ABSTRACT

Title: TURNIP CRINKLE VIRUS COAT PROTEIN EXPRESSION AND PURIFICATION

Kathleen E. Dori, Master of Science, 2005

Directed By: Professor Dorothy Beckett
Department of Chemistry and Biochemistry

Turnip crinkle virus is a (+) sense single strand RNA plant virus consisting of a single copy of the genome surrounded by 180 copies of the 38 kDa coat protein (CP) that comprise the virion capsid. As a structural protein, the CP's main function is assembly of the virion but it has also been linked to suppression of virus-induced gene silencing. The CP gene was cloned and protein expressed in *Escherichia coli*. Protein formed insoluble aggregates packaged into inclusion bodies located in the cell cytoplasm. Cells were mechanically disrupted/lysed and inclusion bodies isolated via 8 M urea extraction. Soluble TCV CP was purified by ion-exchange chromatography using a CM Sepharose column under denaturing conditions. Subsequent step-wise dialysis out of urea and into a high salt buffer refolded the denatured protein, allowing the assembly of a series of aggregates. The purified protein was partially characterized through sedimentation and circular dichroism measurements.

TURNIP CRINKLE VIRUS COAT PROTEIN EXPRESSION AND PURIFICATION

By

Kathleen E. Dori

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:

Professor Dorothy Beckett, Chair
Professor Douglas Julin
Professor David Fushman

Dedication

For my parents

Acknowledgements

I would like to thank everyone who helped me get started in the lab and was there when I needed the support. All of you made my graduate experience a pleasant one.

Thanks goes especially to Dr. Beckett for allowing me to work on this specific project and supporting me through the completion of my degree. I appreciate all the advice both with regards to my lab work and writing techniques. The time in the lab has given me the experience and outlook I needed to pursue a future in research.

To everyone in the lab – Thanks! I would particularly like to thank Anne Colgrove who initially got me started on this project and answered many of my questions about TCV. Much of my gratitude goes to Emily Streaker who was there for me when I had a problem and also when I needed someone to talk to. I would also like to thank Patrick Brown for his help with the analytical ultracentrifugation and his wife Kristy Brown for performing the mass spectroscopy that I needed during a difficult period of my research.

In regards to other techniques performed, I would like to thank Tim Maugel at the University of Maryland College Park Laboratory for Biological Ultrastructure for his help and instruction with using the transmission electron microscope.

Table of Contents

List of Figures

I. Introduction

The goal of this thesis work was to obtain active recombinant viral capsid protein of *Turnip crinkle virus* (TCV) for further experimentation on virus assembly and its relationship to suppression of virus-induced gene silencing (VIGS) in plants. Specifically, the TCV coat protein gene has been cloned and the recombinant protein expressed and purified from *Escherichia coli*. The purified protein has been characterized for future use in determining the mechanism of its assembly into virions and the relationship between virion assembly and the role of the coat protein in inhibiting gene silencing. The topics covered in this introduction are intended to clarify, in general and with respect to TCV, the icosahedral structure, proposed mechanism of assembly and the role of the viral coat protein in VIGS.

A. Virus Classification

A virus is an obligate infectious parasite that uses a host organism during part of its life cycle. Its survival depends on its ability to utilize the infected host's machinery and stored energy in order to replicate its own genome. The infectious form of the virus is the complete virus particle or virion. The interior genome and exterior shell, or capsid, together form the nucleoprotein particle. Some viruses also contain a lipid membrane and associated proteins or viral envelope that surrounds the capsid (*1*). The virus must be able to package its genetic material into a stable virion that can withstand the harsh exterior

environment and disassemble readily once inside a newly infected host cell. Once inside the host it must be able to transport its genetic material from one infected cell to another.

Viruses can be classified based on the symmetry of the protein arrangement that forms the capsid. Three broad classes can be defined including helical, icosahedral and complex virions. In the helical arrangement proteins arrange themselves around a single central axis while an icosahedral arrangement is classified by the virion's spherical outer shell with 2, 3 and 5-fold axes of symmetry. These classes can be broken down further based on the type of nucleic acid, either single-stranded, double-stranded, DNA or RNA, the presence of an outer lipid envelope and various other structural features (*1*).

B. Icosahedral Virus Structure

1. Virion Structure

The structure of the virion is defined by the coat protein arrangement around the interior genome. In the case of the icosahedral virus the proteins arrange to resemble a sphere. Not all viruses are spherical in shape and some, such as the helical *Tobacco mosaic virus*, form rods of different lengths depending on the length of the genome (see Figure 1 a & b). Unlike helical viruses that do not have a set maximum length on the genetic material that can be packaged, icosahedral viruses have a predetermined genome size length due to the limited internal volume of the capsid.

The simplest example of an icosahedral virion is comprised of 20 triangular faces each consisting of three CP subunits. Thus, the smallest number of subunits that can be packaged to form an icosahedron is 60 (triangulation number or T = 1). As the virion size increases it does so by a factor of 60 subunits so that T > 1. In a T = 1 virion, all subunits in each face have identical conformations and subunit- subunit interactions. These virions contain 2-fold, 3-fold and 5-fold axes of symmetry (see Figure 1 c & d). Due to the limited amount of internal packaging space for the genome there are only a few known T = 1 virions. Most icosahedral viruses are larger in size and consist of more CPs with an accompanying increased triangulation number (*1*). The T >1, unlike the T = 1 capsids, consist of hexamer, planar orientation, and pentamer, produced by removing one subunit from a planar hexamer assembly so a convex pentamer can form at each vertex, assemblies of the CPs. Pentamers are inserted into a hexagonal sequence so a closed shell can form with the proper symmetry (*2*). Since the same structural component is used in both subassemblies the interactions are not equivalent. Unlike the T = 1 capsids the subunits of larger viruses are not identical and instead exhibit differences in subunit interactions and conformations, a property known as quasiequivalence (*3*). The TCV capsid is comprised of 180 CP subunits and has T = 3 symmetry.

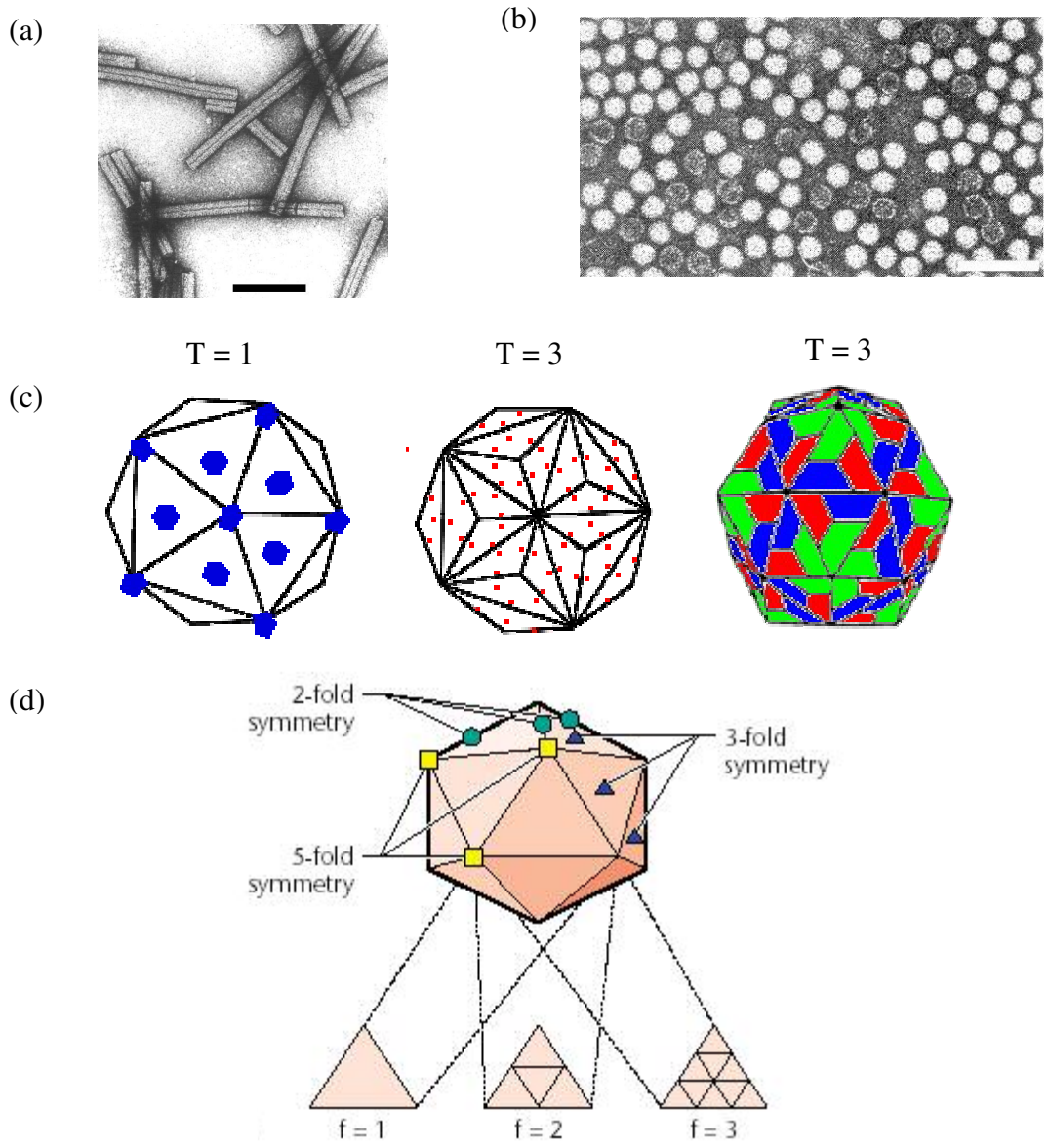

Figure 1. Helical and icosahedral symmetry. (a) Helical TMV virions negatively stained. The bar represents 100 nm . From http://www.dpvweb.net (b) Icosahedral TCV virions stained with potassium phosphotungstate. The bar represents 100 nm. From http://www.dpvweb.net (c) The structures of T = 1 and T = 3 icosahedral virions with the three different coat protein subunits (A, B, C) per face in the T = 3 virion indicated by varying color. Green represents the A subunit, red the B subunit and blue the C subunit. From http://www.tulane.edu/~dmsander/WWW/224/Structure224.html (d) Icosahedral symmetry, with axes of 2, 3, and 5-fold symmetry indicated (*1*).

2. Coat protein structure

In 1986 the x-ray crystallographic structure of TCV was resolved to 3.2Å resolution and was found to be very similar to that of *Tomato bushy stunt virus* (TBSV) another T = 3 icosahedral virus (*4*). The similarity in structure of these two viruses was first reported in earlier work using low resolution x-ray diffraction and electron microscopy (*5*). Once the structure was resolved at high resolution it was observed to contain similar elements to TBSV both with respect to the complete virion and the individual coat protein subunits (*4*). Each face of the icosahedral virion contains three CP subunits that interact with each other and the internally packaged genomic RNA. The structure of these subunits is similar between viruses with specific repeating structural elements or motifs. In the case of TCV the CP contains three main domains, the N-terminal random (R) domain, the shell (S) domain and the protruding (P) domain (see Figure 2). The R domain has a basic N-terminus that can extend into the interior of the virus particle where it can interact with the positive-sense RNA. It is not normally visible in x-ray crystallography due to its conformational flexibility. This domain is linked by an "arm" region to the S domain, which consists of an eight-stranded antiparallel β-sheet or β-barrel jellyroll, with the inserts between these strands pointed outward, forming the virion shell (*4*). There are a limited number of different tertiary structural motifs found in the CPs of different viruses but the β-barrel is the most common. It is not only found in plant but also vertebrate, insect and bacterial viruses (*2*). The P domain, which protrudes out from the virion's surface is connected to the S domain via a hinge (*4*). The only domain that is conserved in all plant RNA viral CPs is the shell (S) domain. The presence

of the protruding (P) domain is a feature of the *Carmovirus* genus in the *Tombusviridae* family (*6*).

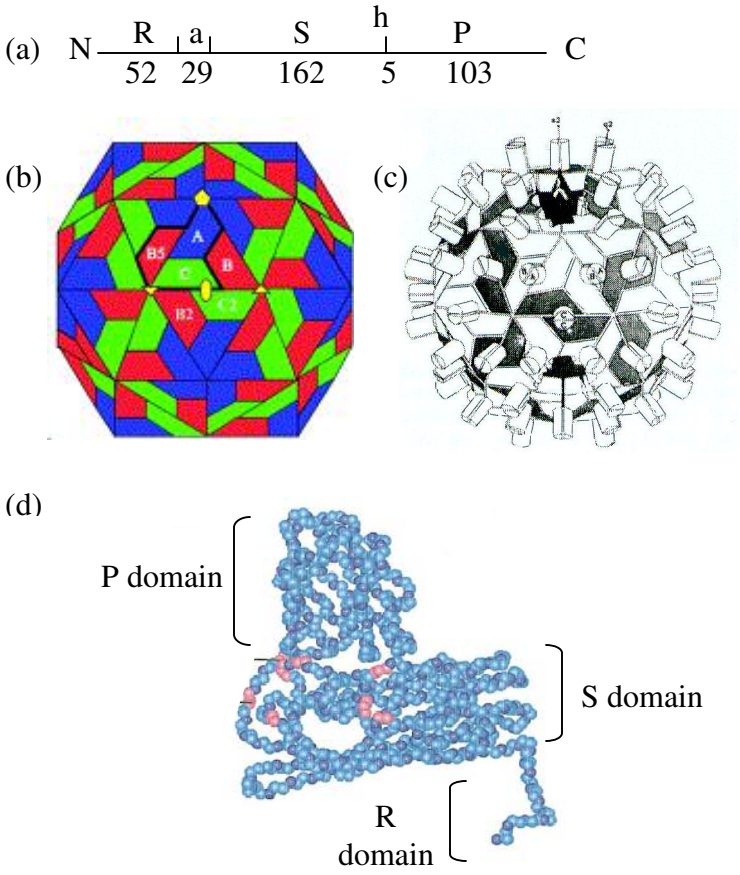

Figure 2. TCV virion and coat protein structure. (a) Linear representation of the order of the domains in the CP from the N-terminus to the C-terminus. (b) Representation of a T = 3 virion with the three different CP subunits indicated by color. The oval represents the two-fold axis, the triangle the three-fold axis and the pentagon the five fold axis (*7*). (c) TCV virion indicating the A, B and C packaging environments of the CP (*8*). (d) A space-filling representation of the carbon-nitrogen backbone of the TCV CP subunit (*9*).

As indicated above, viruses like TCV with larger triangulation numbers exhibit quasiequivalence in their coat protein subunit interactions. These domains can exist in different conformations depending on the packaging environment. There are three distinct CP packaging conformations present: A, B and C, which together form each face

on the capsid surface (see Figure 2b). Conformations A and B have similar S and P domain orientations with the R domain arms in a disordered state. The C conformation has a different hinge angle, resulting in slightly different relative S and P domain orientations, with the R domain arm ordered. From the 180 subunits that comprise the virion structure 120 of these are A and B subunits while the remaining 60 subunits are in the C conformation. Each ordered arm from one C subunit interacts with arms from two other C subunits to form a β-annulus structure that interacts with the genomic RNA at the 3-fold axis. In this manner the C subunits are thought to dictate the size and the T = 3 structure of the virion. Each subunit dimer's P domains form the protruding structures for which they are named, producing a total of 90 protruding structures for each virion (*8*).

3. RNA and replication

Icosahedral viruses can contain RNA or DNA, either single-stranded or double-stranded, although most icosahedral viruses contain single-stranded RNA. Even with this variation in the type/ form of the genetic material, they all contain genomes of a predetermined size. With single-stranded RNA viruses once the virion enters the host cell it dissociates allowing the immediate translation of the viral-RNA dependent RNA polymerase (RdRp) directly from the positive-sense RNA genome. This enzyme can then transcribe the negative-sense RNA template needed to produce more positive-sense RNA for packaging and the other subgenomic RNAs. Those RNA genomes that encode more than one gene have the first gene or open reading frame directly translated from the genome while the other downstream genes are expressed off of subgenomic RNAs (*10*).

In many cases the genome and subgenomic RNAs are divided into a series of open reading frames that each encode a particular required protein. Translation can then occur producing the appropriate proteins needed for assembly, movement through the host and other required functions.

TCV, like many icosahedral viruses, contains one copy of a positive-sense 4054 base single-stranded RNA genome that is encapsulated by 180 copies of the 38 kDa coat protein (TCV CP) in the virion structure (*11*). The RNA consists of five open reading frames (ORFs), encoding the five viral proteins, expressed from the genomic and two subgenomic RNAs, sgRNA1 and sgRNA2 (see Figure 3) (*12*). Both p28 and p88 are directly translated from the genomic RNA. Translation of p88 is by read-through of the p28 amber termination codon, and the two protein products comprise the viral RNA-dependent RNA-polymerase (RdRp) needed for replication of the genome and two subgenomic RNAs (*13*). The movement proteins p8 and p9 are translated from overlapping ORFs of the 1721 base sgRNA1 (*14*). Both are implicated in cell-to-cell and systematic spread of the virus *in vivo* (*13*). The 1447 base sgRNA2 codes for the p38 or CP which has been implicated in: (1) virion assembly (2) movement (*13*) (3) repression of satC replication (*15*) (4) suppression of gene silencing (*16, 17*) and (5) symptom modulation in the presence of satC (*18*).

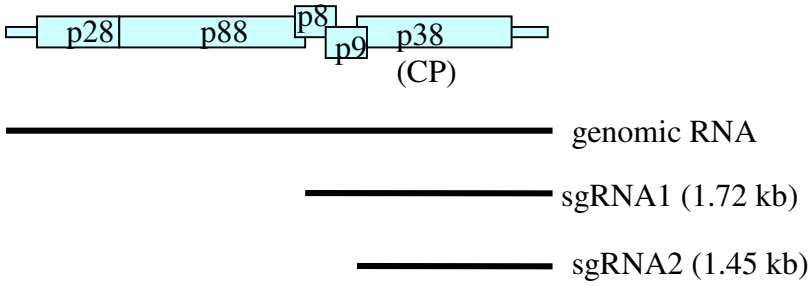

Figure 3. A representation of the coding regions on the TCV genomic RNA that encode for the various viral proteins associated with TCV. The genomic and two subgenomic RNAs are also depicted above.

C. Models of Virion Assembly

A coat protein's main function is assembly of the virion into its appropriate symmetrical structure. The genomic RNA must be packaged and surrounded by an exterior shell to protect the genome from potential degradation outside the host organism. For viruses that do not have an outer lipid envelope encapsidation results in the formation of the complete virion. Since the coat proteins form the capsid they must be able to first recognize the correct nucleic acid sequence or viral genome and package it during the appropriate time frame of the infection process. If packaging occurs too early the genome will be sequestered before it has a chance to act as the template for replication. In order to ensure proper timing of encapsidation the coat protein is translated from a subgenomic RNA, as described previously (*2, 10*). This allows for a pool of new genetic information to be established and ready for encapsidation once the coat protein is produced. Not only must the coat protein be translated after genome replication, once ready for packaging it must also be able to recognize the genomic RNA, bind and initiate assembly. As a

method of identification the protein or protein complex must be able to recognize a particular sequence, termed the nucleation site, on the genome to which it can bind and initiate packaging. Once bound, the protein(s) recruit other coat proteins that then form protein-protein interactions with each other and protein-nucleic acid interactions with the genome (*1, 2*). These latter protein-nucleic acid interactions may be less sequence specific than those formed with the nucleation site for protein binding and instead are a result of the outer folded geometry of the genome. There are different models for how virions assemble and at what point during the process the genome attains its folded conformation.

1. *Tobacco Mosaic Virus* (TMV)

The best-known assembly model and characterized nucleation site is for the helical *Tobacco mosaic virus*. This virion is a right-handed single helix with a 6395 nucleotide RNA genome intercalated between successive turns of a 300nm long 18nm wide helix with an interior space of 2nm. It consists of 131 turns with 16 1/3 subunits per helical turn for a total incorporation of 2130 coat protein subunits (see Figure 4) (*19*).

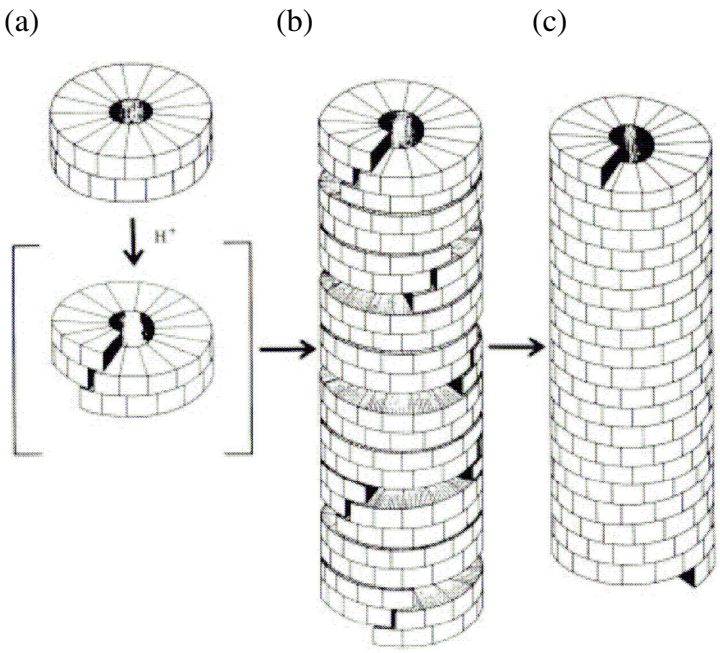

Figure 4. A schematic representation of the different structures produced by TMV during assembly. (a) Depicts the two ring aggregate, (b) the procapsid and (c) the complete TMV particle (*19*).

The nucleation site for assembly of TMV has been identified and characterized through assembly and site-directed mutagenesis studies. The origin of assembly was originally identified through nuclease protection studies (*20, 21*) and is located at about 1000 nucleotides from the 3' end of the genome from nucleotide 5420 to 5546 (see Figure 5) (*22, 23*). This core region forms a stem-loop structure and it was observed that *in vivo*, while coexpressing the TMV coat protein, this region alone was sufficient to allow encapsidation of foreign RNAs (*24, 25*). A series of studies were then carried out in which particular sites in the stem-loop sequence were mutated and the rate of assembly was measured *in vitro*. When the AAGAAGUCG sequence that forms the loop in the stem-loop structure was replaced by either a $(CCG)_3$ or an $(A)_9$ sequence the rate of

assembly decreased by 750 and 3000 fold, respectively. By contrast, a replacement sequence of either $(UUG)_3$ or $(GUG)_3$ where the G was kept in every third position did not have a significant effect on the assembly rate. If the G was either removed or C residues were inserted into the sequence the rate of assembly decreased. It was determined that the triplet G repeat was required in the stem-loop while C residues were unfavorable. To determine if the position of the loop in the genome was an important factor in initiating assembly the stem sequence and stem length were varied. It was found that the overall sequence was unimportant but that the stem length was the determining factor. If either the stem was shortened or the bottom loop was removed the rate of assembly decreased (*26*). The main features required for a nucleation site comparable to wild-type TMV that allowed assembly included a single-stranded loop sequence with a G repeat at every third base, a weakly base paired stem and the correct distance of the lower stem and loop region (*19*).

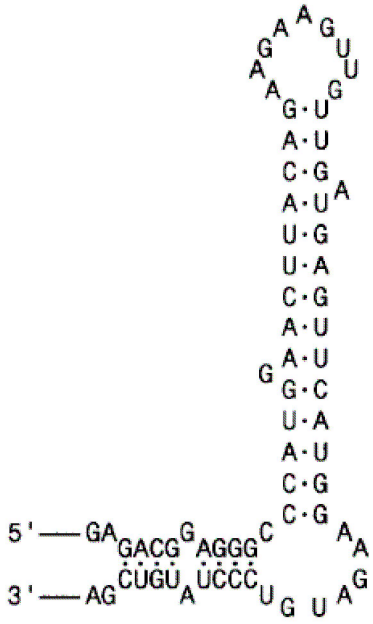

Figure 5. This is the origin of assembly on the TMV RNA(*19*).

In 1984 Butler proposed the model for the nucleation of TMV assembly. It had been observed in earlier studies that a large aggregate or “disk” of the coat protein could form which consisted of two rings of 17 subunits each. The rate of assembly increased with the addition of these disk aggregates compared to assembly studies with a mixture of smaller aggregates termed A-proteins (*19*). The helix was also formed in the absence of RNA at a low pH and by comparing the protein helix to the nucleoprotein helix the structure of the interior genomic RNA was resolved (*27, 28*). From the RNA structure it was determined that each CP bound to a three-nucleotide region on the genomic RNA (*28*). Butler’s proposed model of assembly consisted of a nucleation reaction step and an elongation reaction step (see Figure 6). During the nucleation reaction the nucleation site first inserted itself into the disk and bound causing the weakly base paired stem to melt and generate more single-stranded RNA, which was then available to bind to neighboring subunits until a complete turn was bound. The energy from this interaction allowed the assembly to transition from a disk to a proto-helix structure. This rearranged the protein subunits so the RNA could become trapped between the two layers of the rings allowing the RNA to bind to both the top and bottom surfaces of each protein layer. At this point elongation or the addition of more coat protein subunits proceeded in both directions. The elongation occurred at a faster rate toward the 5' end since disk aggregates were added onto the traveling loop of RNA. The 3' end elongation proceeds at a slower rate since smaller aggregates or A-proteins were added on (*29*).

