecular Biology*, *271* (3), 417-437.

Perutz, M. (1962). Nobel Lecture: X-Ray Analysis of Haemoglobin. *Nobel Lectures, Chemistry 1942-1962.* Elsevier Publishing Company, Amsterdam, 1964.

Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., Ferrin, T. (2004). UCSF Chimera - a Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, *25* (13), 1605-1612.

Pittman, J., & Hirschi, K. (2003). Don't Shoot the (Second) Messenger: Endomembrane Transporters and Binding Proteins Modulate Cytosolic Ca^{2+} Levels. *Current Opinion in Plant Biology*, *6* (3), 257-262.

Poranen, M., Daugelavicius, R., & Bamford, D. (2002). Common Principles in Viral Entry. *Annual Review of Microbiology*, *56*, 521-38.

Prangishvili, D., Forterre, P., & Garrett, R. (2006). Viruses of the Archaea: a Unifying View. *Nature Reviews Microbiology*, *4* (11), 837-48.

Qu, F., & Morris, T. (2000). Cap-Independent Translational Enhancement of Turnip Crinkle Virus Genomic and Subgenomic RNAs. *Journal of Virology*, *74* (3), 1085-93.

Qu, F., & Morris, T. (1997). Encapsidation of Turnip Crinkle Virus is Defined by a Specific Packaging Signal and RNA Size. *Journal of Virology*, *71* (2), 1428-1435.

Ramakrishnan, V. (2009). Nobel Lecture: Unraveling the Structure of the Ribosome. *Les Prix Nobel. The Nobel Prizes 2009.* Nobel Foundation, Stockholm, 2010.

Ravelli, R., & Garman, E. (2006). Radiation Damage in Macromolecular Cryocrystallography. *Current Opinion in Structural Biology*, *16* (5), 624-629.

Reddy, V., Natchiar, S., Stewart, P., & Nemerow, G. (2010). Crystal Structure of Human Adenovirus at 3.5 Angstrom Resolution. *Science*, *329* (5995), 1071-1075.

Reitter, J., Means, R., & Desrosiers, R. (1998). A Role for Carbohydrates in Immune Evasion in Aids. *Nature Medicine*, *4* (6), 679-684.

Rice, R. (1974). Minor Protein Components in Cowpea Chlorotic Mottle Virus and Satellite of Tobacco Necrosis Virus. *Virology*, *61* (1), 249-55.

Robinson, I., & Harrison, S. (1982). Structure of the Expanded State of Tomato Bushy Stunt Virus. *Nature*, *297* (5867), 563-568.

Rochon, D., & Tremaine, J. (1989). Complete Nucleotide-Sequence of the Cucumber Necrosis Virus Genome. *Virology*, *169* (2), 251-259.

Roenhorst, J., Verduin, B., & Goldbach, R. (1989). Virus Ribosome Complexes From Cell-Free Translation Systems Supplemented with Cowpea Chlorotic Mottle Virus-Particles. *Virology*, *168* (1), 138-146.

Rolfsson, O., Toropova, K., Ranson, N., & Stockley, P. (2010). Mutually-Induced Conformational Switching of RNA and Coat Protein Underpins Efficient Assembly of a Viral Capsid. *Journal of Molecular Biology*, *401* (2), 309-322.

Roseman, A. (2004). FindEM - A Fast, Efficient Program for Automatic Selection of Particles from Electron Micrographs. *Journal of Structural Biology, 145,* (1-2), 91-99.

Rosenthal, P., & Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy. *Journal of Molecular Biology*, *333* (4), 721-745.

Rossmann, M., & Blow, D. (1962). Detection of Sub-Units Within Crystallographic Asymmetric Unit. *Acta Crystallographica*, *15*, 24-31.

Rueckert, R., & Wimmer, E. (1984). Systematic Nomenclature of Picornavirus Proteins. *Journal of Virology*, *50* (3), 957-959.

Ruska, E. (1986). Nobel Lecture: the Development of the Electron Microscope and Electron Microscopy. *Nobel Lectures, Physics 1981-1990,* World Scientific Publishing Co., Singapore, 1993.

Ruska, H. (1939). Ubermikroskopische Untersuchungstechnik. *Naturwissenschaften*, *27* (18), 287-292.

Sacher, R., & Ahlquist, P. (1989). Effects of Deletions in the N-Terminal Basic Arm of Brome Mosaic-Virus Coat Protein on RNA Packaging and Systemic Infection. *Journal of Virology*, *63* (11), 4545-4552.

