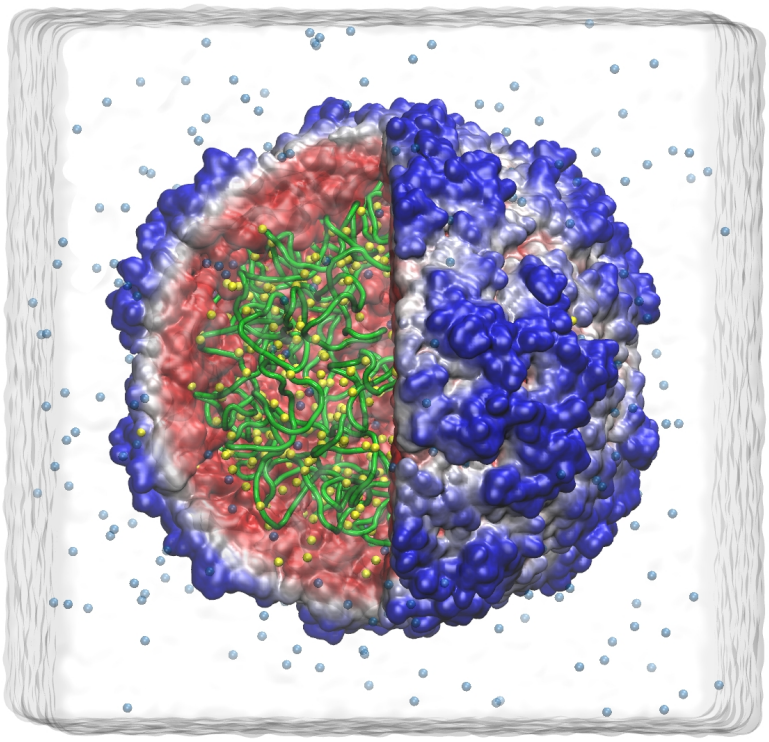
Case Study: Satellite Tobacco Mosaic Virus

Boon Chong Goh and Yaroslav Daniel Bodnar

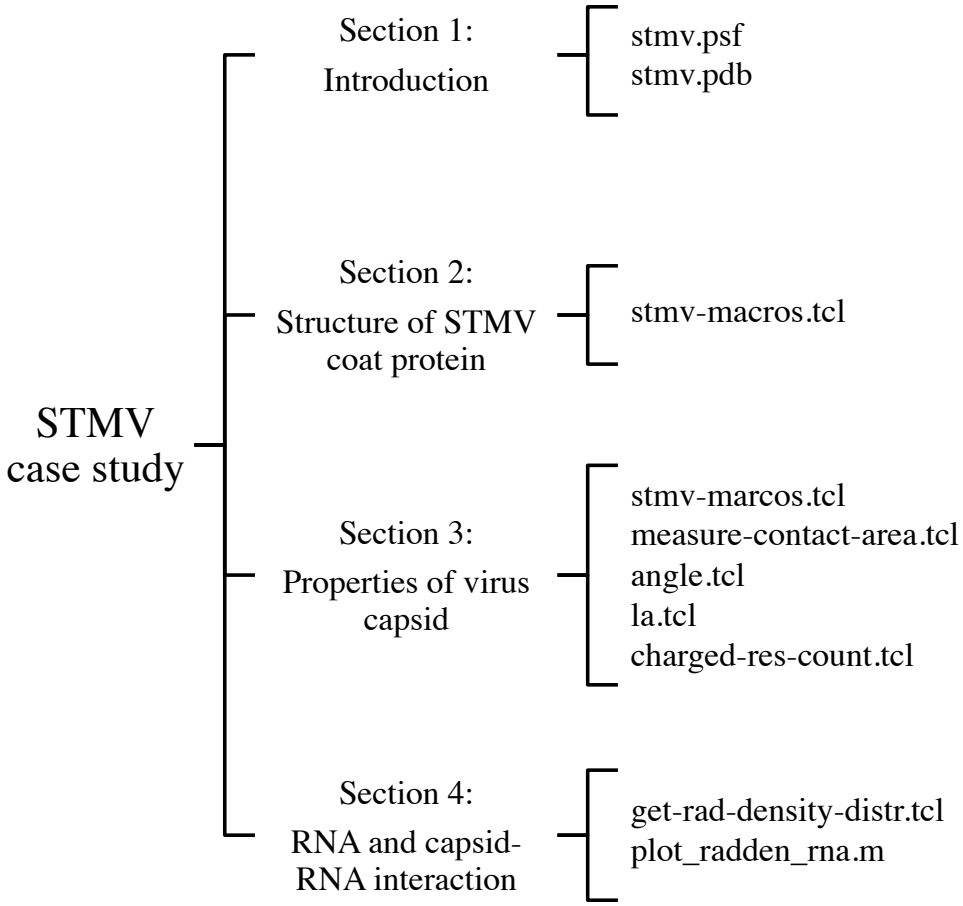
A colored version of this Case Study is available at
`http://www.ks.uiuc.edu/Training/CaseStudies/pdfs/stmv.pdf`.