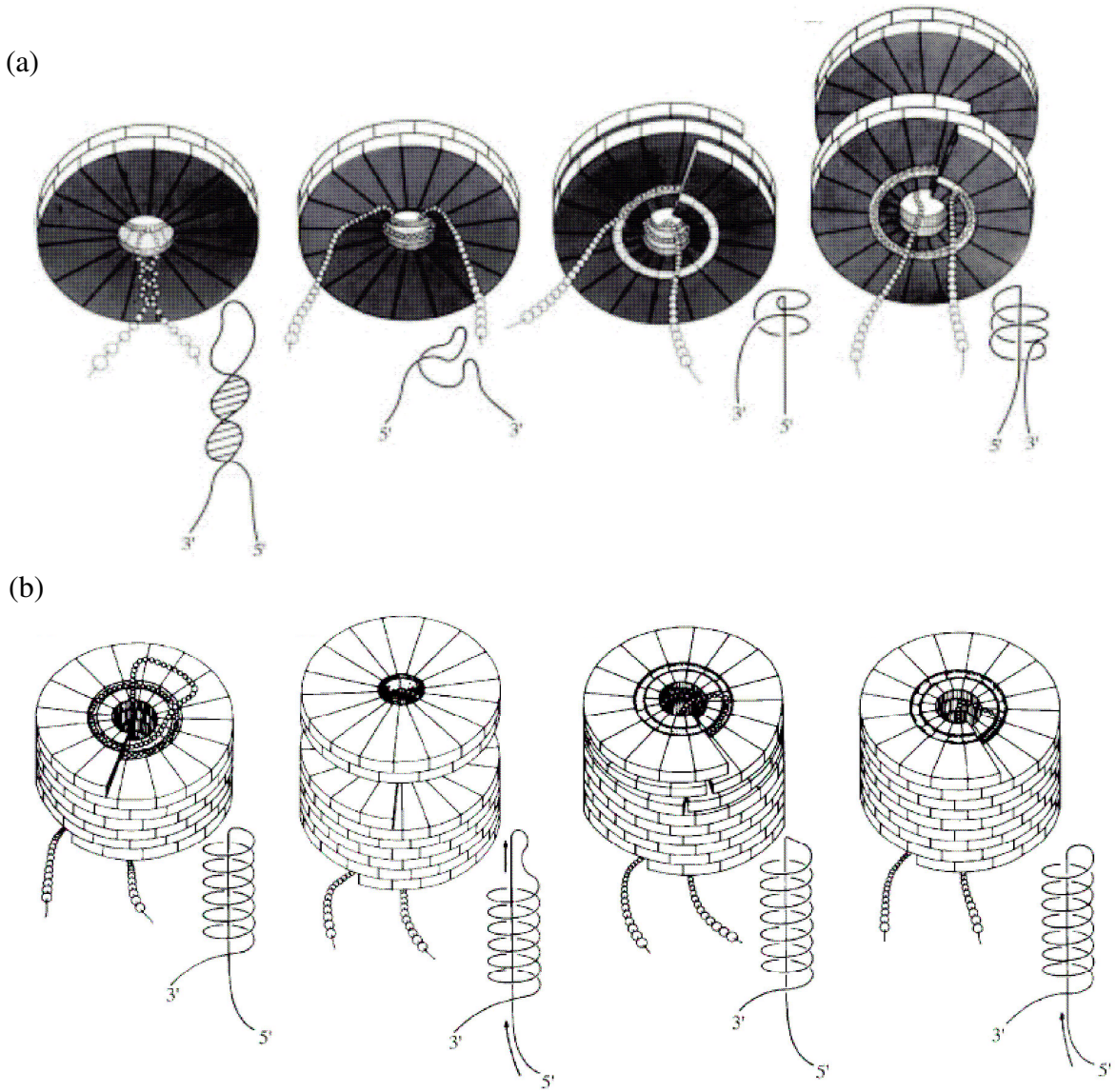

Figure 6. A picture of the proposed TMV assembly process which includes (a) the mechanism of nucleation and (b) the mechanism of elongation (*19, 29*).

2. Icosahedral Virus Assembly

Unlike the well-documented helical virus model of assembly the mechanism of assembly of icosahedral viruses is still not well understood. The proposed model for TMV assembly with its characterized nucleation site has influenced subsequent assembly studies performed with icosahedral viruses. The concept of a nucleation site that must bind CP in order to initiate selective packaging of the viral RNA was adopted as a universal nucleation step for virus assembly. It has consequently been incorporated into various proposed models for assembly of icosahedral viruses. There are two main current assembly models, either the coat protein binds to the nucleation site on the RNA initiating assembly or a procapsid structure is formed which then binds to the genome and draws it into the interior through a gap in the capsid. The second model is observed specifically with DNA viruses such as some bacteriophages while the first model is the more accepted for RNA icosahedral assembly although even this model consists of two possible pathways based on the folding of the RNA (*1*). Either, as proposed for *Cowpea chlorotic mottle virus*, the coat protein or a complex of coat proteins can bind to the RNA and initiate a RNA conformational change or as suggested for *Bean pod mottle virus* and nodavirus, the RNA folds prior to the initial CP binding and the geometry of the RNA influences the coat protein assembly (*7, 30*).

a. *Cowpea Chlorotic Mottle Virus* (CCMV)

One example of a single-stranded positive sense icosahedral virus that has been subjected to assembly studies is the *Bromovirus* CCMV. It contains a tripartite genome and three morphologically identical virions each contain essential genetic information needed for the infection process. The RNA 1 and RNA 2, which each code for proteins needed during replication are packaged in separate virions while RNA 3, the mRNA for the 32 kDa movement protein and RNA 4, a subgenomic mRNA for the 20 kDa coat protein expressed from RNA3, are packaged in a 1:1 ratio in another virion. Since the coat protein can form empty capsids under low pH and moderate ionic strength conditions it has been a useful model for assembly (*30*). In 1995 Johnson determined the virion structure to 3.3 Å resolution by x-ray crystallography (*31*). This crystal structure revealed the presence of a tubular β-barrel structure, termed the β-hexamer, where the N-termini, amino acids 29 to 33, of the three-fold coat protein subunits converged at the quasi-six fold vertices (see Figure 7). From the virion structure it was proposed that the basic assembly unit would consist of a noncovalent dimer. The presence of the β-hexamer and the noncovalent dimer suggested that assembly might be initiated by a hexamer of dimers where the free N-terminus from the three-fold coat protein subunit of each dimer interacts to form the β-hexamer which can then act as a stability factor (*31*). In 2000 Zlotnick, using light scattering and size-exclusion chromatography techniques combined with assembly experiments determined that the hexamer was most likely not the species of initiation. They observed that not only could the coat protein form pentamers of dimers readily in solution but that capsid assembly was nucleated by a pentamer of coat protein dimers. It was then speculated that assembly could progress by

cooperative addition of coat protein subunit dimers (*32*). Once the β-hexamer was determined not to initiate assembly its influence on the assembly process was then investigated. A series of mutational studies in conjunction with assembly assays both *in vitro* and *in vivo* concluded that the structure was not necessary for virion formation but does contribute to overall virion stability (*33*).

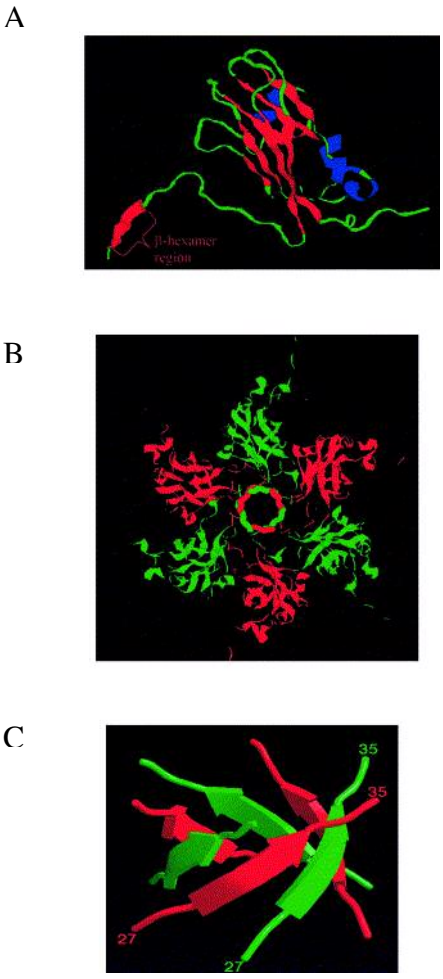

Figure 7. Structural features of CCMV. (A) Diagram of the CCMV CP. The β-strand involved in the β-hexamer formation is labeled. (B) Depicts the hexamer viewed down the three-fold axis. (C) Representation of the β-hexamer motif (*33*).

A new model was proposed in 2004 based on *in vitro* measurements of assembly of the CCMV RNA-containing virus-like particles (VLPs; see Figure 8) (*30*). In the assembly experiments two different time-dependent pathways were observed. For those experiments that lasted 20 minutes a fast pathway was observed while a slow pathway was measured for longer incubation times (20-24 hours). The fast pathway went directly to assembled virions with a ratio of 90 homodimers of CP to one RNA molecule. In contrast, the slow pathway first formed a nucleoprotein structure, termed the C1 complex, in the presence of 10 CP dimers to one RNA molecule. The actual stoichiometry and structure of this complex had not been investigated and is still not known. Once the C1 complex was formed it was then followed by completion of the capsid shell. Folding of the RNA was observed to be independent of its own sequence but dependent on coat protein binding. The formation of the C1 complex was only observed in the slow pathway with gradual development over a six to eight hour period. Based on this data a model was proposed where coat protein first binds to the RNA with low cooperativity in the presence of excess RNA allowing the protein to find specific binding sites. The coat protein and RNA then fold to form the C1 complex although this is not observed via gel electrophoresis in the fast pathway due to the kinetics involved. Once the C1 complex is formed additional coat protein dimers can then be incorporated into the complex with high cooperativity (*30*). An initial site of nucleation has still not been identified for this virus although evidence does suggest that it is initiated with a pentamer of dimers of coat protein subunits instead of a hexameric structure (*32, 33*).

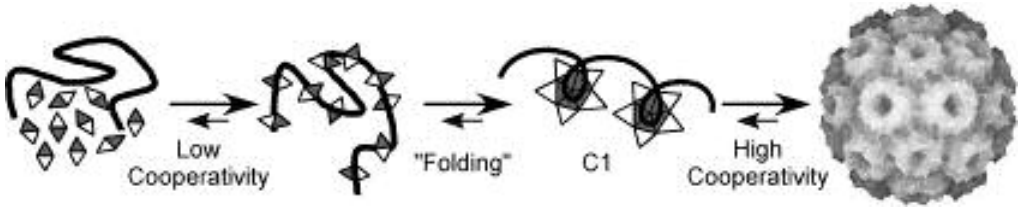

Figure 8. Proposed assembly path for CCMV. (1) CP initially binds excess RNA with low cooperativity allowing the CP to search for high-affinity binding sites. (2) CP and RNA slowly fold to form the C1 nucleoprotein complex which serves as a marker for RNA encapsidation. The complex forms in the presence of 10 CP: 1 RNA molecule although the actual stoichiometry and structure of the C1 complex is not known. (3) Addition of CPs to C1 is highly cooperative yielding complete virions (*30*).

b. *Bean Pod Mottle Virus* (BPMV)

In contrast to CCMV, which is thought to initiate packaging by specific binding of a pentamer of coat proteins, crystallographic evidence of BPMV suggests an alternative mechanism. This virus belongs to the *Comovirus* family and contains a single-stranded bipartite RNA genome where each RNA is encapsidated separately. Its capsid has pseudo T = 3 (P = 3) symmetry and is not a true icosahedron. The ability of the virus to form empty capsids and two different virions allowed for the comparison of contributing factors for assembly and stability. It was observed that 20% of the RNA packaged was structured while the remaining RNA on the interior of the virion not interacting with the coat protein was unstructured. From analysis of the electron density of both the empty capsids and complete virions it was determined that the coat proteins interact with a repeating sequence in the RNA. This sequence consists of six ribonucleotides: adenine, purine, pyrimidine, pyrimidine, pyrimidine, and X. Where X was defined as the base electron density at nucleotide position six consisting of only enough space for a five-membered ring. This sequence was found in both genomic RNAs

with a pyrimidine at nucleotide position six. Evidence suggests that the interactions between the bases and the amino acids are mediated by water molecules (see Figure 9).

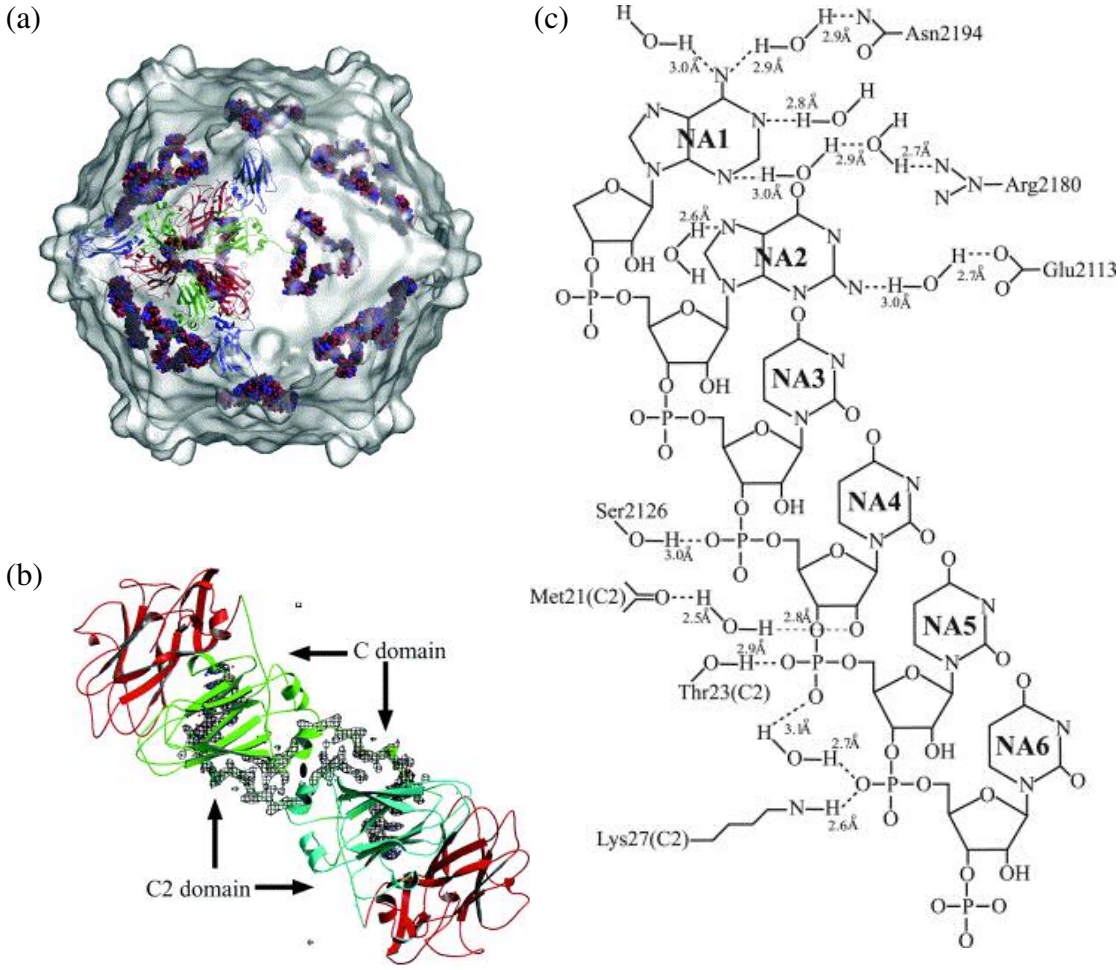

Figure 9. *Bean pod mottle virus* coat protein and RNA interactions. (a) Depicts trefoils of the genomic RNA within the capsid superimposed onto the electron density viewed from the exterior. (b) Represents the CP interacting with the RNA at the two-fold axis. The arrows indicate the CP and their N-termini in relation to the electron density of the RNA. (c) Interactions of the ordered RNA with the protein capsid (*7*).

The structure of the ordered genomic RNA for both virions was observed to be nearly identical indicating that both genomes contained the same RNA packaging signals. Since the CP can assemble without RNA the nucleic acid is not essential but the repeating sequence allows the genome to be targeted for assembly. Although it was determined that both genomes contain the same repeating sequence the actual RNA folding process is still unknown. According to a model proposed by Johnson the RNA first folds loosely obtaining an icosahedral geometry at which point the coat protein subunits, possibly pentamers, then assemble around it (*7*).

3. TCV Assembly Model

Like CCMV, TCV has been used as model for studying icosahedral virion assembly. As a structural protein, the CP's main function is assembly of the virion or encapsidation. The virus has been dissociated and reassembled *in vitro* and the process is most effective when assembling with the genomic RNA or RNA of an equivalent size (*34, 35*). TCV incubated at pH 8.5 and 0.5 M salt dissociates into dimers and a ribonucleoprotein complex or RP complex, consisting of RNA and protein (*8, 34*). In 1986 Harrison obtained the x-ray crystal structure at 3.2Å resolution and performed reassembly experiments *in vitro* from dissociated virions. From both of these studies a model of assembly was proposed although the evidence to support such a model is limited (*4, 34*).

According to the proposed model by Harrison (1986), the assembly process begins with the formation of a trimer of dimers that are connected by a β-annulus structure in which the three ordered arms from the C subunits overlap. This complex was thought to form at a particular site on the TCV genomic RNA called the nucleation site (see Figure 10) (*34*). Interaction between the N-terminal R domain and the genomic RNA at the nucleation site, would facilitate the assembly process *in vitro*. Once a trimer of dimers is formed on the RNA the next dimer to add onto the complex adopts an A/B conformation with the existing CPs. The arrangement of these additional subunits prevents their R domain "arms" from folding and contributing to the established β-annulus. The formation of contacts between B and C faces lock in the extended arms from the complex around the β-annulus. The overall assembly process consists of: (1) β-annulus formation by three dimers with C/C conformation and (2) dimers added onto the complex with A/B conformations. The "arm" between the S and R domains allows the CP subunits to arrange themselves into the proper configurations to form the capsid structure. Once the capsid starts to assemble more dimers are added cooperatively to complete the virus (*34*).

This model was based on low resolution images of both the x-ray crystal structure of TCV and negatively stained samples from the *in vitro* reassembly experiments. The nature of the RP complex was proposed based on its size from gel-filtration of dissociated virions and the capsid structure from x-ray crystallography (*4, 34*). The presence of an RP complex was observed after gel-filtration, when samples were run on SDS/polyacrylamide gels. The RNA samples had a high molecular weight band that

would not load onto the gel but after ribonuclease A digestion the band was eliminated and a lower 38kDa band corresponding to CP was observed. This suggested that the observed high molecular weight band consisted of protein undissociated from the genomic RNA. From the intensities of the bands it was suggested that the composition of the RP complex was between five to six CP subunits per RNA molecule (*34*). The structure of the RP complex was not subjected to further characterization. The structure of the RP complex proposed in the model was based on the x-ray crystal structure arrangement of the CPs on each face of the virion. It was proposed that the C subunit "arms" would interlock to form a β-annulus structure since it was previously known that the C conformational subunits were the only CPs with their "arms" in an ordered state and therefore the only subunits that could interact with the RNA. The actual β-annulus structure has not been observed experimentally or in the crystal structure (*4, 34*). Since the *in vitro* reassembly experiment images supported the idea that the CP forms the capsid around the RNA then the first step proposed in the model was binding of a CP subassembly to the nucleation site on the genomic RNA. From the point of nucleation to completion of the capsid there is no experimental evidence from the *in vitro* reassembly experiments to support Harrison's model (*34*). The sequential addition of the CP subunits described in the model is based primarily on the geometry of the final capsid itself (*4*). In summary, although a model has been proposed for TCV the only known evidence is that the CP binds to the RNA and then forms the capsid around it. The mechanism of the assembly process, including protein-RNA and protein-protein interactions has not been investigated further and is still unknown. Further experimentation needs to be carried out to better define the model of TCV assembly.

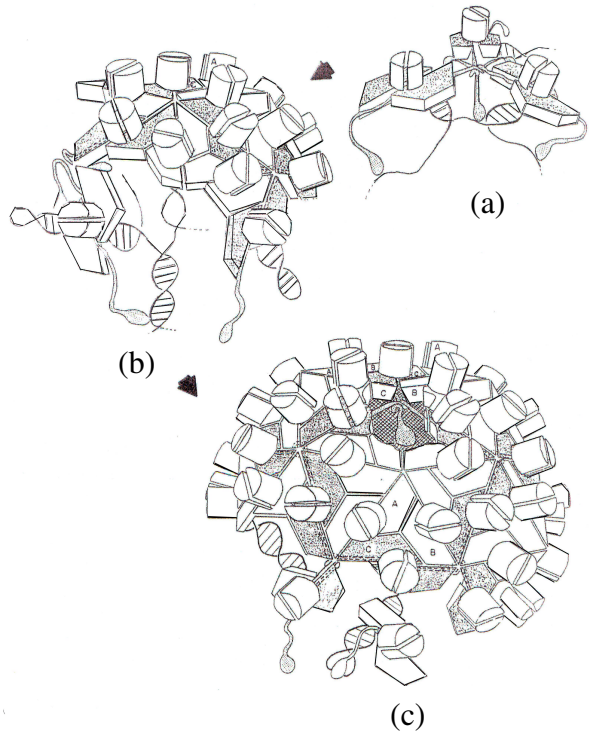

Figure 10. Assembly model of the TCV virion from re-assembled virions *in vitro*. (a) Formation of the RP complex consisting of a 6 CPs connected by the β-annulus structure with the genomic RNA. (b) & (c) Continuation of assembly as described by the Harrison (1986) model (*34*).

a. Initiation of TCV Assembly – Nucleation site

In order to initiate virion assembly the CP must first bind to a particular region on the genomic RNA, called the nucleation site. This region has been identified and characterized for other plant viruses. One classic example is that of *Tobacco mosaic virus* (see Figure 4) (*19*). According to Harrison's (1986) TCV model the CP initiates assembly by binding to the nucleation site and forming the RP complex. Once bound to this region assembly can progress to completion (*34*).

In 1990 the RP complex was isolated and digested with both RNase A and RNase T_1. Five fragments, Fa to Ff, from the TCV RNA genome protected from the digestion were isolated and identified. The fragments clustered in two regions: Ff, Fa and Fc were found in a ~400 nt region that included the leaky stop codon in the polymerase gene while Fe and Fd were located in the CP gene near the 3' end (*36*). Morris (1995), performed binding studies with the 386 nt region, containing Ff, Fa and Fc, of the genomic RNA and determined that the CP had no specificity for this region of the RNA (*37*). The fragments Fe and Fd were later tested using a protoplast infection assay in which mutant viral RNA assembly was examined. A region of 186 nt, containing a part of Fd, near the 3' end of the CP gene was found to be important for packaging. This region contained a 28 nt bulged hairpin loop, nt 3755-3791, that was essential for virion formation (see Figure 11) (*35*). Although this region has been identified, there have been no binding measurements published on the proposed interaction between this nucleation site and the viral CP.

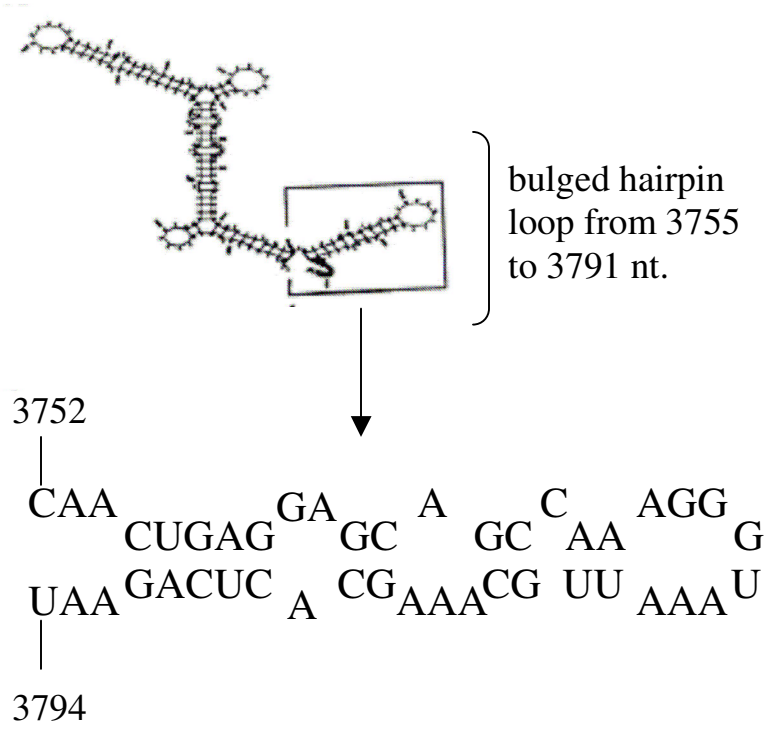

Figure 11. This is a representation of the proposed nucleation site on the TCV genome (*35*).