Scholthof, H. (2006). Timeline - the Tombusvirus-Encoded p19: from Irrelevance to Elegance. *Nature Reviews Microbiology*, *4* (5), 405-411.

Schuck, P. (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal*, *78* (3), 1606-1619.

Schymkowitz J. W., Rousseau F., Martins I. C., Ferkinghoff-Borg J., Stricher F., Serrano L. (2005). Prediction of Water and Metal Binding Sites and their Affinities by Using the Fold-X Force Field. Proc Natl Acad Sci USA, vol 102, p 10147-52.

Shaw, J., Plaskitt, K., & Wilson, T. (1986). Evidence that Tobacco Mosaic-Virus Particles Disassemble Cotranslationally *in vivo*. *Virology*, *148* (2), 326-336.

Shechner, D., & Bartel, D. (2011). The Structural Basis of RNA-Catalyzed RNA Polymerization. *Nature Structural & Molecular Biology*, *18* (9), 1036-42.

Sherman, M., Guenther, R., Tama, F., Sit, T., Brooks, C., Mikhailov, A., Orlova, E., Baker, T., Lommel, S. (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* (21), 10395-406.

Smith, G., Duax, W., Dodson, E., Dodson, G., Degraaf, R., & Reynolds, C. (1982). The Structure of Des-Phe Bl Bovine Insulin. *Acta Crystallographica Section B-Structural Science*, *38*, 3028-3032.

Smith, J., Silvestry, M., Lindert, S., Lu, W., Nemerow, G., & Stewart, P. (2010). Insight Into the Mechanisms of Adenovirus Capsid Disassembly From Studies of Defensin Neutralization. *PLoS Pathogens*, *6* (6), E1000959.

Sorger, P., Stockley, P., & Harrison, S. (1986). Structure and Assembly of Turnip Crinkle Virus .2. Mechanism of Reassembly Invitro. *Journal of Molecular Biology*, *191* (4), 639-658.

Speir, J., Munshi, S., Wang, G., Baker, T., & Johnson, J. (1995). Structures of the Native and Swollen Forms of Cowpea Chlorotic Mottle Virus Determined by X-Ray Crystallography and Cryoelectron Microscopy. *Structure*, *3* (1), 63-78.

Stockley, P., Kirsh, A., Chow, E., Smart, J., & Harrison, S. (1986). Structure of Turnip Crinkle Virus .3. Identification of a Unique Coat Protein Dimer. *Journal of Molecular Biology*, *191* (4), 721-725.

Stockley, P., Rolfsson, O., Thompson, G., Basnak, G., Francese, S., Stonehouse, N., Homans, S., Ashcroft, A. (2007). A Simple, RNA-Mediated Allosteric Switch Controls the Pathway to Formation of a T=3 Viral Capsid. *Journal of Molecular Biology*, *369* (2), 541-552.

Stockley, P., Stonehouse, N., Murray, J., Goodman, S., Talbot, S., Adams, C., Liljas, L., Valegard, K. (1995). Probing Sequence-Specific RNA Recognition by the Bacteriophage-MS2 Coat Protein. *Nucleic Acids Research*, *23* (13), 2512-2518.

Stoscheck, C. (1990). Quantitation of Protein. *Methods in Enzymology, 182,* 50-68.

Stuart, D. (1993). Viruses. *Current Opinion in Structural Biology*, *3* (2), 167-174.

Stupina, V., Meskauskas, A., McCormack, J., Yingling, Y., Shapiro, B., Dinman, J., Simon, A. (2008). The 3' Proximal Translational Enhancer of Turnip Crinkle Virus Binds to 60s Ribosomal Subunits. *RNA*, *14* (11), 2379-93.

Subramaniam, S., & Henderson, R. (2000). Molecular Mechanism of Vectorial Proton Translocation by Bacteriorhodopsin. *Nature*, *406* (6796), 653-657.

Summerfield, A., Guzylack-Piriou, L., Harwood, L., & McCullough, K. (2009). Innate Immune Responses Against Foot-and-Mouth Disease Virus: Current Understanding and Future Directions. *Veterinary Immunology and Immunopathology*, *128* (1-3), 205-210.

Sun, S., Rao, V., & Rossmann, M. (2010). Genome Packaging in Viruses. *Current Opinion in Structural Biology*, *20* (1), 114-120.

Tang, L., Johnson, K., Ball, L., Lin, T., Yeager, M., & Johnson, J. (2001). The Structure of Pariacoto Virus Reveals a Dodecahedral Cage of Duplex RNA. *Nature Structural Biology*, *8* (1), 77-83.