Cover Figure: Viruses are the smallest life form in existence. Satellite Tobacco Mosaic Virus (STMV) is one of the simplest viruses. STMV has a protective outer coat consisting of 60 identical proteins. The coat surrounds the virus's genetic material, in this case a ribonucleic acid (RNA) molecule. STMV will be explored in this case study to illustrate the principles of virus structure.

```
                    Section 1 - Introduction to STMV
Virus case study    Section 2 - Structure of capsid protein monomers and
overview:           dimers
                    Section 3 - Propeties of the whole virus capsid
                    Section 4 - Viral RNA and capsid-RNA interactions
```

In this case study, we'll use the following files as shown below.



All files can be found in `stmv-files.tar.gz`.

1 Introduction

A virus is an infectious agent that can replicate only inside a living cell. Viruses were first discovered by Martinus Beijerinck, a Dutch microbiologist, in 1898. Beijerinck filtered the sap of diseased plants through porcelain, a microporous material known to block the passage of cellular organisms such as bacteria. Through a series of filtration experiments, Beijerinck showed that tobacco mosaic disease is caused by an infectious agent smaller than any known bacterium. Beijerinck described the infectious agent, which was too small to be seen using a light microscope, as a "contagium vivim fluidum" (contagious living fluid) [1]. It was not until development of the electron microscope in the 1930s that the morphology of virus particles was first observed [2]. The successful crystallization of tobacco mosaic virus (TMV) by Wendell Stanley in 1935, for which he earned the 1946 Nobel Prize in Chemistry, eventually lead to visualization of the virus structure at atomic resolution by X-ray crystallography[3].

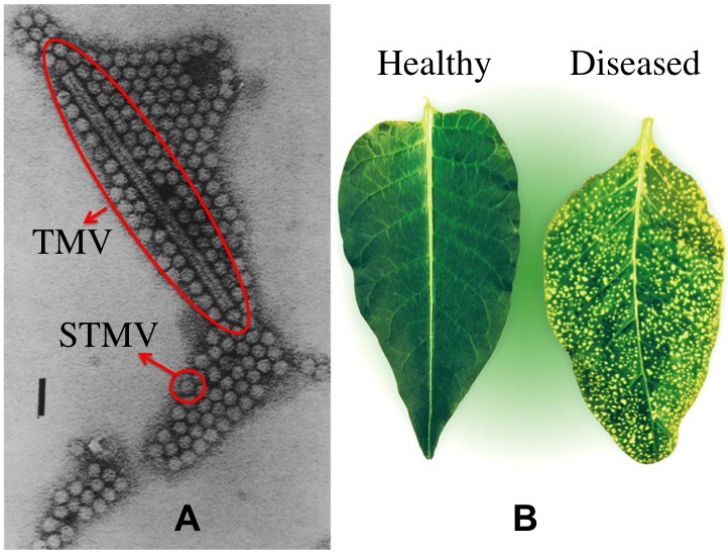


Figure 1: (A) An electron micrograph showing a long tubular Tobacco Mosaic Virus (TMV) surrounded by many small, spherical Satellite tobacco mosaic virus (STMV) particles. Bar marker represents 50nm. (B) Photograph of a healthy and TMV-infected tobacco plant.

Satellite tobacco mosaic virus (STMV) has a spherical shape and, with a diameter of approximately 17 nm, is much smaller than TMV. STMV was initially discovered on tobacco trees in southern California in association with the rod-shaped TMV. In Fig.1, panel A shows a rod-shaped TMV particle, which has a length of approximately 300 nm, surrounded by spherical STMV particles[4]. The infected leaf in panel B forms a mottled pattern of light and dark green areas in the leaf region, better known as the "mosaic" pattern [5]. STMV is a satellite virus since it is entirely reliant on co-infection of a cell with TMV to replicate. Like all non-enveloped viruses, which are not surrounded by a lipid membrane, STMV is composed of a closed protein shell, or the viral capsid, and genetic material (single-stranded RNA in the case of STMV).

Due to its small size, STMV could be studied early on computationally by molecular dynamics simulation [6]. In this case study, we will use STMV as a model system to explore the properties of small RNA viruses, including the structure and key interactions of capsid proteins, the virus

capsid, the RNA genome, and capsid-RNA complexes.

Exercise 1: General structural features of STMV

In this exercise you will use VMD to explore the structural properties of STMV described in this section. Load the provided coordinate files `stmv.psf` and `stmv.pdb` into VMD.

A. Create a representation for the protein component of STMV. You can do so by opening the Graphical Representations window via `Graphics` → `Representations` in the VMD Main window, and creating a new representation with `protein` as the Selected Atoms. Try using `Quick Surf` for the Drawing Method and `Segname` for the Coloring Method. You should be able to see the 60 proteins that make up the virus capsid.

Create a new representation with `nucleic` as the Selected Atoms. Use `Tube` for the Drawing Method, `ColorID` for the Coloring method and color the RNA in green. Hide the protein to view the RNA that is encapsulated by the protein capsid.

B. Reproduce the cover figure as best as you can, and save an image of your view via `File` → `Render...` on the VMD main menu.
*Hint: To color the surface of the virus capsid by the radial distance from the center, use Radial via Coloring Method → `Position` in the VMD `Graphical Representations` window.