D. Introduction into the plant – virus-induced gene silencing (VIGS)

The level of assembly of the TCV virions has been correlated with symptom modulation. As the virion formation decreases in the cell there is an observed increase in the plant's symptoms and a decrease in the gene silencing (*38, 39*). It has been indicated that even though the TCV CP is the main structural component of the assembled virion it also can contribute toward other functions including inhibition of gene silencing in the plant (*16*).

RNA-induced gene silencing is a RNA degradation mechanism, which acts as a host response to the presence of double-stranded foreign RNA (see Figure 12) (*40*). It has been found to occur in animals, fungi and plants and is termed accordingly, RNA-interference or RNAi, quelling, and post-transcriptional gene silencing or PTGS. The process can be triggered by the introduction of dsRNA, viral RNA, and/or single-stranded hairpin RNA into the plant cell. In the case of an infecting virus, TCV, its viral genomic RNA secondary stucture, dsRNA, triggers the RNA-induced gene silencing mechanism also known as virus induced gene silencing (VIGS). Once the dsRNA is introduced into the plant it is degraded into fragments 21-24 nucleotides in length called small interfering RNAs or siRNAs. During this step dsRNAs are cleaved by DICER or a DICER-LIKE homologue (DCL), a ribonuclease III-like enzyme which is ATP dependent (*40*). Dicer was originally discovered as the enzyme involved in RNAi in *Drosophila* (*41*). The specific RNase responsible for siRNA production in plants has not yet been identified, although there are four known DCLs in *Arabidopsis* (*40*). In *Arabidopsis thaliana*, DCL2 participates in viral siRNA production (*42*). In the case of Dicer, the enzyme is a dimer when bound to the RNA allowing it to cleave on either side of its position on the dsRNA (*43*). The siRNA is then incorporated into the RNA-induced silencing complex known as RISC, consisting of both Dicer/DCL and the siRNA. In the presence of ATP RISC becomes activated and unwinds the double-stranded siRNA associated with the complex. Using a single-strand of the siRNA it can detect and recognize the viral RNAs complementary to its sequence, at which point the complex can degrade additional RNAs. At this point, the degradation process is independent of the dsRNA that initiated the gene silencing (*40*).

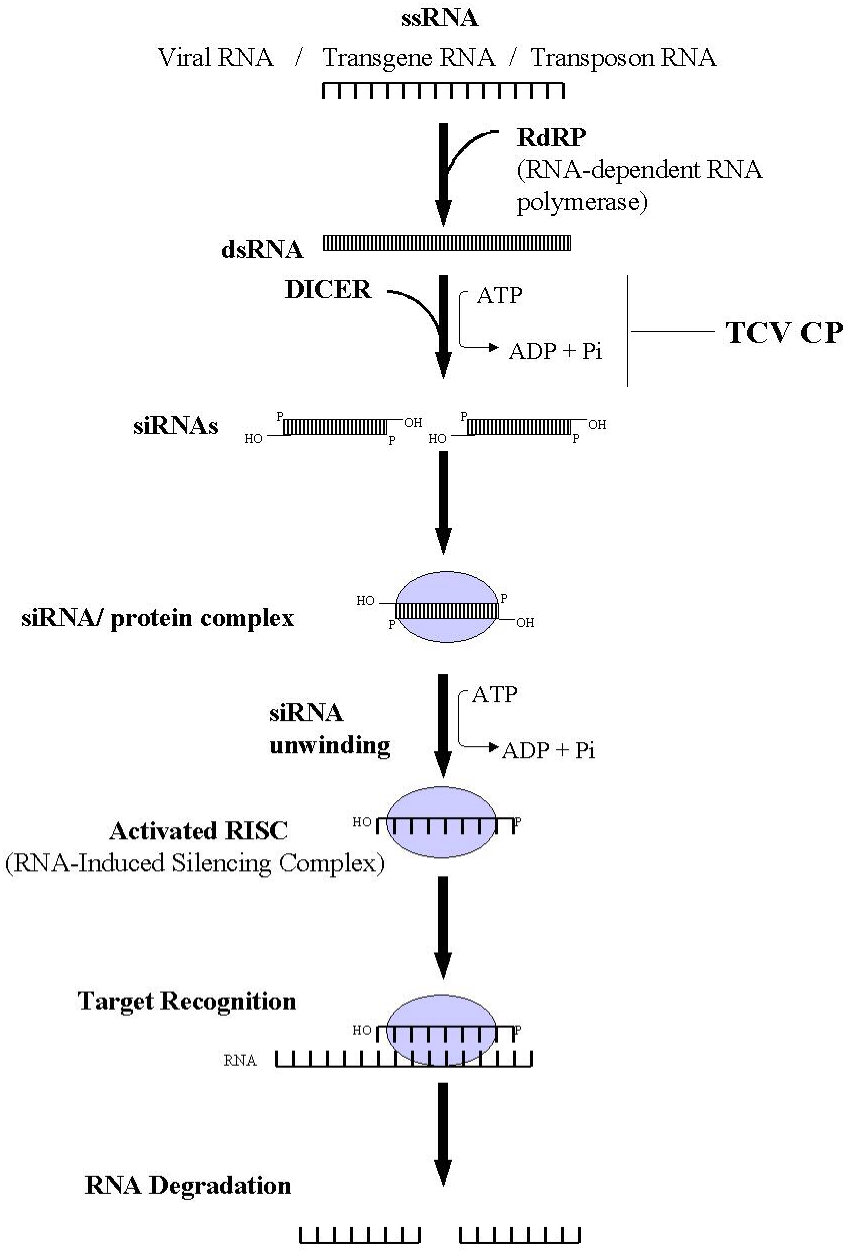

Figure 12. Schematic representation of the RNA-induced gene silencing mechanism in plants, indicating at which point TCV CP acts as a suppressor of the pathway (*16, 40*).

1. The viral counter-defense mechanism – TCV CP as a suppressor of VIGS

Since plants have mechanisms against viral infection it is only natural that viruses have evolved counter-defense mechanisms. There have been a number of studies performed in the last decade describing viral proteins that exhibit suppression activity against the silencing pathway. Each case is unique in that each identified protein targets a different step in the mechanism (see Figure 13). The *Cucumber mosaic virus* (CMV) 2b protein was one of the first to be studied and it was found to block the movement of the silencing signal, although how it can inhibit systemic silencing is still in question (*40*). The *Potato virus X* (PVX) p25 protein required for cell-to-cell movement has been implicated in an initiation step possibly interfering with the cellular RdRp so that the single-stranded RNA cannot be converted to double-stranded RNA (*44*). Another example is the helper-protease component (HC-Pro) of the *Tobacco etch virus* (TEV), which is thought to interfere in a maintenance step blocking the production of siRNAs of a certain size (*45, 46*). Although these suppressors have been identified their mode of action is still not well understood from a mechanistic viewpoint. One exception is the *tombusviruses* p19 protein that was found to bind to siRNAs *in vitro*, suggesting that the protein was binding and sequestering the siRNAs from RISC (*47*). In 2003 the crystal structure of the p19 from *Carnation Italian ringspot virus* (CIRV) bound to a 21 nucleotide siRNA was reported to 2.5 Å resolution (*48*). This structure supported the idea that the protein acts to sequester specific siRNAs but it is still unclear if this function would occur *in vivo*.

In recent studies, it has been observed that the TCV CP can also act as a suppressor of VIGS. It has been found that the TCV CP interferes early in the initiation step when Dicer degrades the dsRNA into siRNAs (see Figure 12). The exact mechanism of this suppression is not known (*16*). In *Arabidopsis thaliana*, the CP has been proposed to interact with the DCL2 but not DCL1 enzymes (*42, 49*). It has been speculated that the N-terminal 25 amino acids of the coat protein R domain may contribute to the suppression activity of VIGS (*17*). This region interacts with the genomic RNA during assembly and may have an effect on dicer's function. This suggests that it is the unassembled CP with free N-termini that participates in the suppression of VIGS.

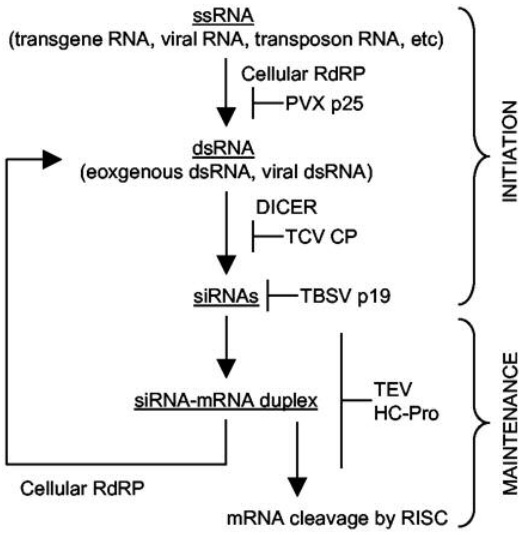

Figure 13. This schematic representation of the VIGS pathway depicts the steps at which each suppressor is proposed to interfere (*16*).

2. TCV Subviral RNAs

The TCV CP can act as a suppressor of gene silencing in the plant only under particular conditions. Suppression has been observed when there is free CP in the cell not being used in the assembly pathway. The levels of free CP in the plant can be increased either by *Agrobacterium* co-infiltration or the presence of subviral RNAs, specifically satellite RNA C (*39, 40*).

Viruses can be associated with subviral RNAs, which consist of defective interfering RNAs (DI RNAs) and satellite RNAs (satRNAs). While DI RNAs are common among animal viruses but rarely found with plant viruses, satRNAs are found specifically with plant viruses. DI RNAs are generated via a sequence of errors during replication and as a result can contain a high degree of sequence similarity with the helper virus. By contrast, satRNAs usually share little if any sequence similarity to the helper virus (*50*). An exception to this is the production of satC in the TCV infection process (*38*).

During replication, satC is produced via a recombination event between satD and two discontinuous segments of the TCV genome. SatC is a chimeric molecule with 189 bases of satD at its 5' end and the two TCV segments, totaling 166 bases, at its 3' end (see Figure 14). Since it shares the same 3' sequence with the genomic RNA it can also be replicated by the viral RNA-dependent RNA polymerase (RdRp) (*38*).

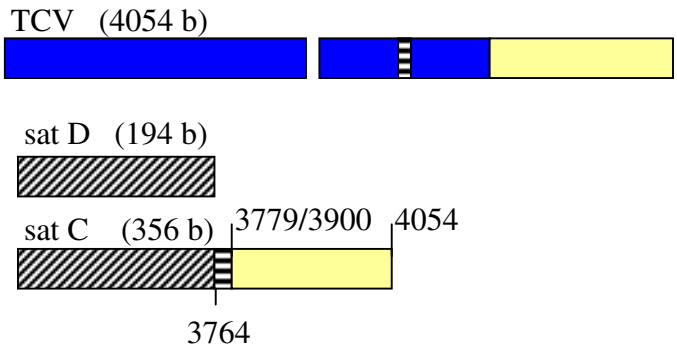

Figure 14. Depicts the TCV genomic RNA relative to both satD and satC. The satC is a chimeric molecule consisting of satD at the 5' end and two segments of the TCV genomic RNA (#3764-3779 & #3900-4045) at the 3' end.

3. TCV SatC & symptom modulation

Subviral RNAs can participate in symptom modulation, either by attenuation or exacerbation of symptoms. The addition of satC in the TCV infection has been observed to cause (1) a reduction in TCV virion accumulation (2) exacerbation of symptoms in the plant and (3) reduced gene silencing (see Figures 15 & 16) (*38, 39*).

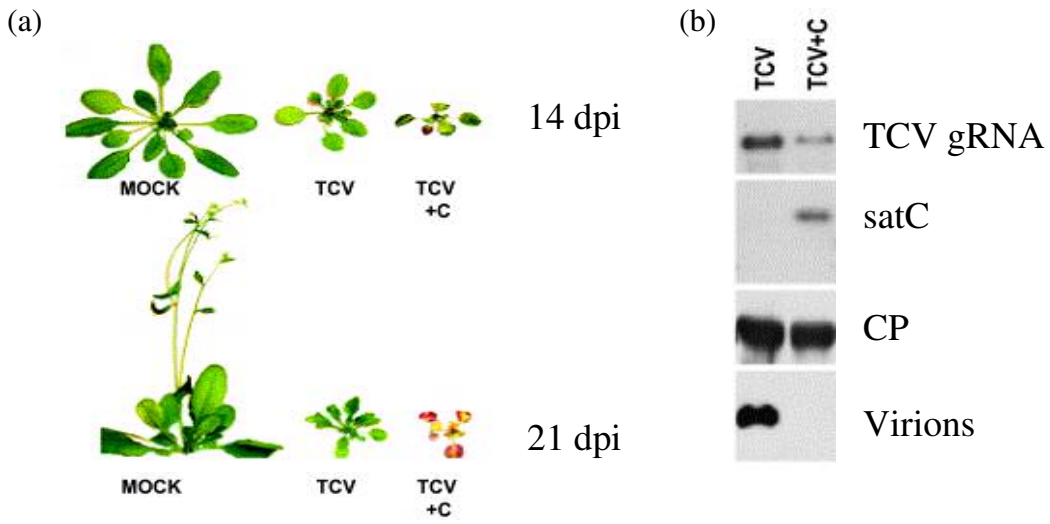

Figure 15. (a) Symptoms of *Arabidopsis* plants inoculated with TCV with (TCV + C) and without satC. The plants in the top row were photographed at 14 days postinoculation (dpi) and those in the bottom row were photographed at 21 dpi. Those plants labeled "Mock" were only treated with infection buffer alone. The plants infected with TCV show mild stunting and delayed bolting. Plants infected with TCV and satC show intensification of symptoms with plant death by 21 dpi. (b) *Arabidopsis* protoplasts were infected with TCV with (TCV + C) and without satC. The genomic RNA and satRNAs were detected by RNA gel-blot analysis while the CP and virions were detected using the anti-TCV CP antibody (*39*).

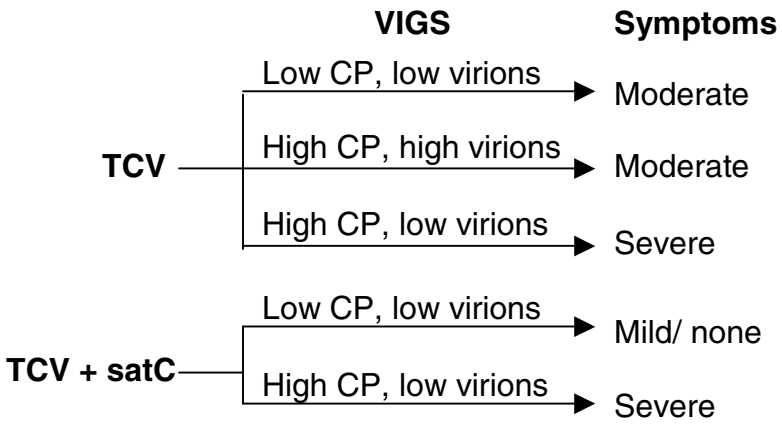

Figure 16. Model for symptom modification by satC. The ds RNA regions in satC and TCV induce VIGS. Since unassembled CP is a suppressor of VIGS, a high level of CP combined with a low level of virions results in an increase in virus movement and thus severe symptoms (*38*).

Previous studies have indicated that the CP does not have a binding affinity for the 3' end of satC, although this region is important in symptom modulation (*51*). Recently an internal plus-strand hairpin of satC, the M1H, which brings together flanking CA-rich sequences, has been found to increase satC fitness and reduce virion accumulation (*52*). How it may interfere with virion formation is not known. Although this hairpin does contain a segment of the sequence found in the essential packaging element of the TCV genome, its ability to reduce virion formation is sequence-nonspecific (see Figure 17) (*53*).

satC M1H (+)-strand

Figure 17. The M1H (+)-strand internal hairpin. The hairpin consists of the 3′ end of satD, 16 bases from the TCV CP coding region and the 3' untranslated region of the TCV genome. The segment corresponding to part of the putative nucleation site is indicated above. Sequences labeled in blue are the original sequences and asterisks denote position differences with wt satC (*53*).

E. TCV Hypothesis

It has been previously theorized that satC interacts with the CP to disrupt the assembly process and thereby increase the amount of free CP in the plant. An increase in the amount of free CP early on in the infection process has been linked to systematic spread of the virus and exacerbation of symptoms already present. Even so, satC will not produce symptoms in plants that are already tolerant (*39*). The increased levels of free CP resulting from inhibition of virus assembly by satC, allows increased suppression of VIGS at the initiation step (see Figure 18) (*16, 38*). As a result the viral RNA is not degraded and the virus is still able to replicate and move throughout the plant.

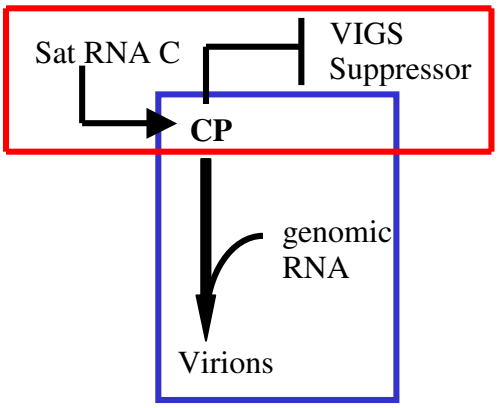

Figure 18. A flow-chart representation of the Hypothesis describing the two functions of the CP. The red box highlights the CP as a suppressor of VIGS while the blue box highlights its ability to form virions.

In order to test the first part of the hypothesis, which proposes that satC interacts with the CP to inhibit virus assembly, a large quantity of recombinant CP from *E. coli* must be obtained and its ability to assemble *in vitro* quantified. An *in vitro* assembly

assay that allows for the observation of the complete assembly process including intermediates needs to be developed so that inhibition of assembly by satC can be measured. If it is determined that satC directly inhibits virion formation *in vitro*, experiments can be performed between CP and satC allowing for the identification and characterization of all RNA-protein interactions relevant to this inhibition. In this thesis the TCV coat protein gene has been cloned and the recombinant protein expressed and purified from *E. coli*. It has also been characterized for future use in the assembly assay.

F. Overview Statement

Chapter 1, the Introduction, provides background on both viruses in general and specifically TCV in regards to virion structure, assembly and virus induced gene silencing in plants. The objective of this thesis research was to clone the CP gene and express, purify and characterize the TCV CP from *E. coli* in order to perform assembly experiments *in vitro*. This was done as foundation work so that the satC and CP hypothesis could be tested *in vitro* in the future.

G. Research Justification

Before the assembly mechanism can be investigated the coat protein had to first be characterized with respect to its folded conformation and assembled state in solution under different conditions, monomer, dimer, ect. All previous *in vitro* experiments with TCV have used the CP produced from either infected plants or protoplasts. In order to

attempt characterization using methods such as analytical ultracentrifugation, circular dichroism and other biophysical techniques, a large amount of protein is required. Extracting the CP from virions in the original manner requires both time and a large quantity of infected material since the virus must first be recovered from plant tissue before the CP can be extracted. It can take an inoculated plant on the average 2-3 weeks before symptoms emerge and infected leaves can be harvested for viral extraction/purification. Once infected leaves have been harvested the virus must first be extracted and purified before it can be dissociated. After dissociation of the virus the CP has to be separated from the genomic RNA, residual nucleic acids, and purified before experimental use. Since the coat protein cannot be directly purified from the leaf tissue the probablilty that a large yield of protein comparable to the original concentration present in the tissue can be recovered after the purification steps is low. During purification, with each additional step protein can be lost. According to Morris (1995), overall CP recovery from virions is about 40% (*37*).

The original method takes an extensive amount of time and labor for the amount of protein recovered. Therefore, a different method of large-scale purification of TCV coat protein had to be devised. It consisted of cloning the CP coding sequence, expressing the protein in an *E. coli* strain and devising a purification procedure that allowed the protein to be produced within a short time period and in quantities appropriate for future biophysical experimentation, including assembly experiments. In the future specific mutants of the CP can also be generated easily via this method. This will allow investigation of interactions between the CP, genomic RNA and satC.

II. Experimental

Chapter II describes all the materials and methods used in this research. It includes all methods associated with the cloning of the TCV CP coding sequence and expression, purification and characterization of the recombinant coat protein. The techniques used to obtain purified virions and CP from infected plant tissue, images of the virions and the synthesis/purification of the TCV genomic RNA are also discussed in detail.

A. Plasmid Constructs

The full-length TCV MS genomic DNA sequence inserted into a pUC19 plasmid was obtained from Anne Simon's lab (University of Maryland, Department of Cell Biology and Molecular Genetics). The CP coding sequence was amplified using polymerase chain reaction (PCR) with a MasterAmp Taq DNA polymerase kit (Epicentre Biotechnologies). The 50 μL PCR reaction contained 1 x PCR enhancer, 3 x PCR buffer, 200 μM of each dNTP, 2.5 mM $MgCl_2$, 1 ng/μL TCV plasmid template, 0.6 μM of each oligo, and 1 μL Taq DNA polymerase enzyme. The CP N-terminal oligo sequence 5' – CAA CAC ATA AGC ATC AAC ACT GCA TAT GGA AAA TGA TCC TAG AG – 3' contained the Nde I cleavage site while the C-terminal oligo sequence 5' – GAG AAG ACT ACA CAG GAT CCG TAC TAA ATT CTG AGT GCT TGC – 3' contained the BamH I cleavage site (see Figure 19).

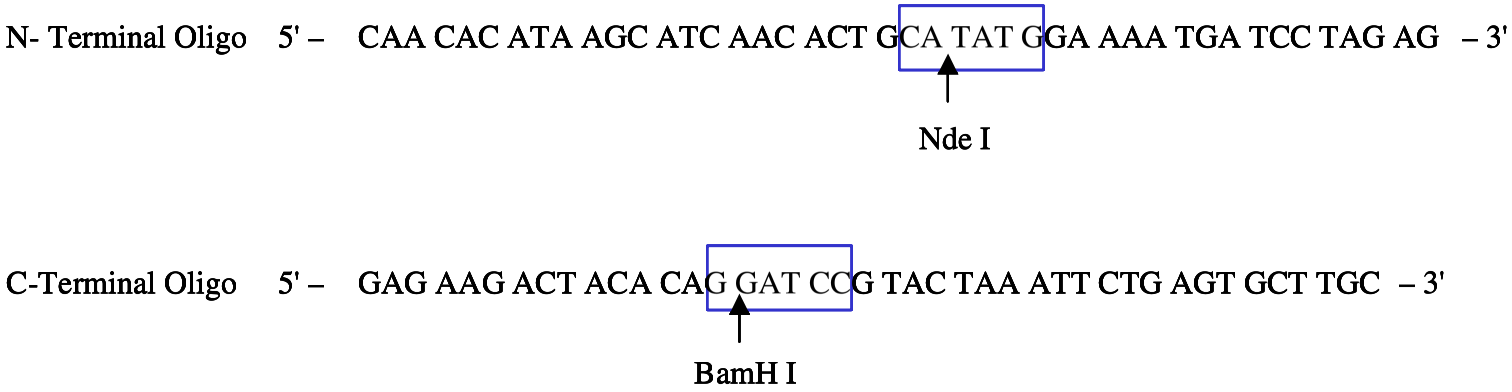

Figure 19. The sequences of the oligos used for amplification of the CP gene are shown with the restriction enzyme cleavage sites indicated.

The PCR protocol used was as follows; the first stage was 5 minutes at 94°C, the second stage was 30 cycles at 94°C 1 minute, 43°C 1 minute and 72°C 2 minutes and the third stage was 1 minute at 72°C with the reaction being held at 4°C upon completion. The complete PCR reaction was loaded onto an agarose gel, the band corresponding to product was excised and the TCV CP DNA eluted via electrophoresis. The DNA was then ethanol precipitated, resuspended in TE buffer containing 10 mM Tris-HCl, 1 mM EDTA pH 8.0, ligated into a pGEM-T Easy PCR cloning vector (Promega), transformed into Top 10 cells and amplified *in vivo*. The insert was excised from the purified plasmid with the restriction enzymes Nde I and BamH I (New England BioLabs) and ligated into the Novagen bacterial expression vector pET-17xb that confers ampicillin resistance. This pET-17xb/TCV CP plasmid was transformed into the *E. coli* cell strain BL21λ(DE3)pLysS, amplified *in vivo* and the DNA recovered and purified for sequencing (see Figure 20A; Anne Colgrove and Beckett unpublished).