Taylor, K., & Glaeser, R. (1974). Electron-Diffraction of Frozen, Hydrated Protein Crystals. *Science (New York, Ny)*, *186* (4168), 1036-1037.

Taylor, K., & Glaeser, R. (2008). Retrospective on the Early Development of Cryoelectron Microscopy of Macromolecules and a Prospective on Opportunities for the Future. *Journal of Structural Biology*, *163* (3), 214-223.

Thomas, C., Leh, V., Lederer, C., & Maule, A. (2003). Turnip Crinkle Virus Coat Protein Mediates Suppression of RNA Silencing in *Nicotiana benthamiana*. *Virology*, *306* (1), 33-41.

Thompson, D., Muriel, P., Russell, D., Osborne, P., Bromley, A., Rowland, M., Creigh-Tyte, S., Brown, C. (2002). Economic Costs of the Foot and Mouth Disease Outbreak in the United Kingdom in 2001. *Rev. Sci. Tech. Off. Int. Epiz.*, *21* (3), 675-687.

Tickle, I., Laskowski, R., & Moss, D. (1998). Rfree and the Rfree Ratio. I. Derivation of Expected Values of Cross-Validation Residuals Used in Macromolecular Least-Squares Refinement. *Acta Crystallographica Section D, Biological Crystallography*, *54* (Pt 4), 547-57.

Tock, M., & Dryden, D. (2005). The Biology of Restriction and Anti-Restriction. *Current Opinion in Microbiology*, *8* (4), 466-472.

Toropova, K., Basnak, G., Twarock, R., Stockley, P., & Ranson, N. (2008). The Three-Dimensional Structure of Genomic RNA in Bacteriophage MS2: Implications for Assembly. *Journal of Molecular Biology*, *375* (3), 824-36.

Toropova, K., Stockley, P., & Ranson, N. (2011). Visualising a Viral RNA Genome Poised for Release From its Receptor Complex. *Journal of Molecular Biology*, *408* (3), 408-19.

Tremaine, J., & Ronald, W. (1978). Limited Proteolysis of Southern Bean Mosaic-Virus by Trypsin. *Virology*, *91* (1), 164-172.

Tsunasawa, S., & Sakiyama, F. (1984). Amino-Terminal Acetylation of Proteins: an Overview. *Methods in Enzymology*, *106*, 165-70.

Tsuruta, H., Reddy, V., Wikoff, W., & Johnson, J. (1998). Imaging RNA and Dynamic Protein Segments with Low-Resolution Virus Crystallography: Experimental

Design, Data Processing and Implications of Electron Density Maps. *Journal of Molecular Biology*, *284* (5), 1439-52.

Tuthill, T., Groppelli, E., Hogle, J., & Rowlands, D. (2010). Picornaviruses. Cell Entry by Non-Enveloped Viruses, *Current Topics in Microbiology, 343*, 43-89.

Tuthill, T., Harlos, K., Walter, T., Knowles, N., Groppelli, E., Rowlands, D., Stuart, D., Fry, E. (2009). Equine Rhinitis A Virus and its Low pH Empty Particle: Clues Towards an Aphthovirus Entry Mechanism? *PLoS Pathogens*, *5* (10), E1000620.

Unge, T., Montelius, I., Liljas, L., & Ofverstedt, L. (1986). The EDTA-Treated Expanded Satellite Tobacco Necrosis Virus - Biochemical-Properties and Crystallization. *Virology*, *152* (1), 207-218.

Valegard, K., Liljas, L., Fridborg, K., & Unge, T. (1990). The 3-Dimensional Structure of the Bacterial-Virus MS2. *Nature*, *345* (6270), 36-41.

Valegard, K., Murray, J., Stockley, P., Stonehouse, N., & Liljas, L. (1994). Crystal-Structure of an Bacteriophage-RNA Coat Protein-Operator Complex. *Nature*, *371* (6498), 623-626.

Valegard, K., Murray, J., Stonehouse, N., Vandenworm, S., Stockley, P., & Liljas, L. (1997). The Three-Dimensional Structures of Two Complexes Between Recombinant MS2 Capsids and RNA Operator Fragments Reveal Sequence-Specific Protein-RNA Interactions. *Journal of Molecular Biology*, *270* (5), 724-738.

Van Den Worm, S., Koning, R., Warmenhoven, H., Koerten, H., & Van Duin, J. (2006). Cryo Electron Microscopy Reconstructions of the Leviviridae Unveil the Densest Icosahedral RNA Packing Possible. *Journal of Molecular Biology*, *363* (4), 858-65.