C. The complete nucleotide sequence of STMV RNA deduced by experiments is approximately 1059 nucleotides in length [7]. Estimate the length of RNA needed to encode all of the capsid proteins. Based on your estimate, argue that the STMV capsid must be composed of identical proteins.
Hint: Three nucleotides encode an amino acid. Note that not all nucleotides are necessarily translated into amino acids. Please refer to [7] for more information.

2 Structure of STMV Coat Protein Monomer and Dimer

2.1 The coat protein is the building block of the STMV capsid

As is demonstrated in Exercise 1.D, genetic efficiency is a requirement of viruses, since the viral genome must be able to encode all of the coat proteins that comprise the capsid. This observation, along with the self-assembly properties of viruses, led Watson and Crick to hypothesize in 1956 that viruses must be highly symmetric assemblies of a structural subunit [8], which has since been confirmed by experimental studies. In the case of STMV, for which one of the highest resolution crystallographic structures of a virus has been solved, the capsid is composed of 60 identical copies of the coat protein (CP), which is shown in Fig. 2. As with most other virus capsid proteins, the STMV coat protein has a distinctive structure known as the jellyroll fold, which has a unique arrangement of eight β-strands that form two anti-parallel sheets, known as the BIDG and CHEF sheets [9]. The CPs are arranged in STMV such that the CHEF sheet is on the outer surface of the capsid and the BIDG sheet is on the inner surface. The eight β-strands that make up the structural core of the CP are highlighted in Fig. 2. The structure of the jellyroll fold has evolved to enable virus capsid proteins to efficiently self-assemble into closed shells.

Although most virus capsid proteins share the same structural core, the unique properties of each virus capsid are mainly due to the variable polypeptide tails and loops that connect the core β sheets. A prominent and functionally crucial feature of the STMV coat protein is the extended N-terminal tail, which is colored purple in Fig. 2. In STMV, the N-terminus securely cross-links two neighboring coat proteins.

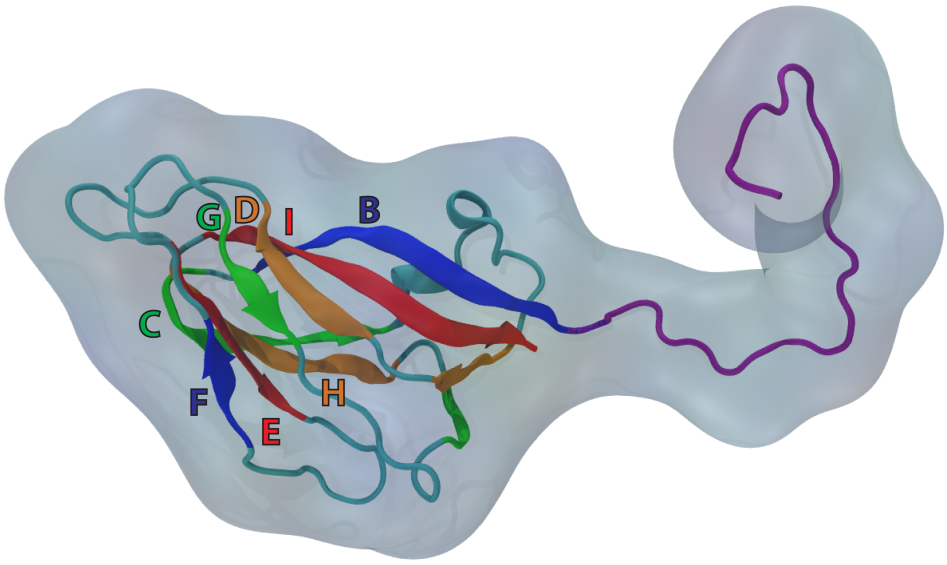


Figure 2: Structure of the STMV coat protein. The eight β-strands of the jellyroll fold, which are colored blue (B, F), red (I, E), orange (D, H), and green (G, C), form two antiparallel sheets, known as the BIDG and CHEF sheets. The variable loops that connect the core β-strands and give the STMV capsid its unique structure and properties, are shown in cyan. The extended N-terminal tail, which securely bridges two neighboring coat proteins, is shown in purple.

2.2 Coat protein dimers are important capsid assembly intermediates

Two STMV coat proteins can associate to form a stable, saddle-shaped dimer, as shown in Fig. 3. The dimeric interface contains extensive favorable hydrophobic contacts, direct intermolecular hydrogen bonds, as well as hydrogen bonding interactions that are mediated by water molecules [10]. In addition, the N-terminus forms an extended β-sheet with the core β-strand of the neighboring CP. STMV coat protein dimers are believed to be an important intermediate in assembly of STMV and contribute to the structural stability of the fully assembled virus.