CP DNA purified from eight different isolates was subjected to DNA sequencing (DNA Sequencing Facility operated by the Center for Biosystems Research at the University of Maryland College Park Campus). The sequencing data was analyzed using the sequence alignment program LALIGN found on the ExPASy website, Expert Protein Analysis System proteomics server of the Swiss Institute of Bioinformatics (www.expasy.org), by comparing the sequences to both coat protein coding sequences MS and B of TCV. The difference in the two CP DNA sequences exists at the genome positions 209, 371 and 1037, which code for the amino acids S70, A124 and L346 for TCV-B and T70, V124 and W346 for TCV-MS, respectively (see Figure 21). Upon

analysis of the sequencing data it was observed that all of the isolates contained numerous silent and missense mutations. By comparing the sequences it was determined that the two isolates #1 and #8 could be digested with the restriction enzymes HinD III and BsaA I (New England BioLabs) and one fragment from each ligated together to decrease the overall number of missense mutations in the full CP gene (1056 nt) to only three. The two isolates were cut with these restriction enzymes HinD III and BsaA I, for which each enzyme only had one cleavage site on the plasmid. A 612 nt fragment from isolate #1 was ligated to a 443 nt fragment from isolate #8 in an effort to form a sequence with the least number of missense mutations (see Figure 20B). These three mutations were then corrected by site-directed mutagenesis using Stratagene's Quick-change site-directed mutagenesis kit and the amplified DNA re-sequenced (see Figure 22). The protein parameters including molecular weights and theoretical isoelectric points were calculated using the ProtParam program also found at the ExPASy website. The TCV CP-MS extinction coefficient used for CP-MS concentration determination was calculated based on its amino acid sequence (*54*).

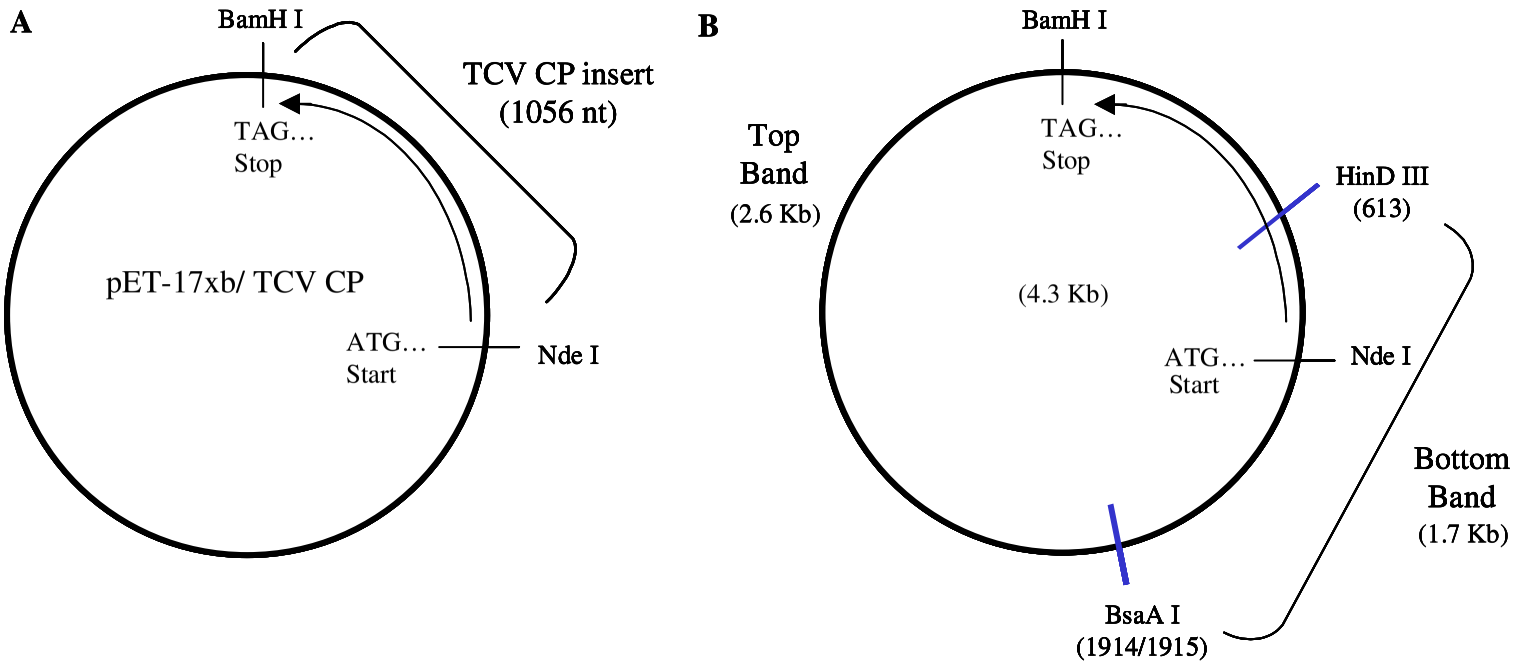

Figure 20. The pET-17xb plasmid containing the TCV CP insert. (A) Indicates the placement of the TCV CP insert in the pET plasmid with the appropriate restriction enzyme cleavage sites used to cut the plasmid before ligation labeled. (B) Represents the plasmid with the restriction sites for HinD III and BsaA I labeled in blue. The bottom band and top band formed through the cleavage process are indicated.

Position (CP gene/ TCV genome)	Codon Nucleotide Mutations (resulting amino acid mutations)		
	Plasmid Construct	TCV CP-B	TCV CP-MS
208/ 2951	ACC (T70)	TCC (S70)	ACC (T70)
371/ 3114	GCC (A124)	GCC (A124)	GTC (V124)
1037/ 3780	TGG (W346)	TTG (L346)	TGG (W346)

Figure 21. The table indicates the mutations in the DNA sequence of the ligated construct that had to be changed in order to construct the two CP sequences MS and B. The nucleotides labeled in blue text are those that were different compared to the plasmid construct. Those nucleotides labeled in red text were corrected with site-directed mutagenesis.

Position (CP gene/ genome)	Mutation	Primers
208/ 2951	A → T	5'– GCC TAC CGC GAG GTT TCC ACC CAG CCT CGG G – 3' 5' – CCC GAG GCT GGG TGG AAA CCT CGC GGT AGG C – 3'
371/ 3114	C → T	5'– GCT CAT TAA GGA GGC GGT CCA GTA TGA AAA ATA CC –3' 5'– GGT ATT TTT CAT ACT GGA CCG CCT CCT TAA TGA GC – 3'
1037/ 3780	G → T	5'– GAG CAG CCA AAG GGT AAA TTG CAA GCA CTC AGA ATT TAG – 3' 5'– CTA ATT TCT GAG TGC TTG CAA TTT ACC CTT TGG CTG CTC – 3'

Figure 22. A table of the DNA primers needed for the site-directed mutagenesis for the changes at the positions as indicated.

B. Induction tests for pET-17xb/TCV CP-MS and B transformed into *E. coli* strains

1. Strains BL21λ(DE3) and BL21λ(DE3)pLysS

Both coat protein MS and B coding sequences inserted into the Novagen pET-17xb plasmid were originally transformed into two *E.coli* cell lines BL21λ(DE3) and BL21λ(DE3)pLysS. These transformants were plated on LB/ 100 μg/mL ampicillin plates and a colony from each was grown in a 5 mL volume of LB/ 100 μg/mL ampicillin media. Induction tests were performed with these cultures to determine which yielded the highest CP expression level. A 5 mL culture of each was grown at 37°C, shaking at 300 rpm until the OD_{600} was ~0.9. At this point a 1 mL sample was transferred to a new culture tube and the expression of the CP in the remaining sample was induced by adding IPTG to a final concentration of 75 μg/mL. All cultures were incubated at 37°C for a remaining one to two hours shaking at 300 rpm after which 500 μL samples of both the uninduced and induced cultures were removed for analysis on SDS-polyacrylamide gels.

2. Strain BL21λ(DE3)

Induction tests of 50 mL with varying incubation temperatures and induction times were performed with both pET-17xb/TCV CP plasmids transformed into the BL21λ(DE3) cell line. Four 50 mL cultures were grown for each transformant with growth temperatures of 37°C 30°C, 25°C, and 20°C respectively. At an OD_{600} of ~0.9 two mL from each were removed and placed into corresponding tubes labeled uninduced

while the expression of the CP in the remaining 48 mL was induced by addition of IPTG to a final concentration of 75 μg/mL. Both the uninduced and induced cultures continued to grow at the same initial temperature for a total of four hours with a 500μL sample removed from the induced culture at each hour past the time of induction. An uninduced sample was removed at the four-hour mark. All samples were analyzed on SDS-polyacrylamide gels to determine the level of expression in relation to the induction time and temperature.

An induction test was performed for the pET-17xb/TCV CP-MS plasmid transformed into the BL21λ(DE3) cell line grown at 25°C varying the initial OD_{600} and length of the induction. Initially a 200 mL culture was grown at 25°C on the shaker at 300 rpm with the OD_{600} measured every 20 minutes. When the cell density reached OD_{600}s of 0.5, 0.6, 0.7, 0.8 and 0.9, a culture volume of 25 mL was removed. The expression of the CP was induced in each of these samples by addition of IPTG to a final concentration of 75 μg/mL. The induced cultures continued to grow at the same initial temperature for a total of five hours with a 500μL sample removed from the induced culture at each hour past the time of induction. A sample was taken from the uninduced culture at the five-hour mark. All samples were analyzed on SDS-polyacrylamide gels to determine the level of expression in relation to the OD_{600} at induction and the length of the induction.

3. Strain Rosetta(DE3)

The recombinant plasmid pET-17xb/TCV CP-MS was transformed into the *E.coli* cell line Rosetta(DE3). The transformants were plated on LB/ 100 μg/mL ampicillin/ 20 μg/mL chloramphenicol plates and a colony from each was grown in a 5 mL volume of LB/ampicillin/chloramphenicol media. Induction tests were performed to determine the CP expression level at different growth temperatures. A series of 5 mL cultures were grown at either 37°C or 25°C with shaking at 300 rpm, until the OD_{600} was between 0.6 to 0.9. At this point a 1 mL sample was transferred to a new culture tube and the expression of the CP in the remaining sample was induced by the addition of IPTG to a final concentration of 75 μg/mL. All cultures were incubated for a remaining one to two hours at the corresponding temperature with shaking at 300 rpm after which 500 μL samples of both the uninduced and induced cultures were removed for analysis on SDS-polyacrylamide gels.

C. Protocol for Expression and Purification from BL21λ(DE3)

1. Expression of Recombinant TCV CP (500 mL protein preparation trial)

E.coli strain BL21λ(DE3) was freshly transformed with the pET-17xb/TCV CP-MS plasmid and grown in LB media containing 100 μg/mL ampicillin with shaking at 300rpm either at 30°C, 25°C or 20°C. Once the cell density achieved an OD_{600} of ~0.9 expression was induced by adding IPTG to a final concentration of 75 μg/mL. Once it

was determined that a growth temperature of 25°C with an induction time of four hours was ideal, the protocol was altered so that the OD_{600} was 0.7 to 0.8 before the cells were induced. After a 4 hour induction period cells were harvested by centrifugation at 7,500*g* for 20 minutes at 4°C. A 500 µL sample of the uninduced and induced cultures were taken for future protein identification using SDS-PAGE gel electrophoresis.

2. Sonication of Cells

The pellet containing the harvested cells was resuspended in 100 mL of lysis buffer containing 100 mM sodium phosphate pH 6.5, 200 mM NaCl, 5% (v/v) glycerol and 0.1 µM DTT, pelleted by centrifugation at 7,500*g* for 20 minutes at 4°C and resuspended in 50 mL of lysis buffer containing 0.1 mM DTT and 0.1 mM PMSF. The resuspended pellet was transferred to a stainless steel 125 mL beaker and put on an ice-salt bath. Once the solution temperature was below 4°C the cells were subjected to sonication, using 30 second bursts and cooling to <4°C between bursts, until the OD_{600} was less than one-tenth the original value. Cells were then diluted to 200 mL with lysis buffer, 0.1mM DTT, 0.1mM PMSF and centrifuged at 12,000 rpm for 45 minutes at 4°C using a SLA-1500 rotor (Sorvall). The resulting pellet was resuspended once in 24.5 mL of lysis buffer, 0.1 mM DTT, 0.1 mM PMSF and another sample was removed for SDS-PAGE gel analysis.

3. Guanidine-HCl Extraction

The method for extraction of insoluble protein using guanidine-HCl was modified from a protocol in Current Protocols in Protein Science (*55*). The resuspended post lysis pellet was centrifuged at 22,000*g* for 1 hour at 4°C. The resulting pellet containing the inclusion bodies was washed twice with 25 mL wash buffer containing 100 mM Tris-HCl pH 7.0 at 4°C, 5 mM EDTA, 5 mM DTT, 0.1 mM PMSF, and supplemented with 2 M urea and 2% (v/v) Triton X-100. The urea and Triton X-100 solubilized proteins were removed by washing the pellet once in 25 mL wash buffer only. After each pellet resuspension, the solution was centrifuged at 12,000 rpm for 30 minutes at 4°C using a SLA-1500 rotor. The washed inclusion bodies were then solubilized in 5 mL of extraction buffer containing 50 mM Tris-HCl pH 7.0 at 4°C, 5 mM EDTA, 8 M guanidine-HCl, and 5 mM DTT. A 1 mL volume of buffer was added per half gram of wet cells. Insoluble material was removed by ultracentrifugation at 29,000 rpm for 1 hour at 4°C using a Beckman 70.1 Ti rotor. The supernatant was then collected and filtered through a 0.22µM filter and stored at 4°C for purification.

4. Urea Extraction

The urea extraction procedure was similar to the guanidine-HCl method except for some changes in the buffers. The wash buffer now contained 1 M Tris-HCl, 50 mM EDTA, 100 mM NaCl, pH 8.0 at 4°C with 1 mM DTT and 0.1 mM PMSF added immediately before use. It was also supplemented with 2 M Urea and 2% Triton X-100.

The extraction buffer contained 50 mM Tris-HCl, 5 mM EDTA, 8 M Urea, pH 8.0 at 4°C and 10 mM DTT. After extraction the solution was centrifuged at 12,000 rpm for 30 minutes at 4°C using a SLA-1500 rotor (Sorvall) instead of at 29,000 rpm for 1 hour at 4°C using a Beckman 70.1 Ti rotor.

5. Purification of recombinant TCV CP

a. Guanidine-HCl Purification Methods

The gel-filtration chromatography and dialysis was modified from a protocol in Current Protocols in Protein Science (*55*). Chromatography was performed under denaturing conditions in the case of gel-filtration. A variety of different resins were tested including Sephacryl S-200 (Amersham Biosciences) and Sephacryl S-1000SF (Amersham Biosciences). The filtered guanidine-HCl solubilized protein was loaded onto a packed column approximately 1.5 by 50 cm with a bed volume of about 80 mL, previously equilibrated with about 2 column volumes of gel-filtration buffer containing 50 mM Tris-HCl pH 7.5, 4 M guanidine-HCl and 5 mM DTT. The sample was then eluted with an additional two column volumes of gel-filtration buffer. Fractions (~3 to 4 mL each) were collected every 10 minutes at a flow rate ~0.35 mL/min. The fractions containing protein were detected by absorbance at 280nm, checked on 12.5% SDS polyacrylamide gels, and then pooled and stored at 4°C.

The protein was renatured by removing the guanidine-HCl and DTT in a step-wise dialysis procedure. The protocol included four 1 hour dialysis steps performed

against high salt buffer containing 10 mM Tris-HCl, 1 M NaCl, 0.1 mM EDTA, 5% (v/v) glycerol, pH 8.0 at 4°C with 1, 0.6, 0.2, and 0 M guanidine-HCl respectively. All dialysis steps were performed at 4°C. The protein concentration was determined via spectroscopic measurements at 280 nm using the TCV CP-MS calculated extinction coefficient of 57280 $M^{-1}cm^{-1}$ (*54*).

b. Urea Purification Methods

Chromatography followed a protocol based on a method by the Stanley Lab (*56*). A CM Sepharose column (Amersham Biosciences) was run under denaturing conditions with solutions containing urea. All solutions were made fresh and used immediately since urea decomposition leads to cyanate production. Tris-base was added to all buffers to act as a cyanate scavanger. The filtered solubilized protein was applied to a packed CM Sepharose column of 1.5 x 30 cm with a bed volume of ~ 40 mL, previously equilibrated with column buffer containing 50 mM Tris-HCl, 5 mM EDTA, 8 M urea, pH 8.0 at 4°C and 10 mM DTT. The column was then washed with 5 column volumes of column buffer at a flow rate of ~ 0.35 mL/min. to remove any protein contaminants that did not bind. Bound material was then eluted with a linear 0 M to 1.0 M NaCl gradient using 5 column volumes at a flow rate of ~ 0.35 mL/min. Fractions were collected at 10 minute intervals (3.5 mL each) and those containing protein detected by absorbance at 280 nm, checked on 12.5% SDS polyacrylamide gels, and then pooled and stored at 4°C.

D. Protocol for Expression and Purification from Rosetta (DE3) strain

1. Expression of Recombinant TCV CP (500 mL protein preparation test)

E. coli strain Rosetta (DE3) cells from Novagen that were transformed with the pET-17xb containing the TCV CP DNA strain MS coding sequence were grown in LB media containing 10 g/L bacto tryptone, 5 g/L bacto yeast extract, 5 g/L NaCl, 1 mM NaOH, 0.02 g/L Chloramphenicol, 0.1 g/L Ampicillin at 25°C until the absorbance of the culture at 600 nm against uninoculated media was 0.7 to 0.8 Units. T7 transcription was then induced with the addition of isopropyl α-D-thiogalactopyranoside (IPTG; FisherBiotech) to a total concentration of 75 µg/mL. After a 4 hour induction period cells were harvested by centrifugation at 7,500*g* for 20 minutes at 4°C. A 500 µL sample of the uninduced and induced cultures were taken for future protein identification using SDS-PAGE gel electrophoresis.

2. Sonication of Cells

The method for the preparation of the harvested *E. coli* cells for sonication was modified from the Current Protocols in Protein Science (*55*). The pellet containing the harvested cells was washed with 100 mL lysis buffer containing 100 mM Tris-HCl, 5 mM EDTA, 500 mM NaCl and 0.1 µM DL-Dithiothreitol (DTT; FisherBiotech). The cells were then pelleted by centrifugation and resuspended in 50 mL Lysis buffer, 0.1 mM DTT, 0.1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma), and 10 µM N-p-

Tosyl-phenylalanine chloromethyl ketone (TPCK; Sigma). The resuspended pellet was transferred to a stainless steel 125 mL beaker and put on an ice-salt bath. Once the solution temperature was below 4°C the cells were subjected to sonication, using 30 second bursts and cooling to <4°C between bursts, until the OD_{600} was less than one-tenth of the original value. Cells were then diluted to 200 mL with lysis buffer, 0.1mM DTT, 0.1mM PMSF, 10 µM TPCK, and centrifuged at 22,000*g* for 45 minutes at 4°C using a SLA-1500 rotor (Sorvall) at 12,000 rpm. The resulting pellet was resuspended once in a 24.5 mL volume and another sample was removed for gel analysis.

3. Urea Extraction of Inclusion Bodies

An inclusion body extraction method was modified from two different protocols published by the Stanley lab for the expression of the Human Papillomavirus Type 16 L1 protein in *E. coli* and the Current Protocols in Protein Science for the preparation and extraction of insoluble protein from *E. coli* (*55, 56*). The resuspended post lysis pellet was centrifuged at 22,000*g* for 1 hour at 4°C. The resulting pellet containing the inclusion bodies was washed twice with wash buffer containing 100 mM Tris-HCl pH 8.0 at 4°C, 5 mM EDTA, 100 mM NaCl, 1 mM DTT, 0.1 mM PMSF, and supplemented with 2 M urea and 2% (v/v) Triton X-100. The urea and Triton X-100 solubilized proteins were removed by washing the pellet once in wash buffer only. After each pellet resuspension, the solution was centrifuged at 12,000 rpm for 30 minutes at 4°C using a SLA-1500 rotor. The washed inclusion bodies were then solubilized in 5 mL of extraction buffer containing 50 mM Tris-HCl pH 8.0 at 4°C, 5 mM EDTA, 8 M urea, and

10 mM DTT. A 1 mL volume of buffer was added per half gram of wet cells. Insoluble material was removed by centrifugation at 22,000*g* for 30 minutes at 4°C. The supernatant was then collected and filtered through a 0.22µM filter and stored at 4°C for future purification steps.

4. Purification of TCV CP

Chromatography and dialysis followed a protocol based on a method by the Stanley Lab (*56*). A CM Sepharose column was run under similar denaturing conditions as described for the CP purification from the BL21λ(DE3) strain except for the following. The filtered solubilized protein was applied to a packed CM Sepharose column of 2.0 x 16 cm with a bed volume of 30 mL instead of 40 mL. Once the protein was bound to the column and washed, the bound material was eluted with a linear 0.025 M to 0.25 M NaCl gradient using 13 column volumes at a flow rate of ~0.35 mL/min. Fractions were still collected at 10 minute intervals (3.5 mL each) and those containing protein detected by absorbance at 280 nm, checked on 12.5% SDS polyacrylamide gels, and then pooled and stored at 4°C.

This purified protein was renatured by removing the denaturing agent and DTT in a step-wise dialysis procedure. Five 2 hour dialysis steps were performed against high salt buffer containing 50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 8.5 at 4°C with 4.0, 2.0, 1.0, and 0 M urea, respectively. Finally the protein was dialyzed once for 2 hours against high salt buffer containing 5% (v/v) glycerol. All the dialysis steps were

performed at 4°C. The protein concentration was determined by UV spectroscopy at 280 nm with a calculated TCV CP-MS extinction coefficient of 57280 $M^{-1}cm^{-1}$. All dialyzed samples were then stored as aliquots at -70°C for further analysis and characterization.

E. Analytical Ultracentrifugation Experiments

Sedimentation velocity and equilibrium experiments were carried out in a Beckman Optima XL-1 analytical ultracentrifuge (Beckman Coulter, Fullerton, CA, USA) with an An 60 Ti rotor and absorption optics at 20°C. Purified stored TCV CP was thawed on ice and dialyzed against either high salt 1 M NaCl buffer, 0.5 M NaCl buffer or 0.25 M NaCl buffer containing 100 mM Tris-HCl, 5 mM EDTA, pH 8.5 at 20°C, in order to remove all glycerol. Samples were then prepared by dilution into the appropriate dialysis buffer at different concentrations.

Sedimentation velocity experiments were performed at speeds of 45,000 rpm and data collected at 280 nm. Absorbance scans were collected at intervals of 15 sec in replicates of three scans with a radial step size of 0.003 and a total of 100-300 scans were recorded per sample. Spectra were analyzed with the DCDT+ version 1.15 software (*57*). Raw data scans were loaded into the program and graphed allowing the meniscus to be set. The dc/dt values for each pair of scans were calculated and the corresponding graphs overlayed. The radial cutoff at the cell base was adjusted so that the curves did not exhibited negative spiking or ‘tails’ at the ends . The time range was then tested for excessive broadening. The number of scans included in the analysis was dependent on the

mass limit value. If this value was near to or less than the expected mass of any species in the sample the number of scans included in the analysis was reduced. Additional scans were added only when the mass limit value far exceeded the expected mass. Once an acceptable value was reached the average dc/dt was calculated and graphed. The upper and lower sedimentation coefficients for the range were set and used to calculate the g(s*) distribution. This was then fit to gaussians allowing the sedimentation coefficient and molecular mass for various species to be derived based on the specific model implemented in the analysis process (*57*). The buffer density, buffer viscosity, TCV CP partial specific volume (vbar) and all temperature corrections were calculated using the protein analysis program SedNTerp (John Philo, 6/2001).

Sedimentation equilibrium experiments were made with rotation speeds of 10,000, 20,000, 22,000 and 26,000 rpm and data collected at 280 nm. Absorbance scans were collected with a radial step size of 0.003 and in replicates of three. Samples were allowed to achieve equilibrium over the course of 18 hours and were determined to be at equilibrium by superimposing two scans collected 1 hour apart. The data were then analyzed with NONLIN software (*58*).

F. Circular Dichroism Experiments

Circular dichroism (CD) measurements were recorded on a Jasco CD J-810 spectrophotometer at 20°C, using a 0.1 cm path length cuvette. Samples were prepared with concentrations of 9-10 μM TCV CP in both sodium phosphate, Tris buffer

containing 100 mM NaH_2PO_4/Na_2HPO_4, 50 mM Tris-HCl, 1 mM EDTA, pH 8.5 at 20°C and sodium phosphate buffer containing 100 mM NaH_2PO_4/Na_2HPO_4 pH 8.5 at 20°C. Far UV spectra were collected from 180 nm to 250 nm. Data below 200 nm could not be analyzed due to buffer interference. Buffer absorbances were collected and subtracted from the samples before analysis, which was performed with the JASCO Spectra Manager program. Far UV CD values were reported in the form of AU (absorbance units).