Van Heel, M., Harauz, G., Orlova, E., Schmidt, R., & Schatz, M. (1996). A New Generation of the Imagic Image Processing System. *Journal of Structural Biology*, *116* (1), 17-24.

Vargason, J., Szittya, G., Burgyan, J., & Hall, T. (2003). Size Selective Recognition of siRNA by an RNA Silencing Suppressor. *Cell*, *115* (7), 799-811.

Von Borries, B., & Ruska, E. (1939). Eigenschaften Der Ubermikroskopischen Abbildung. *Naturwissenschaften*, *27* (18), 281-287.

Von Laue, M. (1915). Nobel Lecture: Concerning the Detection of X-Ray Interferences. *Nobel Lectures, Physics 1901-1921* .

Walker, M., Trinick, J., & White, H. (1995). Millisecond Time Resolution Electron Cryo-Microscopy of the M-ATP Transient Kinetic State of the Acto-Myosin ATPase. *Biophysical Journal*, *68* (4 Suppl), 87s-91s.

Watson, J. (1954). The Structure of Tobacco Mosaic Virus. I. X-Ray Evidence of a Helical Arrangement of Sub-Units Around the Longitudinal Axis. *Biochimica Et Biophysica Acta*, *13* (1), 10-9.

Wei, N., Heaton, L., Morris, T., & Harrison, S. (1990). Structure and Assembly of Turnip Crinkle Virus. VI. Identification of Coat Protein Binding Sites on the RNA. *Journal of Molecular Biology*, *214* (1), 85-95.

Wetz, K., & Kucinski, T. (1991). Influence of Different Ionic and pH Environments on Structural Alterations of Poliovirus and their Possible Relation to Virus Uncoating. *The Journal of General Virology*, *72*, 2541-2544.

Wien, M., Curry, S., Filman, D., & Hogle, J. (1997). Structural Studies of Poliovirus Mutants that Overcome Receptor Defects. *Nature Structural Biology*, *4* (8), 666-674.

Winn, M., Ballard, C., Cowtan, K., Dodson, E., Emsley, P., Evans, P., Keegan, R., Krissinel, E., Leslie, A., McCoy, A., McNicholas, S., Murshudov, G., Pannu, N., Potterton, E., Powell, H., Read, R., Vagin, A., Wilson, K. (2011). Overview of the CCP4 Suite and Current Developments. *Acta Crystallographica Section D-Biological Crystallography*, *67*, 235-242.

Winter, G. (2010). Xia2: an Expert System for Macromolecular Crystallography Data Reduction. *Journal of Applied Crystallography*, *43*, 186-190.

Xiao, C., Kuznetsov, Y., Sun, S., Hafenstein, S., Kostyuchenko, V., Chipman, P., Suzan-Monti, M., Raoult, D., McPherson, A., Rossmann, M. (2009). Structural Studies of the Giant Mimivirus. *PLoS Biology*, *7* (4), 958-966.

Yamazaki, H., & Kaesberg, P. (1961). The Preparation and Some Properties of Protein Subunits Obtained From Wild Cucumber Mosaic Virus. *Biochimica Et Biophysica Acta*, *53*, 173-80.

Yeates, T. (1997). Detecting and Overcoming Crystal Twinning. *Macromolecular Crystallography, Pt A*, *276*, 344-358.

Yoffe, A., Prinsen, P., Gopal, A., Knobler, C., Gelbart, W., & Ben-Shaul, A. (2008). Predicting the Sizes of Large RNA Molecules. *Proceedings of the National Academy of Sciences of the United States of America*, *105* (42), 16153-8.

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

Zhang, X., Jin, L., Fang, Q., Hui, W., & Zhou, Z. (2010). 3.3 Angstrom Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell*, *141* (3), 472-482.

Zhu, Y., Carragher, B., Glaeser, R., Fellmann, D., Bajaj, C., Bern, M., Mouche, F., de Haas, F., Hall, R., Kriegman, D., Ludtke, S., Mallick, S., Penczek, P., Roseman, A., Sigworth, F., Volkmann, N., Potter, C. (2004). Automatic Particle Selection: Results of a Comparative Study. *Journal of Structural Biology, 145,* (1-2), 3-14.

Zlotnick, A., Aldrich, R., Johnson, J., Ceres, P., Young, M. (2000). Mechanism of Assembly for an Icosahedral Plant Virus. *Virology, 227,* (2), 450-456.