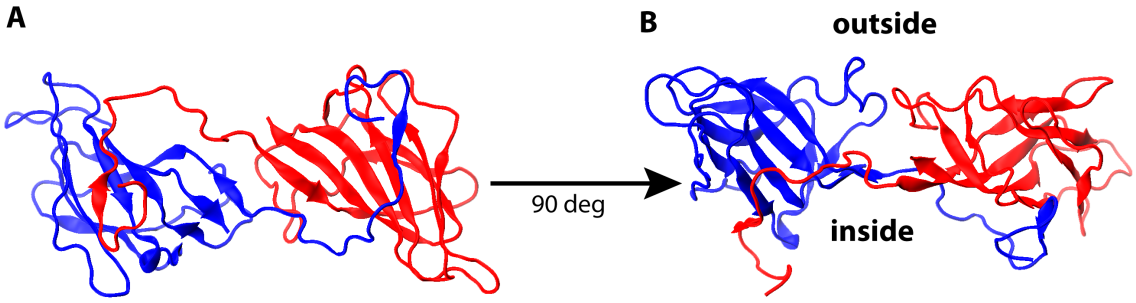


Figure 3: Structure of the STMV coat protein (CP) dimer. Individual CP monomers are colored red and blue. (A) View of the inner surface of the dimer, showing the extended β-sheet formed between the N-terminus and CHEF β-sheet of the neighboring monomer. (B) Side view of the dimer with the outer surface of the capsid on the top and inner surface on the bottom.

Exercise 2: Exploring the structure of the STMV capsid protein monomer and dimer

A. Topology and architecture of the STMV capsid protein. Take a close look at an individual coat protein using the `NewCartoon` drawing method. A CP monomer can be selected by entering ASU1 into Selected Atoms (hint: The atom selection macro file, stmv-macros.tcl, enables any of the 60 subunits to be selected using ASU[1-60]; to use the macros, run the command `source stmv-macros.tcl` in the TkConsole). Identify the BIDG and CHEF β-sheets and use the β-strands cartoon in Fig. 4 to trace the topology of the jelly-roll fold (hint: Individual strands of the CP can be selected using the ASU[1-60][A-J] macros. For example, the alias for the C-strand of the first CP monomer is ASU1C and the ASU1 N-terminus is ASU1A. Aliases are also provided for each of the variable loops, such as ASU1CD for the loop connecting the ASU1 C-and D-strands, and the core β-sheets, such as ASU1CHEF). Label the individual β-strands and draw the connecting loops. Also, label the N-and C-termini.

Visualize a CP monomer without the N-terminal tail. Describe the overall structure of the CP and explain how the shape of the CP enables it to perform its function of associating with 60 equivalent subunits to form an icosahedral shell.

B. Interactions at the dimer interface and non-covalent cross-linking of monomers. Show the individual CP monomers that form a dimer in the STMV capsid. Use the `QuickSurf` drawing method for both monomers and color the surface by `ResType`. Describe the properties of the dimeric interface and use a more detailed molecular representation to identify some of the key residues at the dimeric interface. In addition, observe the non-covalent cross-linking of the monomers due to extended β-sheet formation by the N-terminus. Identify to which core β-strand of the neighboring monomer the N-terminus binds.

Figure 4: Cartoon representation of the β-strands in the jellyroll fold of viral capsid proteins.

3 Properties of the Virus Capsid

3.1 The Whole STMV capsid

In the previous section, we looked at structural features of CP monomers and dimers, the basic building blocks of STMV. In this section, we will discuss the fundamental properties of the whole STMV capsid.

The whole STMV capsid might look complicated at first glance due to its size. To aid visualization, STMV is represented as a regular dodecahedron that is composed of 12 regular pentagonal faces. Figure 5 puts a STMV capsid and a dodecahedron side-by-side with their corresponding pentamers shown in the same color.

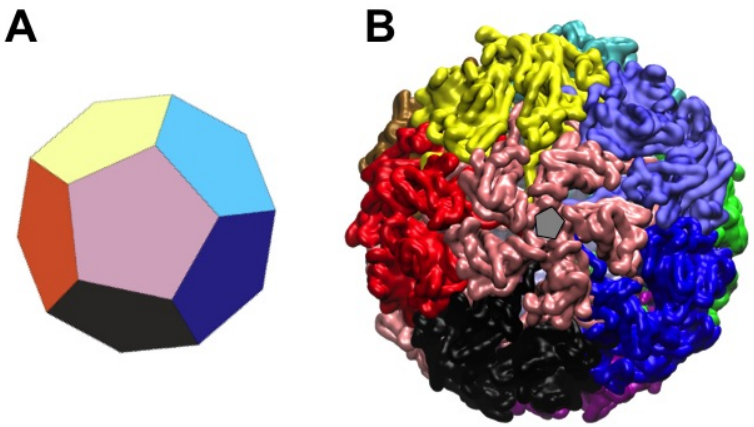


Figure 5: Dodecahedron (A) (image courtesy of commons.wikipedia.org) can be used to represent the protein capsid of STMV (B). The central pentamer of STMV capsid in panel B, which is represented in thick pink tube representation, looks the same upon every 72 degree rotation along the rotation axis, labeled as a grey pentagon. The other pentamers are in thick tube representation of different colors.