G. TCV Genomic RNA Synthesis

The pUC19 plasmid containing the full-length TCV genomic DNA insert, obtained from Anne Simon's Lab, was linearized by digestion with the restriction enzyme Sma I (New England BioLabs) and purified by phenol:chloroform extraction and ethanol precipitation before use as the reaction template. The TCV genomic RNA was synthesized using the MEGAscript kit (Ambion) with transcription of the TCV genomic DNA template by the T7 RNA polymerase. All synthesis and purification steps followed the MEGAscript kit instruction manual. The 20 μL reaction contained 2 μL of each 75 mM ATP, CTP, GTP and UTP (7.5 mM final concentration), 2 μL 10 x Reaction buffer, 1 μg linearized template DNA and 2 μL enzyme mix. It was then incubated at 37°C for about three hours. The synthesized RNA was purified by either lithium chloride precipitation or by a phenol:chloroform extraction followed by an ethanol precipitation. The lithium chloride precipitation solutions were provided in the kit. All other solutions used were prepared from diethyl pyrocarbonate (DEPC)-treated/autoclaved milli-Q

water. Tubes and pipet tips were purchased as nuclease-free (Ambion) and all other glassware/equipment was treated with RNase Away (Molecular Bioproducts) and rinsed with DEPC-treated/autoclaved milli-Q water. For the ethanol precipitation and wash steps a 99.5% ACS grade ethanol reagent was used to decrease the amount of degradation observed in the samples. All synthesized RNA was resuspended in nuclease-free water before determining the absorbance at 260 nm and storing at -20ºC. For long term storage RNA should be ethanol precipitated and stored at -70ºC. RNA samples were checked on formaldehyde-agarose gels to determine the size and efficiency of each step.

H. Polyacrylamide and Agarose Gel Electrophoresis

1. SDS-PAGE Gels

The induction, extraction and purification samples of recombinant CP were monitored by 12.5 % SDS polyacrylamide gels (*59*). Induction samples were collected, pelleted, resuspended in sample buffer containing 0.0625 M Tris-HCl pH 6.8, 3% SDS, 30% glycerol, 0.0125% Bromophenol blue with 5% (v/v) β-mercaptoethanol and heated at 90°C for 10 minutes before loading. Urea extraction and CM Sepharose fraction samples at 8 µL each were added to 10 µL sample buffer with 5% (v/v) β-mercaptoethanol and heated as described previously. All samples were loaded on 12.5% polyacrylamide gels and run at 150 V to 250 V using running buffer containing 3 g/L Tris-base, 14.4 g/L Glycine, 0.1% SDS. Gels were then stained with Coomassie staining solution containing 4.6% acetic acid, 47.5% ethanol, 2.5 g/L coomassie brilliant blue R-

250, for 1 hour, destained with destaining solution consisting of 4.6% acetic acid, 24% ethanol and then dried using the Invitrogen gel drying system.

a. Guanidine-HCl sample treatment

All samples that contained guanidine-HCl had to be treated to remove the denaturant before loading onto polyacrylamide gels (*55*). For each 25µL sample, 225µL of cold ethanol was added, vortexed and stored at -20°C for 5 to 10 minutes. Once the protein had precipitated, the sample was microcentrifuged at 15,000*g* for 5 minutes at 4°C. The supernatant was removed and the pellet washed once with 250 µL cold ethanol and microcentrifuged again. The resulting pellet was resuspended into 25 µL sample buffer as indicated above.

2. Formaldehyde-Agarose Gels

RNA samples were run on formaldehyde-agarose gels using an established protocol, to determine the efficiency of RNA synthesis and analyze the purification procedure (*60*). A 1% agarose-formaldehyde gel was prepared by dissolving 1g agarose into 72 mL DEPC-treated distilled water and cooling to 60°C in the water bath. Under the hood 10 mL of 10 x MOPS buffer, containing 0.4 M MOPS pH 7.0, 0.1 M sodium acetate and 0.01 M EDTA, and 18 mL formaldehyde were added. The gel was then poured and allowed to set at least 2 hours before use. Samples were prepared by adding, 1 to 2 µL RNA to 2.5 µL 10 x MOPS buffer, 4.5 µL 37% formaldehye, 12.5 µL formamide and bringing the volume up to 25 µL with DEPC-treated distilled water. The samples

were then vortexed, centrifuged down and incubated 15 minutes at 65°C before adding 10 μL loading buffer containing 1 mM EDTA pH 8.0, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol and 50% glycerol. After loading the gel under the hood the samples were electrophoresed in 1 x MOPS buffer at 80 V until the bromophenol blue had migrated two-thirds the way down the gel. The gel was then removed and stained in DEPC-treated water containing 0.5 μg/mL ethidium bromide. The protocol was modified with respect to the length of time for destaining in DEPC-treated/autoclaved milli-Q water. The destaining procedure was run overnight for best results. All solutions, tips, tubes and equipment were RNase free.

3. Agarose Gel Electrophoresis of TCV

Virion samples were run on 0.6% high strength agarose (USB) gels prepared with 1 x TAE buffer containing, 40 mM Tris acetate, 2 mM EDTA pH 8.5, as described by both the Young and Zlotnick labs (*30*). All samples were mixed with loading buffer containing 30% glycerol, 0.25% bromophenol blue, and 0.25% xylene cyanol, before loading onto the gel. Once all samples were loaded into the appropriate wells the gel was electrophoresed at 60V in 1 x TAE buffer until the bromophenol blue band had migrated two-thirds the way down the gel. It was then stained for 5 minutes in 5 μg/mL ethidium bromide, destained overnight in distilled water and visualized under a UV light.

J. Virus and CP Purification from plant tissue

1. Plant growth and virus inoculation

Turnip plants were grown in growth chambers at 20°C as described previously by the Simon lab (*51*). At least thirty-three turnip seedlings were required per inoculation with three seedlings as the control group. Seedlings with at least four leaves were transplanted into sterile pots with three plants per pot and were incubated in a growth chamber. All plants were watered 24 hours before inoculation and the leaves cleaned of any soil. Those seedlings with four to six leaves were mechanically inoculated. The two oldest leaves on each plant were inoculated with 10 μL inoculation solution containing 1μg TCV genomic RNA per leaf. The pUC19/TCV plasmid cleaved with the restriction enzyme Sma I was used as the DNA linear template for synthesis of the TCV genomic RNA. The RNA was synthesized at 37°C for 3 hours in a 100 μL reaction volume containing 3 μL 1 M Tris-HCl pH 8.0, 16.5 μL 150 mM $MgCl_2$, 2 μL 100 mM Spermidine, 10 μL 100 mM DTT, 1 μL Triton X-100, 8 μL of each 50 mM dNTP, 16 μL 5 x T7 polymerase buffer, 13 μL linear plasmid TCV DNA at 1μg/μL, 2 μL T7 RNA polymerase and 2.5 μL RNasin. A sample was run on an agarose gel for identification of the TCV product before use in the inoculation process. The concentration of genomic RNA was calculated using the absorbance at 260nm, the RNA length of 4054 bps and an average base weight of 330. An RNA purification procedure was not necessary before inoculation. A 600 μL inoculation solution was prepared with 60 μg TCV genomic RNA, 300 μL 2 x autoclaved infection buffer containing 0.05 M glycine, 0.03 M K_2HPO_4 pH

9.2, 1% bentonite, 1% celite, and 240 μL distilled water. The 10 μL inoculation solution was pipeted onto each leaf and, using a sterile glass, stirring rod, rubbed gently into the leaf surface. All plants remained un-watered for 48 hours past the inoculation process, after which plants were watered heavily once a week. Symptoms became visible in two to three weeks past inoculation. These symptoms included the presence of lesions and yellowing of the older outer leaves while the dark green newer leaves were stunted and severely crinkled. Once at least three to four newer leaves have developed per plant these leaves can be harvested for virus purification.

2. Virus Purification

The method for TCV purification from infected plant tissue was adopted and modified from the method by Skuzeski and Morris (*37*). Approximately 30 grams of turnip leaves were harvested at least 28 days past inoculation (dpi). Every 10 g tissue was added to and ground with 30 mL sap extraction buffer containing 0.2 M sodium acetate pH 5.2, 30 μL β-mercaptoethanol, and protease inhibitor mixture (50 μM TPCK, 0.5 μM PMSF and 10 μg α_2 –macroglobulin per mg virions) in a mortar on ice. Once homogenous the ground mixture was filtered through two layers of cheesecloth and the collected supernatant incubated on ice for 30 minutes. It was then centrifuged at 7,300 rpm for 1 hour at 4°C in a SA-600 rotor (Sorvall) to pellet cell debris. The supernatant was removed and 10 mL of 40% polyethylene glycol molecular weight 6000 (PEG; Sigma) in 1 M NaCl was added per 40 mL supernatant. Once mixed the solution was incubated on ice for 30 minutes to precipitate the virions. It was centrifuged at 7000 rpm

for 30 minutes at 4°C using a SA-1500 rotor (Sorvall) to pellet the virions, which were then resuspended in 20 mL resuspension buffer containing 0.05 M sodium acetate pH 5.2 and incubated on ice for 1 hour. The solution was centrifuged again at 7,300 rpm for 30 minutes at 4°C in the SA-600 rotor and the supernatant removed for further centrifugation. The supernatant was centrifuged at 47,900 rpm for 90 minutes at 4°C using a Type 70 Ti Beckman rotor. The pellet containing the virions was resuspended in 2 mL column buffer containing 0.01 M sodium phosphate pH 7.0 and 2.5 % glycerol. It was then briefly clarified by microcentrifugation before loading onto a DEAE-agarose gel A (Bio-Rad) packed 18 mL column 1.5 by 10 cm pre-equilibrated with column buffer at 4°C. The column was washed with 4 column volumes of column buffer and then eluted with a NaCl gradient (0 to 1 M NaCl) using 6 column volumes with a flow rate of 0.35 mL per minute. Fractions were collected at 10 minute intervals (3.5 mL each) and the virion absorbance was determined and recorded at 280 nm by the UV/Vis detector (ISCO UA-6 with optical unit type 11). All fractions were checked on 12.5% SDS-polyacrylamide gels and those fractions containing only CP were pooled. The virions were re-precipitated by adding a half volume of 40 % PEG supplemented with 1 M NaCl, mixed and allowing the solution to sit overnight at 4°C before centrifugation at 7000 rpm for 30 minutes at 4°C in a SA-600 rotor. The pellet containing the virions was resuspended in 50 mM Tris-HCl pH 8.5 at 4°C with 0.01% sodium azide and stored at 4°C. The concentration was determined by its absorbance at 260 nm, and calculated with a known extinction coefficient of 50 at 260 nm for a 1% solution or 0.01 g/mL virions.

3. Coat Protein Purification

The method for CP purification from TCV was adopted and modified from the protocol by Skuzeski and Morris (*37*). About 4 to 5 mg virions were dissociated by incubating in dissociation buffer containing 0.1 M Tris-HCl, 5 mM EDTA, 1 M NaCl pH 8.5 at 4°C, and the protease inhibitor mixture (50 μM TPCK, 0.5 μM PMSF and 10 μg α_2 –macroglobulin per mg virions) on ice for 1.5 hours. The RNA was precipitated by adding a 5% (w/v) PEI solution by rapid mixing to a final concentration of 0.5 mg/mL or 0.05%. The solution was incubated on ice for 15 minutes before centrifugation at 13,000 rpm for 10 minutes at 4°C in the microcentrifuge. It was then loaded onto a Sephacryl S200HR column 1.0 x 30 cm previously equilibrated with dissociation buffer. Fractions were collected every 10 minutes at a flow rate of 0.22 mL/minute (2.2 mL each) and the CP absorbance was detected at 280 nm using a UV/Vis detector. The corresponding samples were checked on 12.5% SDS-polyacrylamide gels and those fractions with intact CP lacking RNA were pooled. The pooled fractions were then dialyzed for 3 hours against two changes of a high salt storage buffer containing 50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, 5% glycerol, pH 8.5 at 4°C. The CP concentration was finally determined at 280 nm using a calculated extinction coefficient of 57280 $M^{-1}cm^{-1}$ and aliquots were stored at -70°C.

K. Transmission Electron Microscopy

Electron microscopy was used in order to check the size and structure of the virions extracted and purified from infected turnip leaf tissue. Virus concentrations of 0.4 and 0.2 mg/mL in 50 mM Tris pH 8.5 at 4ºC supplemented with 0.01% sodium azide were negatively stained on carbon coated formar grids. A 5 µL drop of 0.1% bacitracin was applied to each grid and allowed to stand 5 seconds then wicked dry. Each diluted sample was then applied in a volume of 3µL and allowed to dry for 5-15 minutes. Once dry one drop of 2.5% uranyl acetate stain, 2.5% (w/v) uranyl acetate in distilled water, was added to each grid for less than 30 seconds then wicked off and dried. All prepared grids were visualized using a ZEISS EM 10CA transmission electron microscope by Tim Maugel at the University of Maryland College Park Laboratory for Biological Ultrastructure. A voltage of 80kV was used consistently while the magnification was varied from 50 to 160 K. Multiple images were recorded on film and used to calculate the approximate virion size.

III. Cloning of CP gene & Expression, Purification and Characterization of Recombinant TCV CP

A. Introduction

Recent research has suggested that the inhibition of assembly of TCV is correlated to the CP's ability to act as a suppressor of gene silencing *in vivo*. In particular it has been observed that the presence of satellite RNA C (satC) in the TCV infection is correlated with (1) a decrease in virion accumulation, (2) an exacerbation of symptoms, and (3) an increase in the suppression of gene silencing (*16, 17, 38, 39*). Suppression of the gene-silencing pathway has also been found to occur with the introduction of unassembled coat protein into plants infected with Green Fluorescence Protein (GFP)-constructs via *Agrobacterium* infiltration. When both unassembled CP and the GFP constructs were introduced into *Nicotiana benthamiana* leaves there was a delayed degradation of these constructs. This was viewed visually by fluorescence of the GFP under long-wavelength UV light. Those leaves co-infiltrated with CP had higher levels of fluorescence that peaked at 5 days past infiltration (dpi) and remained at that level for a two-week period. Leaves infiltrated with GFP alone peaked at 2 dpi and the fluorescence levels started to decrease after 5 dpi. It was also observed that there was a correlation between the increase in the fluorescence and an increase in the levels of GFP mRNA isolated from the infiltrated area. The introduction of the CP suppressed the degradation of the GFP mRNA into siRNAs. This was verified by a lack of siRNAs present, which suggested that the CP is interfering in the early initiation step when Dicer degrades the

dsRNA into siRNAs. It was also determined that in order for the CP to act as an efficient suppressor it had to be present at the time of the silencing initiation (*16*). The TCV CP suppression activity was also observed in *Agrobacterium*-infiltrated *Nicotiana benthamiana* leaves using GFP constructs in another lab. The N-terminal 25 amino acids were determined to be important for this suppression activity (*17*). Based on these results it was hypothesized that the CP must be unassembled in order to act as a suppressor. Therefore satC may be sequestering the CP and allowing it to remain in an unassembled state. This would result in a decrease in the virion assembly while increasing suppression and consequently symptom severity.

In order to determine if satC interferes with virion assembly an *in vitro* assembly assay must be developed and a large quantity of CP purified. The standard protocol for CP purification involves isolating and purifying virions from infected plant tissue, which are then used as the source for CP. As a result the plants must be grown, inoculated and become symptomatic before harvesting for virion purification. Since plants are the main source for virion isolation the complete process is time consuming and inefficient. The CP yields from the dissociated virions are also low. Therefore another method of CP isolation and purification had to be devised.

There have been numerous examples of plant coat protein coding sequences that have been cloned into plasmids and expressed in *E. coli* strains in an effort to obtain recombinant functional protein without the time constraints and low yields associated with isolation from plant tissue. Both *Cowpea chlorotic mottle virus* (CCMV) and

Tobacco mosaic virus (TMV) coat proteins have been expressed in and purified from *E. coli*, specifically the BL21(DE3) strain (*61, 62*). In the case of CCMV, the recombinant coat protein was localized in the insoluble inclusion bodies and the protein was extracted using a denaturant, 8 M urea, before renaturing the CP into a high salt buffer (*61*). The recombinant CP from both of these viruses were tested for reassembly *in vitro* and observed to assemble (*61, 62*).

Based on these previous results it was possible to clone the TCV CP gene into a plasmid, transform it into BL21(DE3) cells, express the recombinant CP and purify it from *E. coli* with it still remaining functional for use in assembly experiments. Once the CP has been purified it will need to be characterized and compared to virus derived TCV CP before use in assembly experimentation.

B. Results

1. Cloning of TCV CP Sequence

In order to obtain recombinant protein the coding sequence was first ligated into an appropriate vector from which it could be expressed. The original TCV-MS genomic DNA inserted into a pUC 19 plasmid was provided by Anne Simon's lab (University of Maryland, Department of Cell Biology and Molecular Genetics). The CP coding sequence was amplified using polymerase chain reaction (PCR) and the reaction run on an agarose gel. The band corresponding to the product was excised and the DNA isolated

via electrophoresis. The CP gene was ligated into a pGEM-T Easy PCR cloning vector and amplified *in vivo*. Once the plasmid was purified the CP coding sequence was excised with restriction enzymes. The coat protein sequence was cloned into the bacterial expression plasmid pET-17xb at which point it was amplified, purified and sequenced (see Figure 20A; Anne Colgrove and Beckett unpublished). Sequence alignment analysis with the original CP-MS and B coding sequences indicated the presence of various mutations in all isolates. In an effort to reduce mutations, isolates #1 and #8 were cleaved with the restriction enzymes HinD III and BsaA I producing two different sized fragments, 1.7 Kb and 2.6 Kb, that could be isolated via gel electrophoresis. The 1.7 Kb fragment consisting of the CP gene nucleotides #1-612 from isolate #1 and the 2.6 Kb fragment consisting of the CP gene nucleotides #613-1056 from isolate #8 were ligated together to form the complete CP DNA sequence (see Figure 20B). Sequence alignment analysis between the new ligated plasmid construct and the TCV CP sequences MS and B indicated sites of silent and missense mutations. The construct contained three missense mutations in the DNA sequence that would result in a change in the amino acid sequence of the CP to a T70, A124 and W346. The differences in the DNA sequence between the CP-MS and B exist at positions 209, 371 and 1037 in the gene (see Figure 21). Site-directed mutagenesis was performed, for the three identified missense mutations, in order to obtain clones of both CP sequences.

2. Expression and Purification of TCV CP

a. Induction of CP expression from *E. coli* cell lines

Once clones were obtained for both TCV CP-MS and B sequences the expression level of the CP in different bacterial strains was investigated. A series of 5 mL CP expression tests were performed using the pET-17xb/TCV CP plasmids transformed into the cell lines BL21λ(DE3) and BL21λ(DE3) pLysS. These cultures were grown at 37°C with CP expression induced at a cell density where the OD_{600} was 0.9 by the addition of IPTG to a final concentration of 75 μg/mL. A 1 to 2 hour induction period resulted in maximum expression levels in the BL21λ(DE3) cell line. In order to determine the optimal conditions for protein expression using this cell line, including growth temperature and length of induction, the pET-17xb/TCV CP plasmids were once again transformed into the BL21λ(DE3) cell line and 50 mL cultures were grown at a range of different temperatures (37°C, 30°C, 25°C, 20°C). Expression of the CP gene was then induced as previously described with samples removed at each hour past the time of induction (1-5 hours). The samples analyzed on 12.5% SDS-polyacrylamide gels indicated that maximum expression of the CP occurred with a culture growth temperature of 25°C and an induction time of 4 hours. These parameters were then implemented in a 500 mL protein preparation test in which it was discovered that the TCV CP was located in the pellet following sonication. In an effort to increase protein solubility a series of 500 mL protein preparation tests were performed whereby the growing conditions were changed. Specifically, the temperature was varied (30°C, 25°C, 20°C, 37°C/25°C), and the ratio of insoluble to soluble protein expressed was compared (see Figure 23).

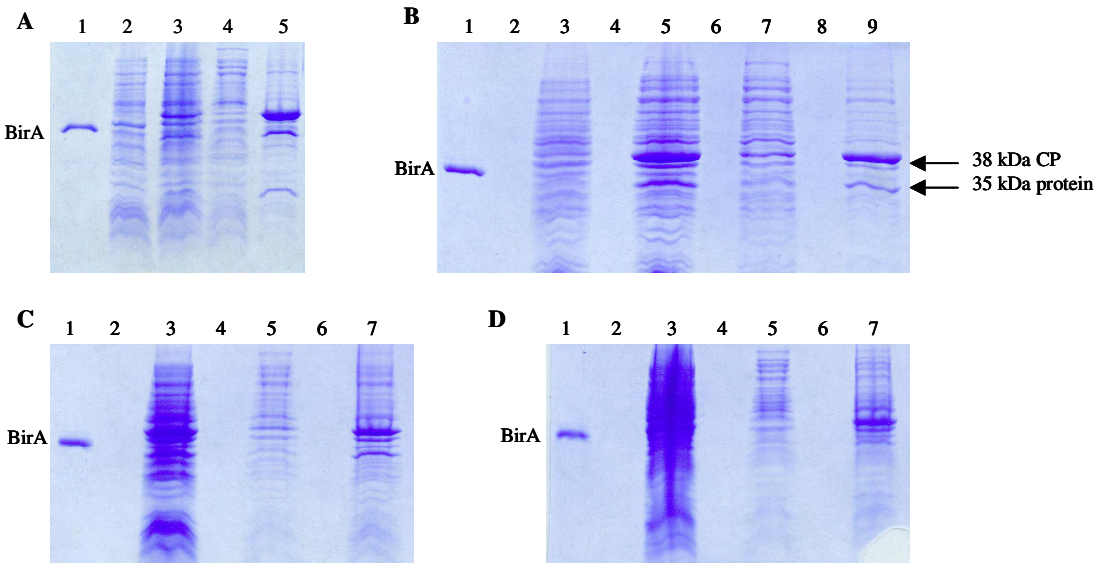

Figure 23. Samples from the 500 mL protein preparation tests of pET-17xb/TCV CP-MS in BL21λ(DE3) grown at varying temperatures were run on 12.5% polyacrylamide denaturing gels for analysis. Cultures were grown at temperatures of 30°C, 25°C, 20°C and 37/25°C with the samples from these protein preparation tests run on the gels A, B, C, and D respectively. The samples indicated in each lane are as follows for gel A: Lane (1) BirA marker, (2) uninduced, (3) induced, (4) post lysis supernatant, and (5) post lysis pellet. The sample sequence for gel B was: Lane (1) BirA marker, (3) uninduced, (5) induced, (7) post lysis supernatant, and (9) post lysis pellet. The samples for gels C and D were loaded as follows: Lane (1) BirA marker, (3) induced, (5) post lysis supernatant, and (7) post lysis pellet.

It was observed that changing the temperature for bacterial cell growth did not improve the solubility of the coat protein. Since the solubility of the CP could not be successfully increased the objective became to obtain all the parameters needed for maximum CP expression using this strain. Therefore a series of tests were performed in order to determine the optimal OD_{600} value for induction. The main 200 mL culture was grown at 25°C as previously described but the initial OD_{600} at the time of induction was varied from an OD_{600} of 0.5 to 0.9. At each point a 25 mL volume was removed from the main culture and expression of the CP was induced by addition of IPTG. Samples taken every hour after the time of induction for each culture were run on denaturing polyacrylamide

gels. The maximum level of CP expression was observed to occur by addition of IPTG to cultures at an OD_{600} between 0.7 to 0.8. The optimal induction time was 4 hours (see Figure 24).