An object with rotational symmetry looks the same after a specific amount of rotation about a symmetry axis. For example, figure 5B exhibits a 5-fold rotational symmetry because the capsid looks the same after a $360/\mathbf{5} = 72$ degree rotation about the 5-fold axis. STMV, like many other viruses, overall exhibits icosahedral symmetry. Icosahedral symmetry is the highest order of symmetry for any object. The dodecahedral capsid of STMV which displays icosahedral symmetry has six 5-fold rotational symmetries (pentameric faces), ten 3-fold (trimeric vertices) and fifteen 2-fold (dimeric edges). When resolving the structure of a virus capsid, X-ray crystallographers make use of the icosahedral symmetry of the virus to refine the structures. Additionally, the rotational axes are useful points of reference when one first looks at an icosahedral virus capsid.

Next, let us look at the role of electrostatics in maintaining structural stability of the STMV capsid. The total charge of the capsid is +300e. With this amount of charge packed into the small capsid, the repulsive electrostatic force is large. The capsid would not assembly into a stable structure if it were not for the RNA inside the capsid. We will unravel the details behind the stable

capsid of STMV in the following exercise.

Exercise 3: Icosahedral Symmetry and electrostatics of the STMV capsid

In problems A, B and C, we will explore the symmetry of STMV; in problem D, we will investigate the electrostatic properties of the capsid.

A. In a new VMD session, open the visualization state file `symmetrical-axes.vmd` to load the coordinates of the proteins. An alias file `stmv-macros.tcl` contains the macros for selecting the pentameric protein complexes in STMV. Instead of typing `segname C0 C1 C2 C3 C4` in `Selected Atoms` to display one of the pentamers, you could type `penta1` in the `Selections` tab of the `Graphical Representations` window. The pentamers are named `penta1`, `penta2`, ..., `penta12`. Can you find the 3-fold and 2-fold rotational axes of STMV? Save the respective views of STMV displaying 3-fold and 2-fold rotational symmetry via `File` → `Render...` on the VMD main menu. After you have printed out the images, draw a triangle and an ellipse to highlight the positions of the 3-fold and 2-fold rotational axes.

B. We can gain useful insight into the assembly process of the capsid by measuring the contact surface area between the CPs at the 5-fold, 3-fold, and 2-fold rotational axes. Using the first CP (`segname C0`) as the reference, find the corresponding CPs that are involved in 5-fold, 3-fold, and 2-fold rotational symmetries relative to `segname C0`. Next, measure the contact surface area between the involved CPs for the respective rotational symmetries. For example, the pentamer that exhibits 5-fold rotational symmetry in figure 5B is composed of 5 CPs, namely `segname C0 C1 C2 C3 C4`. The contact surface area among the 5 CPs could be calculated as follows:

1. Run `source measure-contact-area.tcl` and `measure_contact_area 0 {segname C0} {segname C1} contact_c0-c1.txt` in the VMD TkConsole. The output file contains the contact surface area between `segname C0` and `segname C1`.

2. Repeat the calculation for `segname C2`, `C3`, and `C4` respectively to obtain the total contact surface area for the pentamer.

3. Divide that total contact area value by 5 to get the contact surface area contributed by one CP in the 5-fold symmetrical plane.

Repeat step 1 to 3 for the CPs that display the 3-fold and 2-fold symmetrical axes. Now with the total contact area per CP for different symmetries, propose a model of the assembly process of the STMV capsid.

C. Next we will measure the angles between two nearby pentamers. With the same VMD session, open the VMD TkConsole and type `source angle.tcl`. Make sure the files `angle.tcl`, `la.tcl`, and `orient.tcl` are in the working directory. The output file `angle_stmv.txt` contains the distances between the centers of mass of the nearby pentamers and their angles. Plot a histogram with bin size of 1 degree to show the angle distribution. Did you expect the angle distribution of the STMV capsid to look like the histogram?

D. To have a better understanding of the electrostatics of the capsid, we will calculate the radial distribution of the charged residues. In the same VMD session, type `source charged-res-count.tcl` in VMD TkConsole to calculate the number of positively (basic) and negatively-charged (acidic) residues at different radial distances from the nucleic acid. Two text files will be generated: `basic-residues.txt` and `acidic-residues.txt`. Plot the radial distribution for basic and acidic residues. Comment on the distribution and suggest what role the RNA inside the capsid might play in the stability of capsid.

3.2 More complex viruses

While STMV can be represented as a regular dodecahedron, many viruses are larger and more complex than STMV. Larger capsids are typically constructed of both hexameric and pentameric capsid subunits. One of Euler's theorems states that exactly 12 pentagons are needed to form a closed surface from a 2-dimensional hexagonal network. Thus, to build larger capsids, more hexamers are incorporated into the capsid structure, while the number of pentamers remains exactly 12. For example, Dengue virus and Cowpea Chlorotic Mottle Virus take on a truncated icosahedral structure, with 20 hexagons and 12 pentagons (the same shape, incidentally, as Buckminsterfullerene). However, not all virus capsids have a symmetric shape. An example of an irregular virus capsid is the human immunodeficiency virus (HIV) capsid. Figure 6A shows the fullerene cone model of an HIV capsid that has 216 hexagons and 12 pentamers [11].

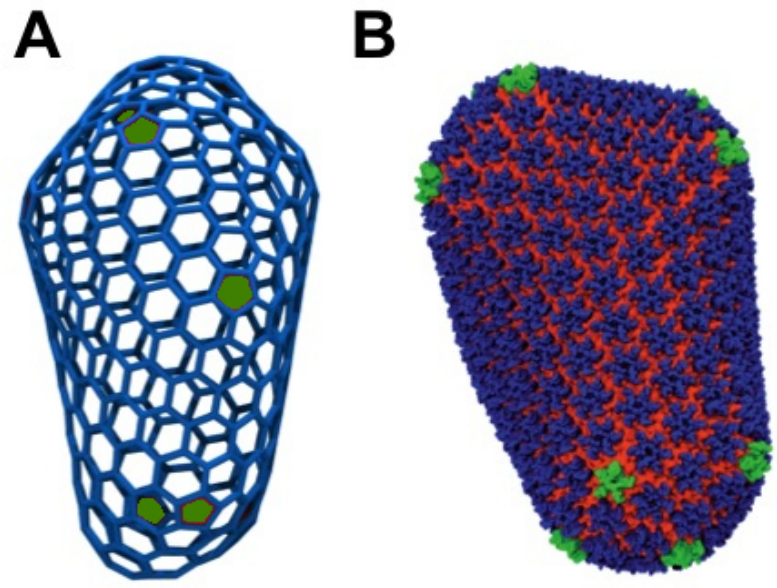


Figure 6: The human immunodeficiency virus (HIV) has an irregular capsid. (A) A fullerene cone model of an HIV capsid with the pentamers colored in green (image courtesy of Juan R. Perilla). (B) An all-atom structure of HIV capsid comprising 216 hexamers (amino-terminal domain in blue; carboxy-terminal domain in red) and 12 pentamers (green).

4 Structure of encapsidated viral RNA and capsid-RNA interactions

Due to limitations in experimental methods, the atomic structures of most encapsidated viral genomes are not known. STMV is one of the few viruses for which a high resolution atomic structure of a significant portion of the genome, approximately 59%, has been solved [10]. The STMV genome, which is shown in Fig. 7, consists of a single-stranded RNA molecule that folds inside the capsid to form 30 icosahedrally arranged double helical stemloops. The 30 helical segments are assigned in the X-ray atomic model of STMV. However, due to their flexibility, the linkers that connect the stemloops were not resolved. The structures of the linkers joining the stemloops were modeled computationally in order to construct a complete model of the STMV encapsidated genome [6].

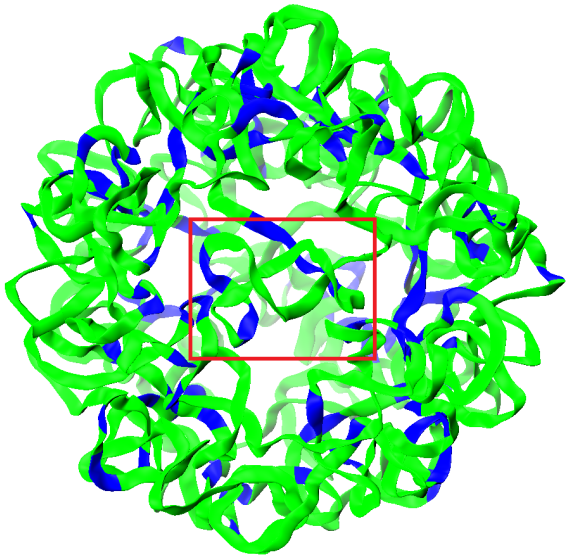


Figure 7: Structure of the STMV RNA genome. Segments of the RNA that correspond to double helical stemloops, of which there are 30, are shown in green and linkers between the helical segments are shown in blue. The red box highlights a single stemloop in the RNA structure.

Since the structure of STMV has been solved at 1.8 Å resolution, it permits a detailed look at interactions between a virus capsid and the genetic material that it contains. An important feature of STMV is that each of the helical stemloops is intimately associated with an RNA binding site on the inner surface of a coat protein dimer and is oriented perpendicular to a 2-fold symmetry axis of the capsid. As can be seen in Fig. 8, the saddle-shape of the inner surface of the CP dimer enables it to wrap around the double helical RNA segment.

A detailed examination of the interactions at the CP dimer-RNA stemloop interface shows that the coat proteins primarily interact with the RNA via salt bridges and hydrogen bonds to the sugar-phosphate backbone of the RNA. Since the capsid-RNA interactions mainly involve the RNA backbone, association of the viral genome with coat proteins is independent of its sequence of nucleotides. Due to the high variation in length and nucleotide sequence of the genome between different STMV virus particles, the ability of the coat proteins to encapsidate a variety of RNA molecules is crucial to the successful replication of STMV [10].