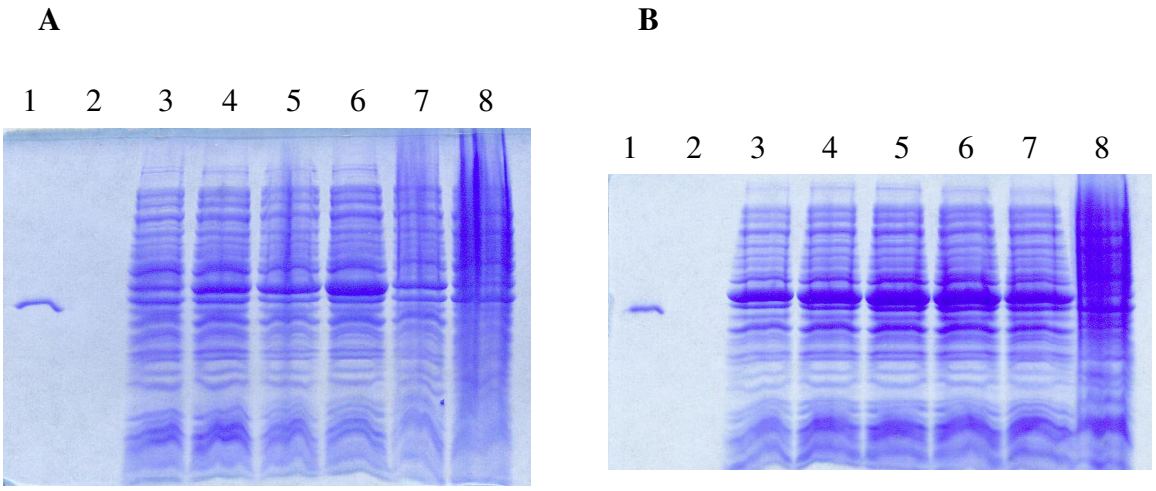

Figure 24. Induction tests at 25°C of pET-TCV CP-MS transformed into BL21l(DE3). Gel (A) represents samples removed after induction of the CP gene at an OD600 of 0.8 at various time points. The lanes are indicated as follows: Lane (1) BirA marker, (3) 1 hr., (4) 2 hr., (5) 3 hr., (6) 4 hr., (7) 5 hr., and (8) uninduced. Gel (B) represents samples taken from cultures 4 hours past the initial time point where the CP gene was induced. The OD600 at the time of induction varied as indicated: Lane (1) BirA marker, (3) 0.5, (4) 0.6, (5) 0.7, (6) 0.8, (7) 0.9, and (8) uninduced.

b. Protein Extraction and Chromatography (pET-17xb/TCV CP-MS in BL21λ(DE3))

Since protein solubility could not be increased by changing the growth conditions, the insoluble protein found in the inclusion bodies of the sonicated pellet was extracted using a denaturant, guanidine-HCl. Most of the protein was extracted with only a small amount left in the residual pellet. Once soluble the denatured protein was purified via chromatography. First the solubilized protein was subjected to gel-filtration chromatography in 4 M guanidine-HCl. Both Sephacryl S-200 and S-1000SF columns

were not useful in separating the CP from another protein that migrated to the 35kDa mark (see Figure 25). The samples containing CP were pooled and dialyzed out of guanidine successfully by step-wise dialysis from the 4 M guanidine sample into high salt buffer. The original protocol consisted of eight 1 hour dialysis steps performed against high salt buffer containing 3, 2, 1, 0.8, 0.6, 0.4, 0.2 and 0 M guanidine-HCl, respectively. Due to the amount of guanidine needed to prepare these solutions the dialysis procedure was modified to six steps starting with the 1 M guanidine-HCl high salt buffer. Since this was also successful in renaturing the protein the dialysis steps were modified to four steps of high salt buffer containing 1, 0.6, 0.2 and 0 M guanidine, respectively.

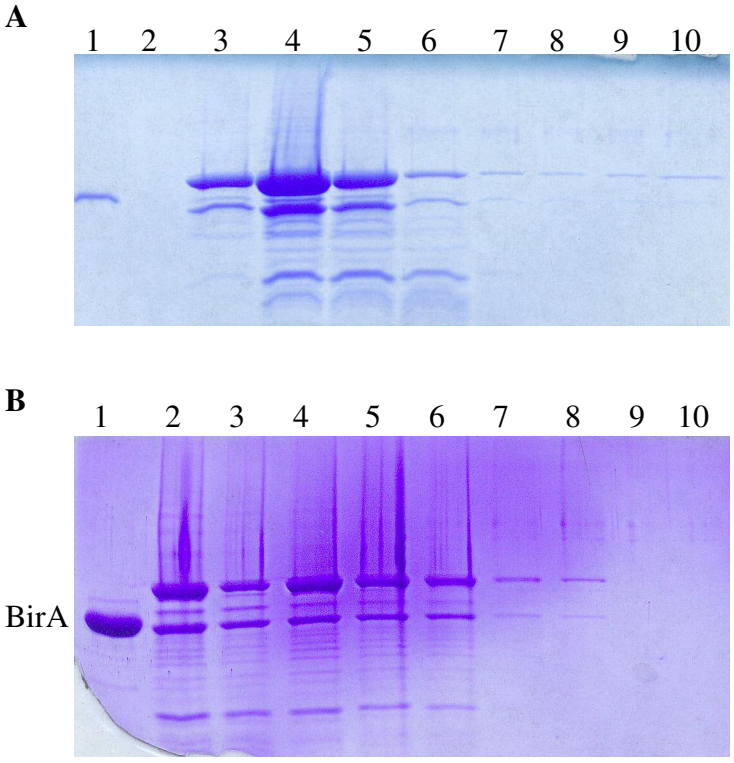

Figure 25. The guanidine-HCl extracted pellet of a 500 mL pET-TCV CP-MS/BL21λ(DE3) protein preparation test was run on different gel-filtration resins. The gels represent samples of fractions collected from both a Sephacryl S-200 (gel A) and a Sephacryl S-1000SF column (gel B) with a BirA marker in lane 1 of each gel. Both columns were run under guanidine-HCl denaturing conditions.

In an effort to separate CP from the 35 kDa contaminant, the dialyzed protein was then tested under non-denaturing conditions on both Q-Sepharose and HAP resins but it was unable to bind to the columns. As the sample's salt concentration was decreased from 1 M NaCl to 200 mM KCl the protein precipitated out of solution. As a consequence, once the sample was removed from denaturant the protein needed to remain in high salt or 1 M NaCl. This limited the type of resins that could be used for purification to either ion-exchange under denaturing conditions or gel-filtration under denaturing/high salt conditions. The resulting solubilized fraction from the guanidine extraction was observed to consist of the 38 kDa TCV CP and a 35 kDa protein which could not be separated from the sample by chromatography, either gel-filtration (S-200, S-1000) or ion-exchange (Q-Sepharose, HAP) under different buffer conditions, or by using a different denaturant, 8 M urea, extraction procedure with a CM-Sepharose column under denaturing conditions (see Figure 26). The resins SP and Q-Sepharose and HAP were also tested under denaturing, 8 M urea, conditions but no separation of the CP from the 35 kDa protein was achieved.

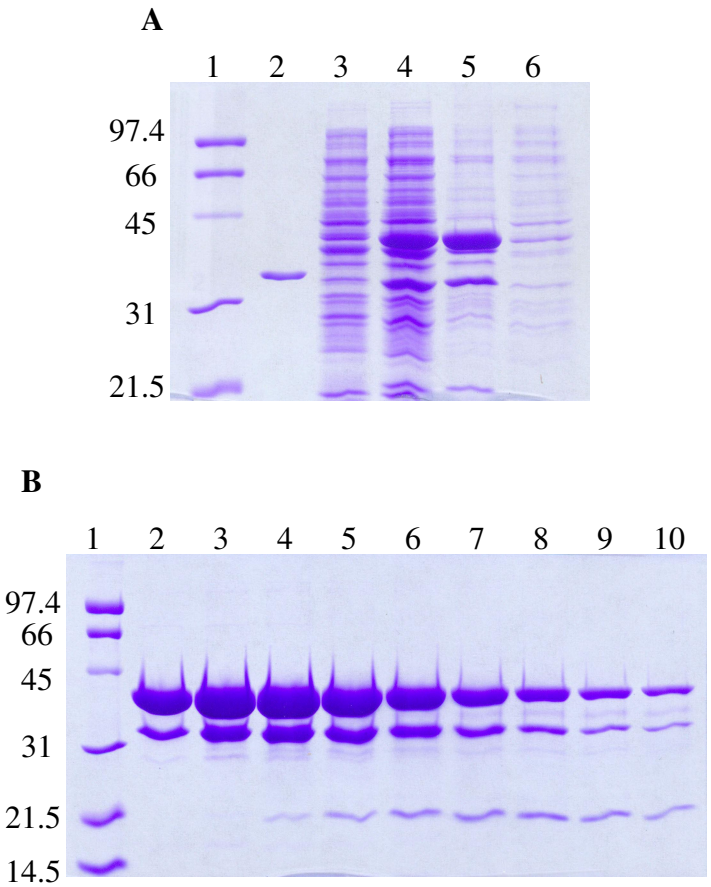

Figure 26. All samples taken during the 500 mL protein preparation trial of pET-TCV CP-MS/BL21λ(DE3) were run on 12.5% polyacrylamide denaturing gels for analysis. Gel (A) represents the samples taken from the expression and urea extraction procedures while gel (B) indicates the fractions # 12 - 20 collected during the denaturing CM Sepharose column. The samples indicated in each lane are as follows for gel (A): Lane (1) low range standard, (2) BirA marker (3) uninduced, (4) induced, (5) post lysis pellet, and (6) post lysis supernatant.

c. Characterization of the 35kDa protein

Since the two proteins were inseparable with these techniques, the 35 kDa protein was characterized using both N-terminal sequencing performed by Macromolecular Resources at Colorado State University and mass spectroscopy performed by Dr. Kristy Brown. It was suggested that the protein was a cleavage product of the TCV CP since proteases are known to cleave the N-terminal region of the CP and the 35 kDa protein

was behaving similar to the target. The N-terminal sequencing indicated that the 35 kDa protein contained a N-terminus with the first seven amino acids MENDPRV, which is identical to the N-terminus of the CP amino acid sequence. Mass spectroscopy was performed with both protein samples digested with chymotrypsin. The mass of each protein fragment was compared to the actual predicted digestion fragments for the CP-MS sequence. The data indicated that the 35 kDa protein contained the same sequence as the TCV CP-MS except for a missing region at the C-terminus starting at the lysine at position 325 in the coat protein sequence (see Figure 27).

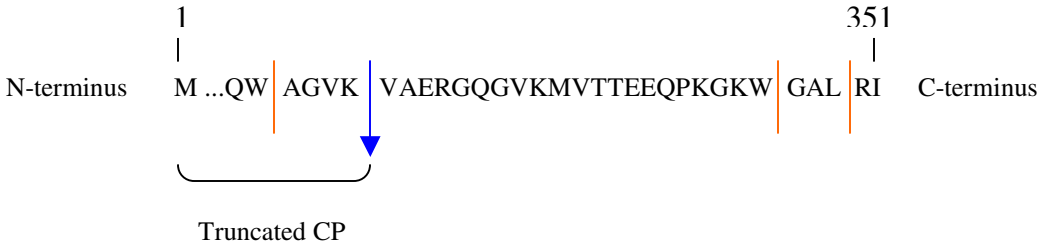

Figure 27. Chymotrypsin digestion of 35 kDa protein. The protein was initially digested with chymotrypsin producing fragments that were then weighed with mass spectroscopy. The red lines indicate the predicted sites of cleavage for the TCV CP-MS sequence. Analysis determined that the 35 kDa protein was a truncated form of the CP (amino acids #1-325) missing part of the C-terminus. The blue arrow indicates the end of the weighed fragments.

From these data the calculated theoretical pIs for both the 38 kDa and 35 kDa proteins were determined to be 9.30 and 9.15, respectively. These values were almost identical making separation by ion-exchange chromatography impractical, while the size similarity between the proteins eliminated the possibility of separation via gel-filtration. All

attempts at separation using ion-exchange and gel-filtration chromatography were unsuccessful with either denaturant extraction method and a new approach to the problem had to be devised.

d. Expression and purification of the CP from the *E. coli* strain Rosetta(DE3)

The logical next step was to determine the reason for the presence of the 35 kDa protein and find a way to inhibit its production *in vivo*. These included two main ideas. Either there was an early termination site in the coat protein sequence, where the AAG codon for the K325 was being interpreted as a UAG stop codon, or that there were not enough rare transfer RNAs present in the BL21λ(DE3) *E. coli* strain population to allow for complete translation. Analysis of the TCV CP DNA sequence allowed identification of 30 rare codons, which with the low rare tRNA population could result in translational stalling and premature translational termination. Based on this data the coat protein gene was expressed in a different bacterial cell strain, Rosetta(DE3), that provided a larger pool of these rare tRNAs. A series of induction tests at the optimal growing conditions resulted in expression of the 38 kDa coat protein without the production of the 35 kDa early translational termination product (see Figure 28).

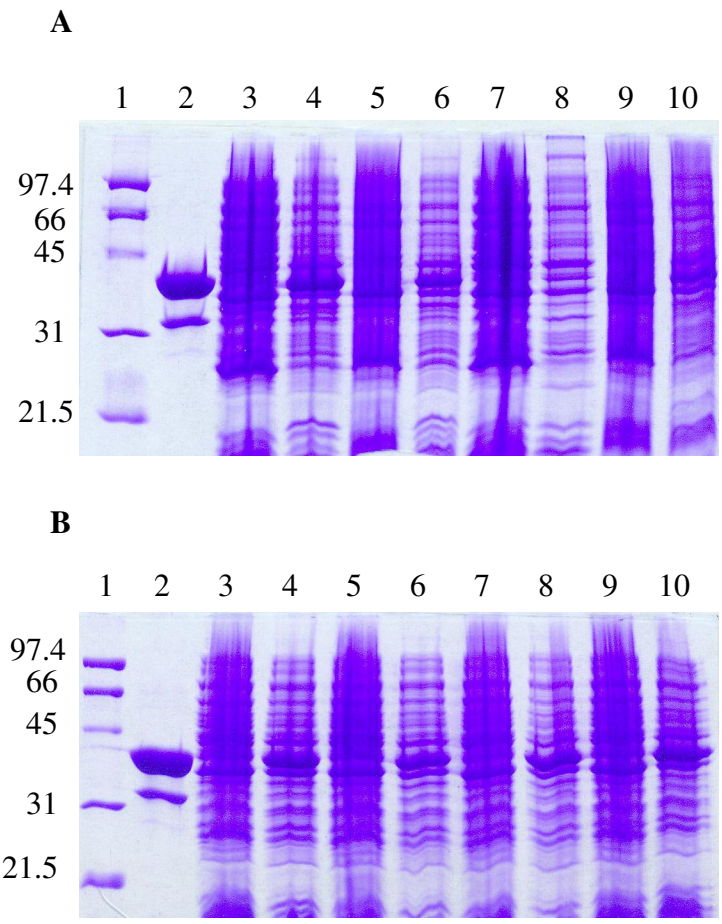

Figure 28. Induction tests of pET-TCV CP-MS transformed into Rosetta(DE3) grown at 37°C (gel A) and 25°C (gel B) with a 4 hour induction of the CP gene starting at a cell density with an OD_{600} of 0.7 to 0.8. Four cultures (# 1 – 4) were grown at each temperature with induction by addition of IPTG to a sample from each. For each gel the samples are as follows: Lane (1) low range protein standard, (2) CM sepharose fraction from pET-TCV CP-MS/BL21λ(DE3) prep, (3) culture #1 uninduced, (4) culture #1 induced, (5) culture #2 uninduced, (6) culture #2 induced, (7) culture #3 uninduced, (8) culture #3 induced, (9) culture #4 uninduced, and (10) culture #4 induced.

A 500 mL protein pET-TCV CP-MS/Rosetta(DE3) preparation test using the same initial parameters as described for BL21λ(DE3) was implemented. The protein was extracted with an 8 M urea extraction procedure and ion-chromatography performed on a CM Sepharose column under 8 M urea denaturing conditions. The production of the 35 kDa

protein had been eliminated and purification of the 38 kDa TCV CP was >98% (see Figure 29). HAP chromatography under denaturing 8 M urea conditions was tested to help increase the purity of the CP. Although the CP bound to the column and eluted, the purity of the sample did not increase and there was an overall dilution of the protein (see Figure 30). This presented a problem since the CP under these conditions could not be concentrated without loss in the form of aggregates. Therefore the HAP column was not used as an additional step in the purification procedure.

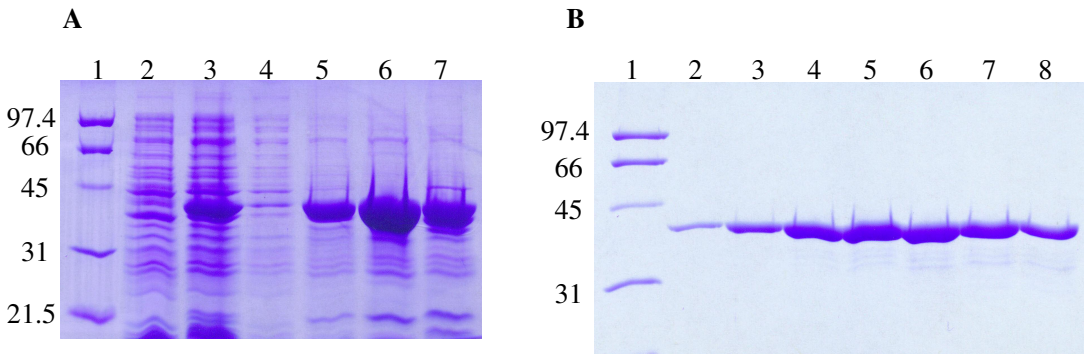

Figure 29. All samples taken during the 500 mL protein preparation test of pET-TCV CP-MS/Rosetta(DE3) were run on 12.5% polyacrylamide denaturing gels for analysis. Gel (A) represents the samples taken from the expression and urea extraction procedures while gel (B) indicates the fractions # 18 - 24 collected from the denaturing CM Sepharose column. The samples indicated in each lane are as follows for gel (A): Lane (1) low range standard, (2) uninduced, (3) induced, (4) post lysis supernatant, (5) post lysis pellet, (6) extraction supernatant, and (7) extraction pellet.

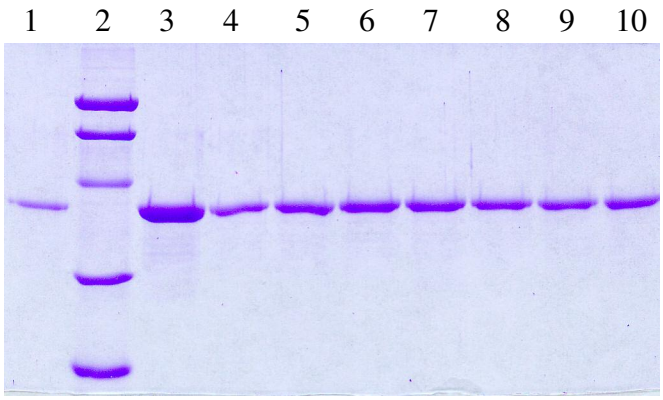

Figure 30. The purified CP from the CM Sepharose pooled fractions of the 500 mL pET-TCV CP-MS/Rosetta(DE3) protein preparation test was run on a denaturing urea HAP column. The fractions # 33 - 40 collected from this column were run on 12.5% polyacrylamide denaturing gels for analysis. The samples indicated in each lane are as follows: Lane (1) fraction # 33, (2) low range standard, (3) CM Sepharose pooled fractions, (4 – 10) fractions # 34 – 40.

The purified protein from the CM Sepharose column was renatured via step-wise dialysis out of urea and into a high salt storage buffer. Spectroscopic measurements at 280 nm indicated that the protein was free of both aggregates and nucleic acid. All attempts at decreasing the salt concentration of the CP sample resulted in precipitation and aggregation of the CP over time. It was determined that once the CP was removed from the denaturant it had to remain in a high salt buffer containing 1 to 0.5 M NaCl at a pH ≥ 8.0 at 4°C. A CP sample with a concentration of 11μM once dialyzed into 0.5 M NaCl buffer pH 8.5 at 20°C contained a small fraction of protein that precipitated out of solution. Therefore all samples were stored at -70°C in a 1M NaCl high storage buffer pH 8.5 at 4°C supplemented with 5% glycerol.

3. Characterization of TCV CP

a. CP folding

Circular dichroism was used in order to determine if the renatured TCV CP had folded properly. Data were collected with the protein in a 100 mM sodium phosphate buffer and 100 mM sodium phosphate buffer containing 50 mM Tris-HCl, 1 mM EDTA, pH 8.5 at 4°C (see Figure 31). Qualitative data analysis (200-250 nm) indicated that the protein was folded and contained the expected secondary structures, β-sheet and α-helix. The lower wavelength range (180-200 nm) was not analyzable due to buffer interference and low protein concentration.

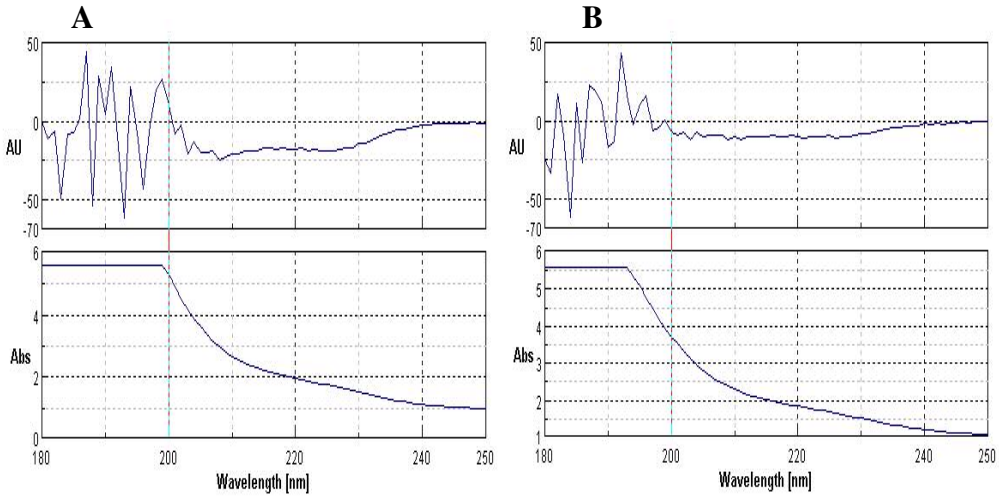

Figure 31. CD spectra of purified TCV CP in (A) sodium phosphate, Tris buffer (100 mM NaH_2PO_4/Na_2HPO_4, 50 mM Tris-HCl, 1 mM EDTA, pH 8.5 at 20°C) and (B) sodium phosphate buffer (100 mM NaH_2PO_4/Na_2HPO_4, pH 8.5 at 20°C). The wavelength region 180-200 nm cannot be analyzed due to buffer interference.

The protein concentration could not be increased above a concentration of about 10 μM without precipitation. Since buffers especially those containing sodium chloride can interfere with measurements at the lower wavelengths a range of buffers were tested for dialysis against the CP. The protein was dialyzed into different concentrations of sodium phosphate buffers containing 50 mM Tris-HCl and 1 mM EDTA with some success, although it was observed that these buffers still interfered with the lower wavelength measurements and protein solubility. The CP precipitated out of solution either as the concentration was increased, the temperature was varied or the ionic strength of the buffer was decreased. Therefore it was determined that sodium phosphate was not an ideal buffer for CD analysis of the CP due to both its inability to buffer the higher pH region above 7.8 and its continued interference with the instrumentation. The protein had the highest solubility when present in a buffer containing a high concentration of sodium chloride.

In order to attain measurements in the far UV wavelength region the protein sample was in a 0.1 cm path length cuvette, which dictated the higher level of CP needed to obtain a clear signal. Since the protein concentration could not be increased without increasing the buffer interference and maintaining protein solubility it has been suggested to run CD measurements in the near UV region from 350 to 250nm where sodium chloride would not interfere with the protein signal. Within this range the level of the individual secondary structures present in the sample cannot be quantified but it is possible to compare the spectrum of the renatured recombinant CP with protein isolated directly from virions. Overlap of the two spectra would indicate that the recombinant CP

had folded properly. Some of the advantages of obtaining measurements from this region include the ability to use sodium chloride as a buffer component, allowing a higher concentration of soluble protein to be maintained during the measurement process.

b. CP assembly state in solution

Sedimentation experiments, both velocity and equilibrium, were used in an effort to characterize the assembly state of the protein in solution. High salt buffer was used for these measurements. Velocity sedimentation data analysis indicated the presence of monomer and dimer along with other larger aggregates. The data from equilibrium sedimentation experiments could not be fit to a model and must be repeated. There were some obvious problems with the data fit during the analysis of both the velocity and equilibrium experiments, which was attributed to the fitting programs. Velocity sedimentation was repeated with protein in a 0.5 M NaCl buffer and a portion of the data were fit to a 1 species model, which indicated the presence of monomer (see Figure 32). Fitting of earlier timed scans indicated monomer, dimer and possibly larger aggregates. This indicates that over time the larger protein aggregates migrated to the bottom of the cell. There were still some problems with data fit and salt effects. These experiments will need to be repeated under different conditions (low salt and slower speed). The CP had been dialyzed into 1, 0.5 and 0.25 M NaCl buffer containing 100 mM Tris-HCl, 5 mM EDTA, pH 8.5 at 20°C. As the concentration of the sodium chloride was decreased the dialyzed CP started to form aggregates and precipitate out of solution. This was first observed from data analysis of the measurements collected from the sedimentation experiments using these buffers and was later confirmed with dialysis trials.