The intimate association of STMV coat proteins with the RNA genome suggests that the coat protein-RNA interactions play an essential role during assembly of the virus [10]. An assembly

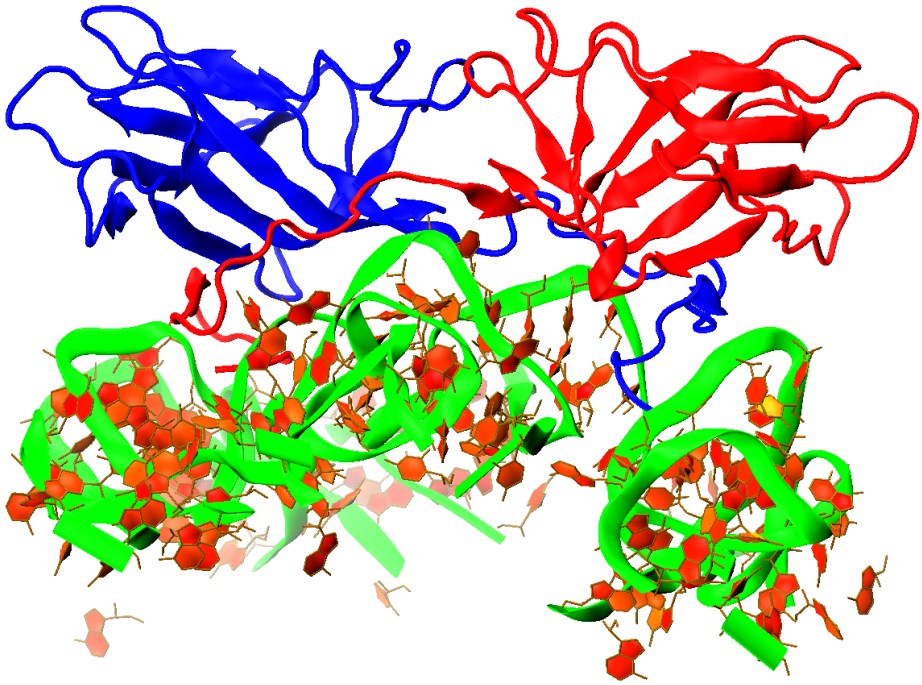


Figure 8: A STMV coat protein dimer (individual CP monomers are colored red and blue) interacting with an RNA double helical stemloop (backbone shown in green and nucleotide bases in orange) of the encapsidated viral genome. The orientation of the helical axis of the RNA stemloop is perpendicular to the 2-fold symmetry axis of the capsid.

mechanism for STMV has been proposed in which CP dimers bind to double helical segments of the viral genome and, in a concerted process, condense the RNA by forming higher orders of capsid symmetry elements, specifically associating into the trimeric and pentameric CP structures at the 3-fold and 5-fold symmetry axes, respectively. Hence, association of CP dimers bound to helical segments of viral RNA may drive STMV assembly. In addition to playing a critical role during STMV assembly, the viral RNA may also be essential for the mechanical stability of the STMV particle. Molecular dynamics simulations of empty STMV capsids with the RNA removed have shown that the capsid collapses, as shown in Fig. 9 [6]. The essential role of the RNA in STMV assembly and, possibly, stability is also supported by the fact that empty STMV capsids have not been observed by electron microcopy studies.

Exercise 4: Secondary structure of the STMV genome, ionization of the viral RNA, and capsid-RNA interactions.

A. Secondary structure of the STMV genome.

Load the provided structure and coordinate files, stmv.psf and stmv.pdb, into VMD. Calculate the maximum number of hydrogen bonds that can be formed between the bases of the viral RNA (Hint: The hydrogen-bond donor atoms are (`"(resname ADE and name N6) or (resname URA and name N3)"`) and the acceptor atoms are (`"(resname ADE and name N1) or (resname URA and name O4)"`) in the RNA bases). Now, measure the actual number of hydrogen bonds between the bases of the RNA (Hint: Use `"resname ADE and not backbone"` for selection 1 and `"resname URA and not backbone"` for selection 2) using the Hydrogen Bonds analysis extension (`Extensions` → Analysis → `Hydrogen Bonds`). Check the `write output to files?` option and run `Find hydrogen bonds!` using a Donor-Acceptor distance of 3.5Å and Angle cutoff of 30 degrees. The output file `hbonds.dat` will be produced in your current working directory. Open `hbonds.dat` using a text editor, which will contain the number of measured hydrogen bonds for frame 0 of the selected molecule. What fraction of the possible base pairing hydrogen bonds are formed in the encapsidated genome and what does the extent of base pairing indicate about the secondary structure of the viral RNA? Select the RNA bases inside the capsid and visualize the base pairing hydrogen bonds using the `HBonds` drawing method with a `Distance Cutoff` of 3.5Å and Angle Cutoff of 30 degrees. You will need to increase the `Line Thickness` to easily see the hydrogen bonds. Comment on the hydrogen bonding pattern and secondary structure of the encapsidated RNA.

B. Ionization of the STMV genome.
Measure the total charge of the viral RNA. In the VMD TkConsole, use `atomselect` to select the RNA and run the command `measure sumweights $your-RNA-selection weight charge` to obtain the total charge. As you should expect, the RNA is highly negatively charged. How does the viral RNA remain stable despite having such a high negative charge (Hint: Take a close look at other components of the STMV system surrounding the RNA)? Next, measure the radial density distribution of the protein, RNA, and ions in the STMV particle by running the `get-rad-density-distr.tcl` script, which produces the output files `radden_prot.dat`, `radden_rna.dat`, `radden_mg.dat` and `radden_cl.dat` in the current working directory. Plot the radial density distribution of protein, RNA, magnesium ions, and chloride ions (Hint: a Matlab script `plot_radden.m` is provided. If you do not use Matlab, you can plot the graph with your preferred graphing software). Do you notice anything in the radial density distribution profiles that supports your hypothesis on the stability of the viral RNA?

C. Capsid-RNA interactions.
Look at the interactions between the RNA and the inner surface of the capsid, paying specific attention to interactions at the two-fold, three-fold, and five-fold symmetry axes. At which symmetry axis of the capsid does the RNA make the most contact with the capsid? Take a close look at the RNA and capsid proteins at the main points of contact and describe the structure of the RNA at these contact points. Identify the type of interactions between the RNA and the capsid proteins. What regions of the RNA do the capsid proteins interact with and what implications might this have on assembly of the STMV particle?

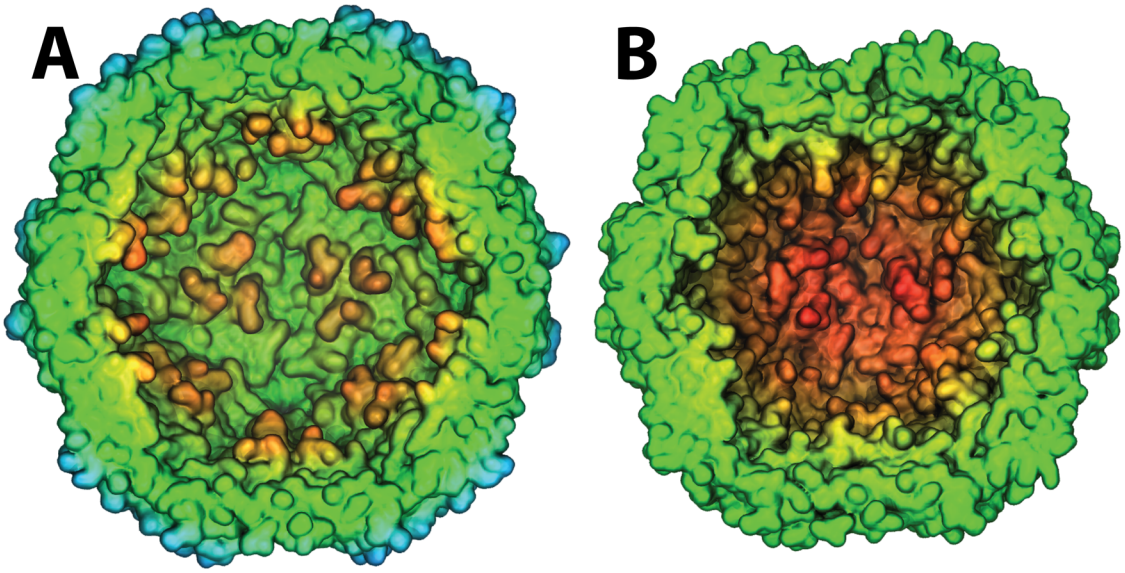


Figure 9: (A) Cut through the center of the starting point of a molecular dynamics simulation of the STMV capsid with RNA removed. (B) Endpoint of molecular dynamics simulation of STMV capsid with RNA removed, showing that the capsid collapses without the necessary internal support from the RNA [6]. The surfaces are colored by distance from the center of the capsid.

References

[1] M.W. Beijerinck. Concerning a contagium vivum fluidium as a cause of the spot-disease of tobacco leaves. *Reprint from: Phytopathology Classics, Number 7. 1942. James Johnson, translator, APS Press.*, 1898.

[2] Phillippe Roingeard. Viral detection by electron microscopy: past, present, and future. *Biol. Cell*, 100:491–501, 2008.

[3] Milton Zaitlin. The Discovery of the Causal Agent of the Tobacco Mosaic Virus Disease. In *Discoveries in Plant Biology*, pages 105–110. Elsevier, 1998.

[4] RA Valverde and JA Dodds. Some properties of isometric virus particles which contain the satellite RNA of tobacco mosaic virus. *Journal of general virology*, pages 965–972, 1987.

[5] E Lam, N Kato, and M Lawton. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature*, 411(6839):848–53, June 2001.

[6] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.

[7] T.E. Mirkov, D.M. Mathews, D.H. Du Plessis, and J.A. Dodds. Nucleotide sequence and translation of satellite tobacco mosaic virus {RNA}. *Virology*, 170(1):139 – 146, 1989.

[8] F.-H.-C. Crick and J.-D. Watson. Structure of small viruses. *Nature*, pages 473–476, 1956.

[9] Shanshan Cheng and Charles L.-Brooks. Viral capsid proteins are segregated in structural fold space. *PLoS Comput. Biol.*, 9(2):e1002905, 2013.

[10] S. B. Larson, J. Day, A. Greenwood, and A. McPherson. Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.*, 277:37–59, 1998.

[11] Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy, Xin Meng, Bo Chen, Jiying Ning, Jinwoo Ahn, Angela M. Gronenborn, Klaus Schulten, Christopher Aiken, and Peijun Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646, 2013. doi:10.1038/nature12162.