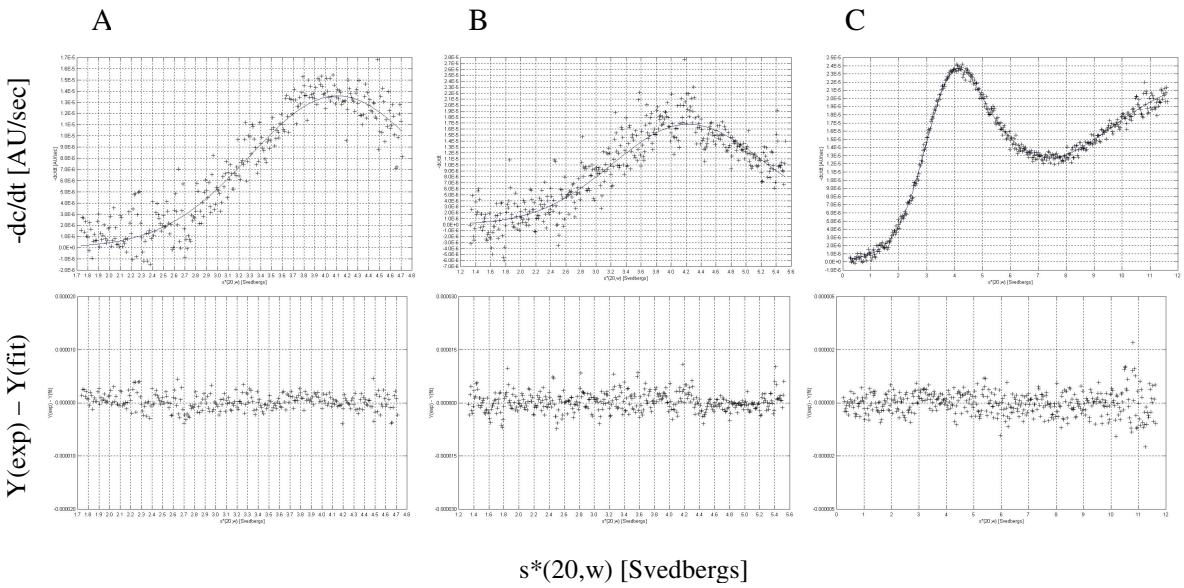

Figure 32. Velocity sedimentation data of TCV CP in 0.5 M NaCl buffer (0.5 M NaCl, 100 mM Tris-HCl, 5 mM EDTA, pH 8.5 at 20°C) was obtained for a total of 90 scans. The later scans were fit to a 1 species model using (A) scans 76-90 and (B) scans 61-70. Data in graph (A) was fit to a monomer of molecular weight 42.26 kDa [37.45, 46.77] while (B) was fit to a monomer of 36.34 kDa [33.57, 39.12]. Earlier scans were fit to a 3 species model using (C) scans 25-60. Data in graph (C) was fit to possibly monomer and dimer: 69.33 kDa [60.65, 76.11], 26.38 [23.55, 28.58] and 7.49 kDa [6.55, 9.43]. The top graphs show the time derivative -dc/dt analysis with the distribution of -dc/dt as a function of the sedimentation coefficient. The distribution of residuals are shown in the bottom panels.

C. Discussion

1. Cloning the CP gene

Characterization and assembly of the CP, requires a large amount of material obtained from dissociated virions. The CP gene was therefore cloned and the recombinant CP expressed and purified from *E. coli*. The CP DNA sequence was first amplified using PCR and the purified DNA was then inserted into a pET-17xb plasmid under control of the T7 promoter. Once the DNA was amplified *in vivo* and sequenced,

analysis of the data indicated the presence of both silent and missense mutations. These mutations were produced as a result of PCR mistakes and/or cross-contamination. Problems with PCR may be attributed to a wide-range of factors including old dNTPs, buffers, and varied temperatures. Error may be attributed to using the PCR products as template for amplifying the DNA, in which case carry-over contamination would be responsible. Contamination of the original TCV genomic RNA sample used in performing reverse transcriptase PCR (RT-PCR) was most likely the main contributor to the mutations in the CP DNA sequence. There is also the possibility that the Taq polymerase introduced the mistakes into the DNA sequence since it lacks 3' to 5' proofreading exonuclease activity. A polymerase that contains a proofreading mechanism such as Pwo or Pfu could have been used instead to limit mutations (*63*). All missense mutations were corrected with site-directed mutagenesis producing both TCV CP coding sequences MS and B.

2. Early Translational Termination

Initially the recombinant CP was expressed in the BL21λ(DE3) cell strain as previously performed with other plant recombinant CPs (*61, 62*). Inducing the CP gene resulted in co-expression of the full-length CP with a translational termination product that could not be separated from the target protein. Synthesis of this truncated product was correlated to the presence of specific codons in the viral plant gene. Eukaryotic organisms use a specific subset of the 61 available codons with a frequency different from that found in prokaryotes. Those major codons that occur in highly expressed *E. coli*

genes are used with higher frequencies than rare codons present in genes expressed at relatively low levels within *E. coli*. The codons with a low frequency of use in *E. coli* can be found to occur with a higher frequency in other organisms. As a result most *E. coli* strains do not contain the necessary concentration of transfer RNAs (tRNAs) needed for expression of the heterologous protein from its mRNA (*64*). Analysis indicated the presence of seven rare codons used repeatedly throughout the CP coding sequence. The DNA contained a total of 30 codons that *E. coli* normally does not require the appropriate tRNAs for in order to keep up with translation (see Figure 33). This resulted in the synthesis of the 35 kDa translational termination product missing the C-terminus of the CP starting after K325 (see Figure 34). The early termination may be correlated to either the presence of the downstream rare codon cluster AGG/GGA, such clusters have been shown to terminate translation, or the misinterpretation of the AAG lysine codon as a UAG stop codon (*64*). Two different approaches to eliminate synthesis of the truncated product included either replacing all 30 of the rare codons with major codons or transforming the plasmid into an *E. coli* strain designed to supply the necessary tRNAs for complete translation. Since it would be too time consuming to change all the appropriate codons, a different strain was implemented in the expression process. The protein was therefore expressed in the Rosetta(DE3) strain (Novagen) which supplied six of the seven rare tRNAs needed for complete translation of the CP gene (see Figure 33). It was not necessary to change the four rare CGA codons, which the corresponding tRNAs were not supplied by the Rosetta(DE3) strain, since expression in the new strain eliminated production of the 35 kDa protein.

Rare Codons	Encoded Amino Acids	Frequency per 1000 codons	Rosetta(DE3)	TCV CP-MS
AGG	Arg	1.4	*	4
AGA	Arg	2.1	*	5
CGA	Arg	3.1	-	4
CUA	Leu	3.2	*	1
AUA	Ile	4.1	*	2
CCC	Pro	4.3	*	5
CGG	Arg	4.6	-	-
GGA	Gly	7.0	*	9

Figure 33. A table indicating the codons used by *E. coli* at a frequency of <1% and the amino acids encoded by these rare codons. The * identifies those codons that Rosetta(DE3) supplies tRNAs for. The rare codons found in the TCV CP sequence are recorded by the number of each present (*61*).

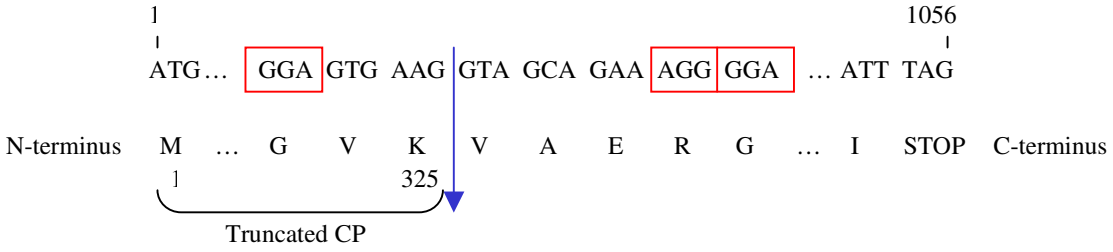

Figure 34. This represents the site of truncation for the 35 kDa translational termination product.

3. Expressed CP insolubility

After expression the CP was still found in the insoluble inclusion bodies within the post-lysis pellet. This insolubility is not uncommon in the overexpression of recombinant proteins. Normally in the cell the native fold and solubility of the expressed protein is mediated by chaperone proteins. Solubility is influenced by both the protein's affinity for chaperone(s) and the ratio of the two present in the cell. A high level of expressed protein may exceed the chaperone system's capacity allowing some overexpressed protein to sequester all available chaperones while the rest has a chance to misfold and consequently become packaged into inclusion bodies (*65*). The solubility level may also be dictated by the protein structure. It has been observed that the capsid protein has a high affinity for self-association resulting in aggregation and association with other RNA, decreasing the chance of the protein staying soluble for an extended time in the cell. The basic positive charge associated with the protein N-terminus attributes to the overall protein insolubility. It has been observed that exported proteins translocated across the cytoplasmic membrane in gram-negative bacteria have either a net positive or neutral charge at the N-terminus. When the net charge increases in positive charge the exportation of the protein across the cytoplasmic membrane decreases (*66*). Studies have suggested that a net positive charge on a 18 to 30 residue sequence at the protein N-terminus will decrease the degree of translocation across the membrane (*66, 67*). The first 41 residues of the CP basic N-terminus R domain has a net charge of +5 while the first 30 amino acids contains a +3 charge. This suggests that the positive charge on the CP N-terminus prevents protein secretion into the cytoplasm. From the CP structure and behavior it was predictable for the overexpressed protein to remain

insoluble in the post-lysis pellet. Once the cells were lysed, as with CCMV CP, the recombinant TCV CP was extracted via 8 M urea (*61*). Protein was then purified under these same denaturing conditions on a CM Sepharose column (*56*). The fractions containing purified CP were pooled and dialyzed step-wise out of urea and into a high salt 1 M NaCl buffer. From experimentation it had been observed that the CP starts to aggregate and precipitate out of solution as the salt concentration is decreased below 1 M NaCl. Therefore the CP was stored in a 1 M NaCl high salt storage buffer pH 8.5 at 4°C supplemented with 5% glycerol.

4. CP Structure

a. CP interactions in solution

Before the CP could be used in assembly experiments it needed to be determined if the recombinant CP had attained its native conformation once it was dialyzed out of the denaturant. The CP was partially characterized with respect to its renatured folded and assembly state in solution using both CD and ultracentrifugation measurements, respectively. Both CD and sedimentation analysis exhibited limitations in characterizing the protein. From sedimentation it was observed that the protein in a 0.5 M NaCl buffer consisted of monomers and larger CP aggregates. This was not unexpected since the CP's main function is to assemble into the larger capsid structure and such aggregation has been observed with other plant capsid proteins such as *Tobacco mosaic virus* (*19*). It has been previously observed that TCV dissociated under mild conditions, 0.5 M NaCl, 0.1 M Tris, 0.01 M EDTA, pH 8.3, will produce CP dimers in solution. This has been

previously determined both with gel-filtration chromatography, Ultrogel ACA-54, under these conditions and DMS cross-linking of the dissociated TCV (*8*). The protein structure also suggests dimerization due to the P domain's susceptibility for hydrophobic interactions with polar groups forming hydrogen bonds and salt bridges (*68*). The formation of these dimers through interaction between the P domains was observed by DMS cross-linking experiments (*8*). Analysis of the surface contacts between adjacent P domains indicate that TCV has a high percentage of hydrophobic and aromatic interactions, with a lower percentage of polar and charge interactions. These percentages were 57, 29 and 7 %, respectively. These contact residues are not conserved between different viruses even within the same family as observed with *Carnation mottle virus* (CarMV), *Tomato bushy stunt virus* (TBSV), and TCV from the family *Tombusviridae* (*69*). CP P domain deletions have resulted in both a lack of assembly and systematic movement (*13, 70*). The P domain and its dimerization is essential for viral assembly. The CP has been observed to form dimers in solution under certain buffer conditions and concentrations that do not favor further association (*8*). Since a reasonable amount of protein is required for sedimentation experiments this concentration may be too high to exhibit only dimerization of the CP. With the increase in the concentration there is a higher probability for the presence of larger CP assembly states. The concentration threshold for the transition of the CP association from dimers to larger aggregates needs to be identified for future experimentation.

b. Native Conformation

The folded state of the recombinant CP was also not completely defined although the CD spectra in the far UV wavelength region contained the characteristics of β-sheet and α-helix, both of which are found in the known protein structure. The spectra did not contain the characteristic for random coil, which would have indicated that the protein was still denatured. It is not known if the protein had folded into its native conformation. The native CP structure consists of mainly β-sheet where the S domain contains the eight stranded anti-parellel β-sheet in the β-barrel jelly roll motif with 2 small α-helices and the P domain contains ten β-strands (*69*). The R domain is unstructured in unassembled CP and as a result it has not been able to be crystallized (*4*).

In order to compare the folded structure of both CPs the spectra from the CD analysis of the proteins in the near UV wavelength region must be analyzed for overlap. If the spectra are identical then the recombinant CP has been folded properly and will function/assemble in the same manner as the virion isolated CP. It is important that the protein folds properly so that the putative calcium-binding site (CBS) is formed. Each pair of CP subunits must bind a calcium ion in order to form and stabilize the virion structure. The CBS is located adjacent to the 5 amino acid hinge between the S and P domains. Each Ca^{2+} is bound by the 3 residues Asp155, Asp157 and Lys160 from one subunit and the two residues Gln127 and Asp199 from the other (*9*). It has been suggested that the CBS is stabilized by two salt bridges: one cation-mediated salt bridge and the other a Lys-Asp salt bridge (*12*). The presence of the CBS allows the virus to disassemble and release its RNA in the low calcium environment of the cytoplasm (*9*).

Therefore the folded structure of the S domain with the CBS is required for assembly. To determine if the recombinant CP contains the CBS site virions and CP needed to be isolated and purified from infected plant tissue before beginning the investigation into the TCV assembly mechanism.

IV. Virion and Coat Protein purification from plants

A. Introduction

Since the recombinant CP was obtained and purified from *E. coli* it had to be characterized both with respect to its folded and assembled state in solution. Comparison between the CP isolated from virions and recombinant CP would determine if the recombinant CP could reassemble *in vitro*. Therefore, before characterization can continue virions and CP needed to be extracted and purified for analysis.

Previously CP has been purified from virions isolated from infected tissue of both turnip (*brassica campestris*) and *Arabidopsis thaliana* plants. The plants were inoculated with TCV and once symptomatic, harvested for the isolation/purification of the virions. The purified virions were then dissociated allowing for the extraction and purification of the CP. The Morris lab developed a protocol for the purification of both the virions and CP from infected turnip plants (*37*). The original chromatography resin Ecteola cellulose (Sigma) involved in the purification of the virions was substituted with a DEAE-agarose gel A resin (Bio-Rad) by the Simon lab (*37, 51*). The remainder of the virion and CP purification followed Morris' protocol (*37*).

B. Results

1. Virus purification

Virions were purified from plant tissue using a modified version of the purification procedure from the Morris lab (*37*). Plants were grown as previously described although the time and part of the plant to be harvested was not well defined in the previous method. Leaves were not harvested until at least 28 days past inoculation (dpi) when the newer leaves had completely developed. These leaves were severely crinkled and stunted while the older leaves became yellowed with visible lesions. Unlike the original method by Morris, in which the whole turnip plant was homogenized with extraction buffer, only the newer leaves were harvested (*37*). The main vein from each leaf was removed with a razor blade prior to weighing. This allowed for easier pulverization of the plant material into a homogenous mixture with the extraction buffer containing 0.2 M sodium acetate pH 5.2, β-mercaptoethanol and protease inhibitors.

The method of extraction and chromatography followed the protocol described in the Experimental section. After extraction the virions were precipitated using a 40% PEG/ 1 M NaCl solution and centrifuged producing a dark green and white pellet. This pellet was resuspended in resuspension buffer containing 0.05 M sodium acetate pH 5.2, incubated on ice and centrifuged again in order to remove large debris present in the sample. The light green supernatant containing the virions was then pelleted using the ultracentrifuge and resuspended with column buffer in preparation for purification on a

DEAE-agarose gel A column (see Figure 35). The virions were bound to the column, washed with column buffer containing 0.01 M sodium phosphate pH 7.0 with 2.5% glycerol and finally eluted with a linear sodium chloride gradient. Fractions containing pure virions, identified by absorbance at 260 nm and CP migration on SDS-polyacrylamide gels, were pooled and the virions precipitated with PEG (see Figure 36).

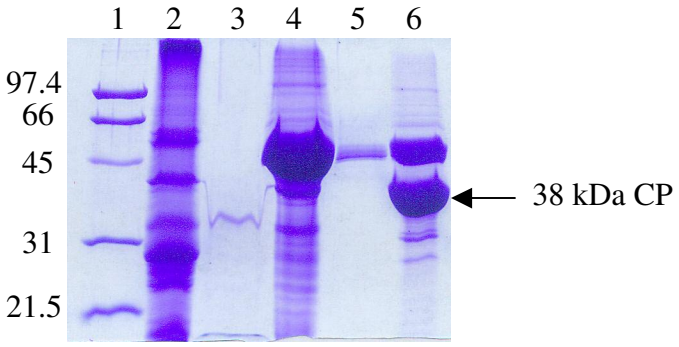

Figure 35. The TCV purification samples run on a 12.5% SDS-polyacrylamide gel. The samples were loaded onto the gel as follows: Lane (1) low range standard, (2) pellet # 1, (3) supernatant # 1, (4) pellet # 2, (5) supernatant # 2, and (6) resuspended pellet. The pellet/supernatant were produced as indicated: (# 1) after the first sample centrifugation at 7, 300 rpm for 1 hr. at 4°C, and (# 2) after addition of the 40% PEG/1 M NaCl solution and centrifugation. The resuspended pellet was loaded onto the DEAE-agrose gel A column for purification.

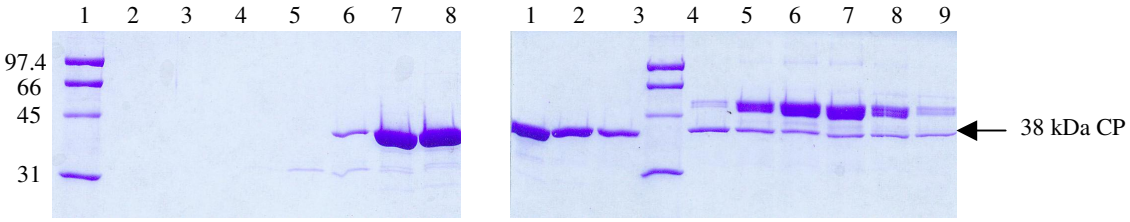

Figure 36. Fraction samples from the DEAE-agarose gel A column were run on 12.5% SDS-polyacrylamide gels. The fractions # 36 – 52 were run on these gels although only the fractions # 42 – 46 (Gel A lanes # 8, 9 and Gel B lanes # 1-3) contained purified TCV.

The method precipitated the virions by addition of ¼ volume 40% PEG/1 M NaCl solution with mixing, and incubation on ice for 30 minutes followed by centrifugation. The pellet was then resuspended in a 2 mL volume of 50 mM Tris-HCl pH 8.5 at 4°C supplemented with 0.01% sodium azide. It was determined by the absorbance at 260 nm that the pooled fractions contained a total of 5.55 mg of virions. After precipitation the resuspended pellet contained only 0.804 mg of virions, with the remainder still present in the supernatant. In order to extract all the virions the PEG precipitation procedure was repeated and the combined total of both trials resulted in 4.82 mg of isolated virions from the initial 5.55 mg of virions. In a second preparation the following modification was used: addition of ½ volume of 40% PEG/1 M NaCl solution with mixing and incubation overnight at 4°C followed by centrifugation as previously described. The DEAE-agarose gel A pooled fractions containing 11.97 mg of virions once precipitated and resuspended using the modifications resulted in a solution containing 10.554 mg of virions. The overall loss was decreased significantly by increasing the ratio of PEG added to the sample volume. The resuspended pellet was checked on both SDS-polyacrylamide and agarose gels. The ethidium bromide stained agarose gel indicated that there was other nucleic acid present in the resuspended virion sample. Treatment of the TCV sample with either DNase I or RNase resulted in identification of the nucleic acid as additional RNA (see Figure 37). The presence of this nucleic acid may be due to the increased PEG concentration in the precipitation. Previously the final resuspended pellet was not checked on agarose gels using ethidium bromide staining, so there is no precedent for comparison. Since the sample contained additional nucleic acid besides the genomic

RNA the concentration calculations dependent on the absorbance level at 260 nm were not accurate.

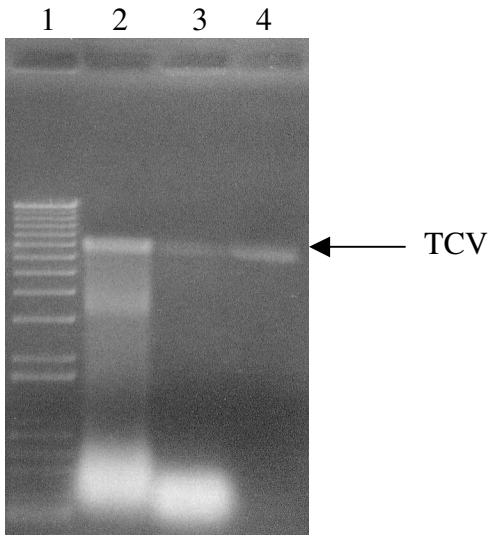

Figure 37. Purified TCV on a native 0.6% High Strength Agarose gel. The samples were loaded as follows: Lane (1) 1 Kb marker, (2) TCV purified, (3) TCV treated with DNase I, and (4) TCV treated with RNase.

Virion samples stored at 4°C for several months resulted in noticeable degradation of the capsid over time. The samples could not be stored at -20°C due to degradation of the virion capsid structure. Images of negatively stained virion samples on carbon coated formar grids were obtained using transmission electron microscopy (TEM) and captured on film (see Figure 38). A virion of ~ 30nm diameter was calculated from these images and agrees with the known capsid size. These data can be compared in future experimentation to the images of reassembled virions from recombinant CP.

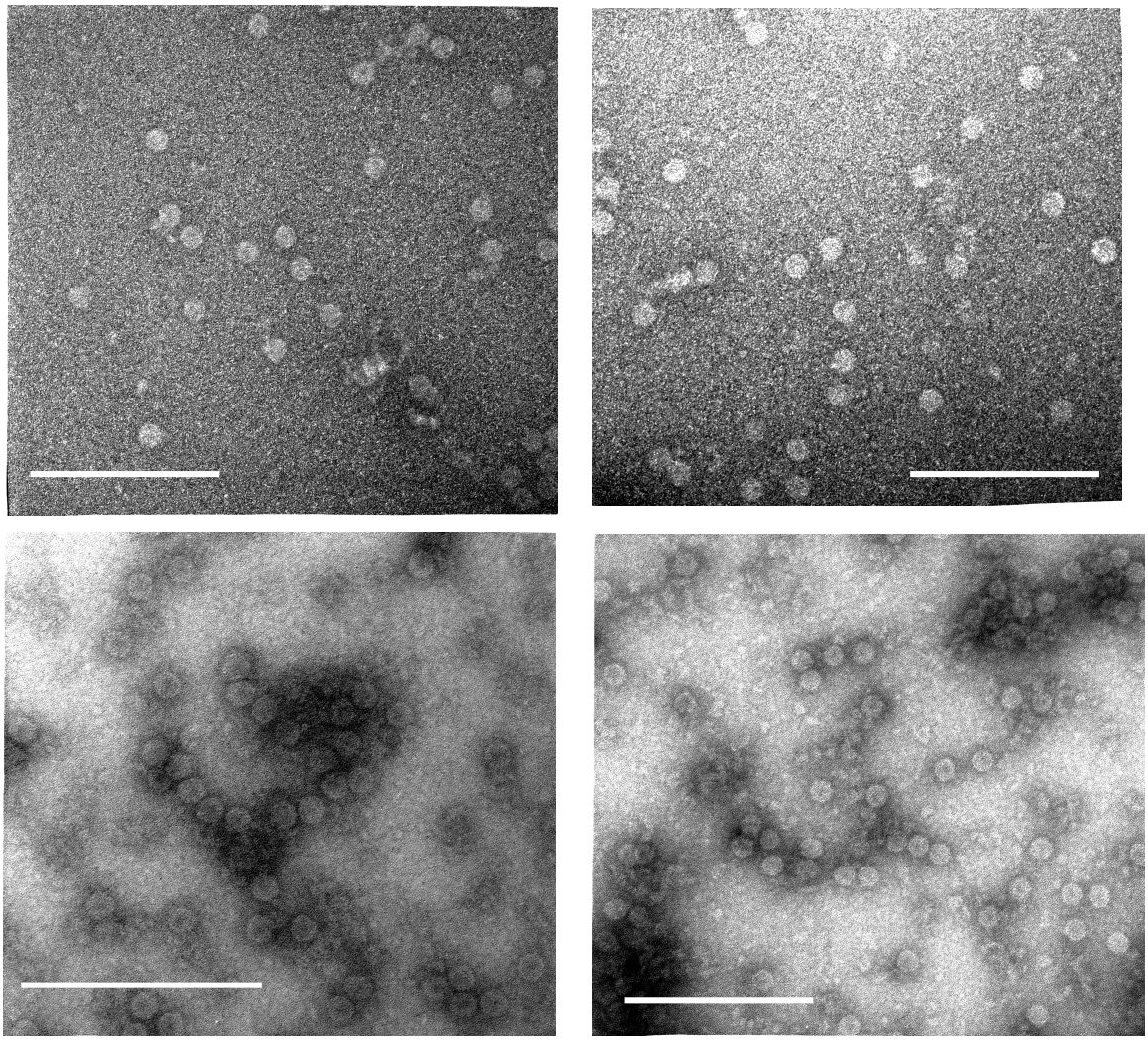

Figure 38. TEM images of purified TCV virions 0.2 to 0.4 mg/mL. The bars represent 300 nm.

2. Coat protein purification from virions

The CP was purified from the dissociated virions following the method by the Morris lab (*37*). After the virions were dissociated a few modifications in the method were implemented in order to increase the efficiency of the procedure. After the virions were dissociated and precipitated with a 5% PEI solution the sample was centrifuged for 10 minutes at 13,000 rpm in a microcentrifuge and the supernatant loaded onto the Sephacryl S-200HR gel-filtration column. However, it has been observed that CP will aggregate as the concentration of sodium chloride is decreased below 1 M. In order to increase the efficiency of separation the 0.5 M NaCl was replaced with a 1 M NaCl buffer. The column was run with the 1 x dissociation solution containing 1 M NaCl, 0.1 M Tris-HCl, 5 mM EDTA, pH 8.5 at 4°C. Once the fractions containing CP were pooled (see Figure 39) instead of reprecipitating the CP, resuspending the pellet and dialyzing the sample 16 hours against two changes of a 0.01M Tris-HCl, 25 mM NaCl, pH 8.0 buffer as indicated in the original method, the pooled fractions were dialyzed for 3 hours against two changes of a high salt storage buffer containing 50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, 5% glycerol, pH 8.5 at 4°C. This is the same high salt storage buffer used for recombinant CP. From an initial sample containing 4.22 mg of virions the amount of CP successfully extracted and purified was about 2.81 mg. This is close to the theoretical value of 3.54 mg which is based on the assumption that 84% of the virion is composed of CP. The difference between the two values may be attributed to either a loss of the CP during precipitation of the RNA or an incorrect estimate of the initial virion concentration.

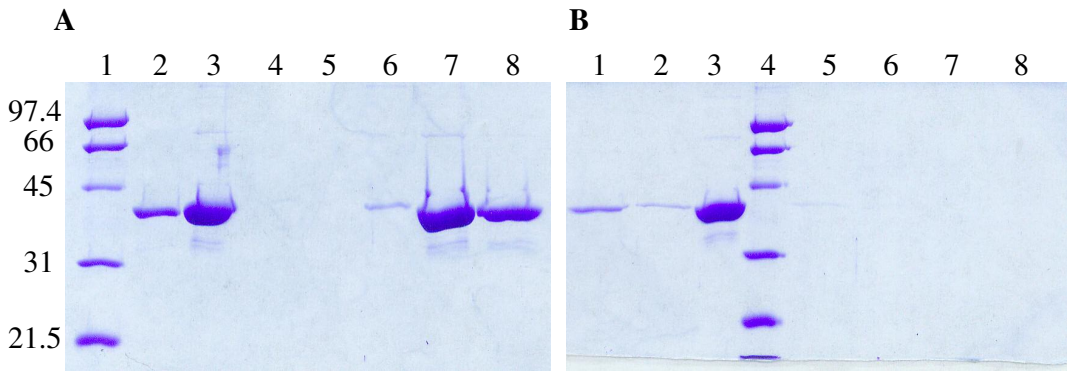

Figure 39. All samples from the fractions # 4 – 14 of the gel-filtration Sephacryl S-200HR column were run on 12.5% polyacrylamide denaturing gels for analysis. The CP eluted off the column in fractions # 6 – 10 although only fractions # 7 and 8 contained the majority of the protein. Samples were loaded as follows: Gel A lane (1) low range standard, (2) purified recombinant CP, (3) loaded supernatant, (4 – 8) Fractions # 4 – 8. Gel B lane (1 – 2) Fractions # 9 – 10, (3) loaded supernatant, (4) low range standard, (5 – 8) Fractions # 11 – 14.

3. TCV genomic RNA

a. Isolation of genomic RNA from virions

In order to perform assembly experiments both with the recombinant and functional CP genomic RNA needed to be isolated and purified. The Harrison lab developed a method for the isolation and extraction of the genome from dissociated virions (*34*). A sample containing purified virions was incubated in a buffer containing 1 M Tris-HCl, 5 mM EDTA, pH 8.0 preincubated with 10 µg proteinase K per mL. The EDTA caused the virions to swell and the proteinase K degraded the protein allowing for extraction of the interior RNA. The RNA was extracted twice with phenol: chloroform and the aqueous layers containing the nucleic acid were moved to a separate eppendorf tube. Any residual chloroform was extracted once with ether and the RNA was ethanol

precipitated and resuspended in a 1 x TE buffer. The purified RNA was checked on a native 1% agarose gel and although the RNA was observed to migrate as one main band above the 2 Kb marker there was evidence of RNA degradation (see Figure 40A). Even though solutions and equipment were treated for RNase prior to use degradation was still visible. There were two main concerns with this method including the presence of RNA degradation and the low concentration of purified RNA. Since only about 16% of each virion consists of the genome it would require a large concentration of virions to acquire the necessary amount of intact purified RNA for experimentation.

b. Synthesis and purification of genomic RNA

As an alternative the genomic RNA was synthesized using the pUC19/TCV plasmid cut with Sma I as the linearized DNA template. The genomic RNA was then synthesized using this template under control of the T7 promoter. Initially all solutions were prepared in the lab using DEPC-treated autoclaved distilled water. The RNA was synthesized in a 100 µL reaction volume containing 3 µL 1 M Tris-HCl pH 8.0, 16.5 µL 150 mM $MgCl_2$, 2 µL 100 mM Spermidine, 10 µL 100 mM DTT, 1 µL Triton X-100, 8 µL of each 50 mM dNTP, 16 µL 5 x T7 polymerase buffer, 13 µL linear plasmid TCV DNA at 1µg/µL, 2 µL T7 RNA polymerase and 2.5 µL RNasin. It was then incubated in a water bath at 37°C for 3 hours. Once run on a native agarose gel the RNA was observed to migrate as a main single band although after purification the RNA showed signs of degradation (see Figure 40B).

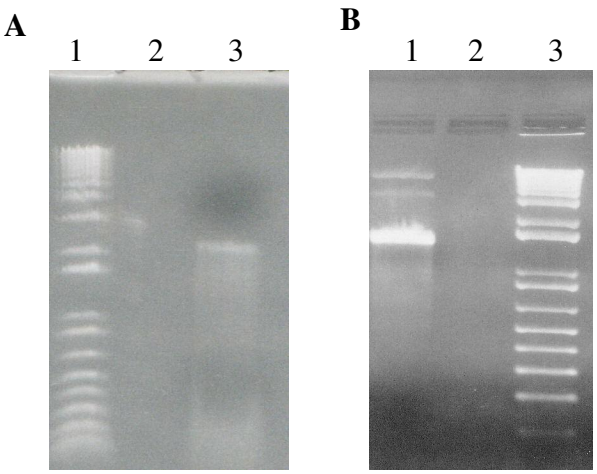

Figure 40. TCV genomic RNA was run on 1% native agarose gels. The RNA extracted from the virions was run on gel A while the RNA synthesized for plant inoculations was run on gel B. Both samples were run against a 1 Kb marker.

In an attempt to increase synthesis efficiency and limit degradation a MEGAscript kit (Ambion) was used in conjunction with RNase free tubes and tips (Ambion). As a result there was an increase in the concentration of the synthesized RNA, which was observed on denaturing 1% agarose-formaldehyde gels. Degradation of the RNA was still visible after purification (see Figure 41). Three different purification techniques were used in an attempt to decrease degradation. These included a lithium chloride precipitation, a phenol: chloroform extraction followed by an ethanol precipitation and the use of a MEGAclear kit (Ambion). All three procedures introduced degradation into the sample, which indicated that a common component, the 100% ethanol, was the most likely cause of the continued degradation (see Figure 42). Once the 100% industrial brand was replaced with 99.5% ACS grade ethanol, the lithium chloride and phenol: chloroform purification techniques had a noticeable decrease in the RNA degradation while the MEGAclear kit still showed signs of degradation (see Figure 43). Since the lithium

chloride precipitation technique is shorter and consists of fewer steps for potential loss of RNA it was chosen as the final purification method for the RNA.

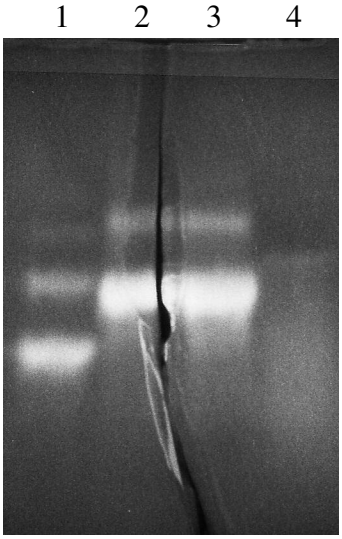

Figure 41. Samples of the genomic RNA from the MEGAscript and MEGAclear kit were run on a 1% denaturing fromaldehyde-agarose gel. Lane (1) positive control, (2) TCV RNA synthesized, (3) TCV RNA DNase I treated, and (4) TCV RNA purified with kit.

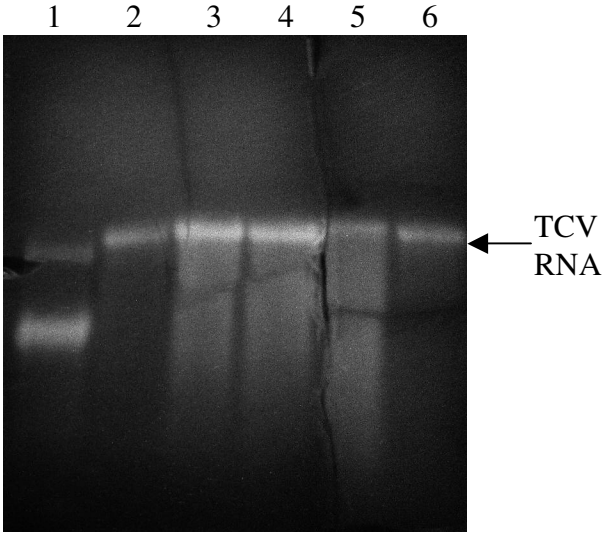

Figure 42. Samples of the synthesis and different purification techniques of the genomic RNA were run on a denaturing 1% Formaldehyde-agarose gel. The samples were loaded as follows: Lane (1) positive control, (2) TCV RNA DNase I treated, (3) LiCl precipitated, (4) phenol: chloroform/ ethanol precipitated, (5) MEGAclear kit, and (6) TCV RNA synthesized.

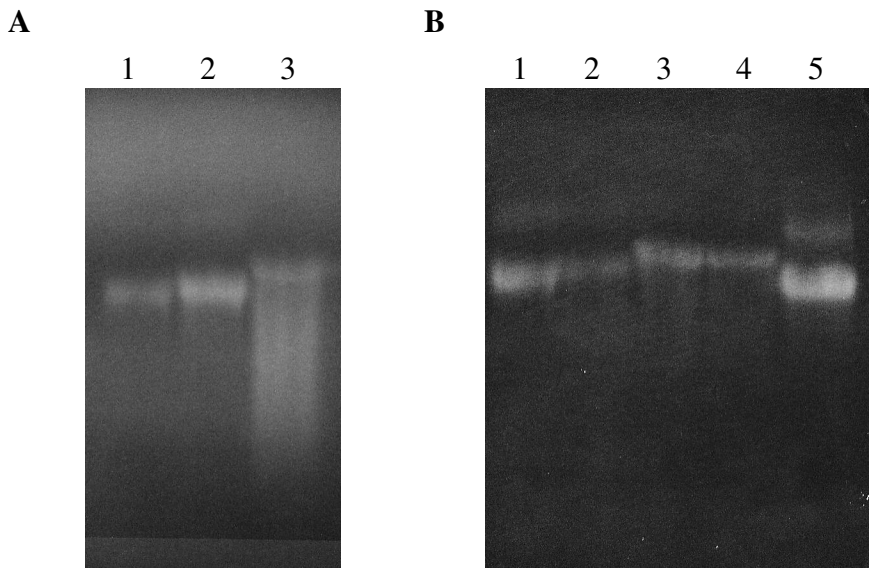

Figure 43. RNA samples from different purification procedures using the 99.5% ACS grade ethanol were run on 1% denaturing formaldehyde-agarose gels. Gel A lane (1) TCV RNA synthesized, (2) TCV RNA DNase I digested, and (3) TCV RNA purified with MEGAclear kit. Gel B lane (1) TCV RNA DNase I digested, (2) TCV RNA DNase I digested at room temp., (3) LiCl precipitated, (4) phenol: chloroform/ ethanol precipitated, and (5) TCV RNA synthesized.

C. Discussion

1. CP Purification

The CP has been successfully extracted and purified from infectious virions. A few problems with the method were identified and modified for both virion isolation and CP purification. The procedure for virion purification was essentially the same except for the ratio of 40% PEG/ 1 M NaCl added to the final purified sample to reprecipitate the virions. This volume was doubled since it was observed that the majority of the virions were retained in the supernatant upon addition of a ¼ volume of 40% PEG solution. This

volume was increased to an addition of ½ volume of 40% PEG, which succeeded in precipitating all of the virions. Analysis identified the presence of other nucleic acid in the final purified sample, which affected the concentration calculations based on the absorbance at 260 nm. This nucleic acid was identified as supplementary RNA since it was successfully degraded with the addition of RNase. There was the possibility that any degraded genomic RNA that was not completely purified from the virions was reprecipitated with the PEG addition. Since the final sample has not been previously checked on agarose gels for identification of a single band corresponding to the virus there is no precedent for comparison. Therefore it cannot be determined if the supplementary RNA present is from the increase in the PEG precipitation. Although the RNA did have an effect on the virion concentration calculations the presence of this additional RNA did not affect either the electron microscopy images or purification of the CP. In order to keep the purified CP under the same conditions as the recombinant protein the column buffer was changed so that the CP was eluted in a 1 M NaCl high salt buffer. It has been observed that with decreasing salt the CP aggregates. To limit this aggregation and allow the CP to elute off the column a higher salt concentration was used for the buffer. Both the purified CP and the recombinant form were stored in the same buffer conditions, which will allow for easier comparison.

2. Isolation of Genomic RNA

Characterization of this protein compared to the recombinant will determine if the recombinant CP is in its native conformation and retains its ability to reassemble. Both can also be compared through virion reassembly experiments *in vitro*. Reassembly occurs between the CP and genomic RNA in a ratio of one molecule of the genomic RNA to 180 CPs to form a single virion. To visualize the virions on agarose gels about 1 μg RNA is needed compared to electron microscopy, which requires a virion concentration of 0.4 to 0.2 mg/mL or 0.064 to 0.032 mg/mL of genomic RNA. The high yield of purified recombinant CP was successfully expressed and purified from *E. coli*. When the genomic RNA was purified directly from dissociated virions the presence of RNA degradation resulted in a low final yield. Since only 16% of each virion consists of the genomic RNA a large initial concentration of purified virions would be required to obtain enough intact purified RNA for experimentation. Therefore the RNA was synthesized from the TCV genomic DNA template with T7 RNA polymerase, specifically using a MEGAscript kit, and the RNA was purified via a lithium chloride precipitation technique.

3. Reassembly of TCV *in vivo*

The purified synthesized genomic RNA and the recombinant CP can be used in reassembly experiments in an effort to develop an *in vitro* assembly technique. This can be used for identification of the role of satC in the inhibition of the virion assembly mechanism. These experiments may start off using the reassembly conditions for

dissociated virions (*34*). Once conditions for the functional CP and RNA assembly are established these parameters can then be implemented with the recombinant CP. Comparison between the virions assembled in both cases to actual TCV virions by electron microscopy will indicate the success of the procedure. Once an *in vitro* method has been developed the addition of satC into the reaction will determine if the satellite RNA can directly interfere with assembly. This will give some insight into the first part of the current hypothesis, which proposes that satC interferes with virion assembly, thereby decreasing the virion accumulation.

References

1. Lucas, W., and Knipe, D. M. (2002) in *Encyclopedia of Life Sciences*, Macmillan Publishers Ltd, Nature Publishing Group.
2. Johnson, J. E., and Speir, J. A. (1999) in *Encyclopedia of Virology* (Granoff, A., and Webster, R., Eds.) pp 1946-1956, Academic Press, CA.
3. Caspar, D. L., and Klug, A. (1962) *Cold Spring Harb Symp Quant Biol 27*, 1-24.
4. Hogle, J. M., Maeda, A., and Harrison, S. C. (1986) *J Mol Biol 191*, 625-638.
5. Finch, J. T., Klug, A., and Leberman, R. (1970) *J Mol Biol 50*, 215-222.
6. Callaway, A., Giesman-Cookmeyer, D., Gillock, E. T., Sit, T. L., and Lommel, S. A. (2001) *Annu Rev Phytopathol 39*, 419-460.
7. Lin, T., Cavarelli, J., and Johnson, J. E. (2003) *Virology 314*, 26-33.
8. Golden, J. S., and Harrison, S. C. (1982) *Biochemistry 21*, 3862-3866.
9. Lin, B., and Heaton, L. A. (1999) *Virology 259*, 34-42.
10. Miller, W. A., and Koev, G. (2000) *Virology 273*, 1-8.
11. Carrington, J. C., Heaton, L. A., Zuidema, D., Hillman, B. I., and Morris, T. J. (1989) *Virology 170*, 219-226.
12. Carrington, J. C., Morris, T. J., Stockley, P. G., and Harrison, S. C. (1987) *J Mol Biol 194*, 265-276.
13. Hacker, D. L., Petty, I. T., Wei, N., and Morris, T. J. (1992) *Virology 186*, 1-8.
14. Li, W. Z., Qu, F., and Morris, T. J. (1998) *Virology 244*, 405-416.
15. Kong, Q., Wang, J., and Simon, A. E. (1997) *Plant Cell 9*, 2051-2063.
16. Qu, F., Ren, T., and Morris, T. J. (2003) *J Virol 77*, 511-522.

17. Thomas, C. L., Leh, V., Lederer, C., and Maule, A. J. (2003) *Virology 306*, 33-41.

18. Wang, J., and Simon, A. E. (1999) *Virology 259*, 234-245.

19. Butler, P. J. (1999) *Philos Trans R Soc Lond B Biol Sci 354*, 537-550.

20. Zimmern, D., and Butler, P. J. (1977) *Cell 11*, 455-462.

21. Jonard, G., Richards, K. E., Guilley, H., and Hirth, L. (1977) *Cell 11*, 483-493.

22. Zimmern, D., and Wilson, T. M. (1976) *FEBS Lett 71*, 294-298.

23. Guilley, H., Jonard, G., Kukla, B., and Richards, K. E. (1979) *Nucleic Acids Res 6*, 1287-1308.

24. Sleat, D. E., Gallie, D. R., Watts, J. W., Deom, C. M., Turner, P. C., Beachy, R. N., and Wilson, T. M. (1988) *Nucleic Acids Res 16*, 3127-3140.

25. Sleat, D. E., Plaskitt, K. A., and Wilson, T. M. (1988) *Virology 165*, 609-612.

26. Turner, D. R., Joyce, L. E., and Butler, P. J. (1988) *J Mol Biol 203*, 531-547.

27. Mandelkow, E., Stubbs, G., and Warren, S. (1981) *J Mol Biol 152*, 375-386.

28. Stubbs, G., and Stauffacher, C. (1981) *J Mol Biol 152*, 387-396.

29. Butler, P. J. (1984) *J Gen Virol 65 (Pt 2)*, 253-279.

30. Johnson, J. M., Willits, D. A., Young, M. J., and Zlotnick, A. (2004) *J Mol Biol 335*, 455-464.

31. Speir, J. A., Munshi, S., Wang, G., Baker, T. S., and Johnson, J. E. (1995) *Structure 3*, 63-78.

32. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P., and Young, M. J. (2000) *Virology 277*, 450-456.

33. Willits, D., Zhao, X., Olson, N., Baker, T. S., Zlotnick, A., Johnson, J. E., Douglas, T., and Young, M. J. (2003) *Virology 306*, 280-288.

34. Sorger, P. K., Stockley, P. G., and Harrison, S. C. (1986) *J Mol Biol 191*, 639-658.

35. Qu, F., and Morris, T. J. (1997) *J Virol 71*, 1428-1435.

36. Wei, N., Heaton, L. A., Morris, T. J., and Harrison, S. C. (1990) *J Mol Biol 214*, 85-95.

37. Skuzeski, J. M., and Morris, T. J. (1995) *Virology 210*, 82-90.

38. Simon, A. E., Roossinck, M. J., and Havelda, Z. (2004) *Annu Rev Phytopathol 42*, 415-437.

39. Zhang, F., and Simon, A. E. (2003) *Virology 312*, 8-13.

40. Roth, B. M., Pruss, G. J., and Vance, V. B. (2004) *Virus Res 102*, 97-108.

41. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) *Nature 409*, 363-366.

42. Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E., and Carrington, J. C. (2004) *PLoS Biol 2*, E104.

43. Waterhouse, P. M., and Helliwell, C. A. (2003) *Nat Rev Genet 4*, 29-38.

44. Voinnet, O., Lederer, C., and Baulcombe, D. C. (2000) *Cell 103*, 157-167.

45. Llave, C., Kasschau, K. D., and Carrington, J. C. (2000) *Proc Natl Acad Sci U S A 97*, 13401-13406.

46. Mallory, A. C., Ely, L., Smith, T. H., Marathe, R., Anandalakshmi, R., Fagard, M., Vaucheret, H., Pruss, G., Bowman, L., and Vance, V. B. (2001) *Plant Cell 13*, 571-583.

47. Silhavy, D., Molnar, A., Lucioli, A., Szittya, G., Hornyik, C., Tavazza, M., and Burgyan, J. (2002) *Embo J 21*, 3070-3080.

48. Vargason, J. M., Szittya, G., Burgyan, J., and Tanaka Hall, T. M. (2003) *Cell 115*, 799-811.

49. Chapman, E. J., Prokhnevsky, A. I., Gopinath, K., Dolja, V. V., and Carrington, J. C. (2004) *Genes Dev 18*, 1179-1186.

50. Roossinck, M. J., Sleat, D., and Palukaitis, P. (1992) *Microbiol Rev 56*, 265-279.

51. Wang, J., and Simon, A. E. (2000) *J Virol 74*, 6528-6537.

52. Sun, X., and Simon, A. E. (2003) *J Virol 77*, 7880-7889.

53. Zhang, G., and Simon, A. E. (2003) *J Mol Biol 326*, 35-48.

54. Gill, S. C., and von Hippel, P. H. (1989) *Anal Biochem 182*, 319-326.

55. Coligan, J. E., Dunn, B. M., Speicher, D. W., and Wingfield, P. T. (2004) (Taylor, G. P., Ed.), John Wiley & Sons, Inc.

56. Zhang, W., Carmichael, J., Ferguson, J., Inglis, S., Ashrafian, H., and Stanley, M. (1998) *Virology 243*, 423-431.

57. Philo, J. S. (2000) *Anal Biochem 279*, 151-163.

58. Johnson, M. L., Correia, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) *Biophys J 36*, 575-588.

59. Laemmli, U. K. (1970) *Nature 227*, 680-685.

60. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J., Smith, J., and Struhl, K. (1999), John Wiley & Sons, Inc., New York.

61. Zhao, X., Fox, J. M., Olson, N. H., Baker, T. S., and Young, M. J. (1995) *Virology 207*, 486-494.

62. Hwang, D. J., Roberts, I. M., and Wilson, T. M. (1994) *Proc Natl Acad Sci U S A 91*, 9067-9071.

63. Wikipedia, t. f. e. (2002), Free Software Foundation, Inc.

64. Kane, J. F. (1995) *Curr Opin Biotechnol 6*, 494-500.

65. Kurland, C., and Gallant, J. (1996) *Curr Opin Biotechnol 7*, 489-493.

66. Kajava, A. V., Zolov, S. N., Kalinin, A. E., and Nesmeyanova, M. A. (2000) *J Bacteriol 182*, 2163-2169.

67. Andersson, H., and von Heijne, G. (1991) *Proc Natl Acad Sci U S A 88*, 9751-9754.

68. Harrison, S. C. (1980) *Biophys J 32*, 139-153.

69. Ke, J., Schmidt, T., Chase, E., Bozarth, R. F., and Smith, T. J. (2004) *Virology 321*, 349-358.

70. Heaton, L. A., Lee, T. C., Wei, N., and Morris, T. J. (1991) *Virology 183*, 143-150.