Title
The Self-assembly of Cowpea Chlorotic Mottle Virus

Author
Garmann, Rees Frederik

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The self-assembly of Cowpea Chlorotic Mottle Virus

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Chemistry

by

Rees Frederik Garmann

2014

ABSTRACT OF THE DISSERTATION

The Self-assembly of Cowpea Chlorotic Mottle Virus

by

Rees Frederik Garmann

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2014

Professor William M. Gelbart, Chair

The first half of this dissertation (Chapters 1-3) deals with the *in vitro* self-assembly of Cowpea Chlorotic Mottle Virus (CCMV) from its single-stranded (ss) RNA genome and its capsid protein (CP). We have systematically investigated how the assembly depends on three previously unaddressed factors: (1) the length of the RNA; (2) the strength of attraction between the CP and the RNA; and (3) the electrostatic charge of the CP. Our results support a theoretically predicted but previously unverified self-assembly mechanism in which CP initially binds RNA in a disordered manner, and then reorganizes into the final icosahedral capsid structure. Additionally, we find that any length of RNA (from 140 to 12,000 nucleotides) can be efficiently packaged so long as two key requirements are met. First, the reaction mixture must contain a specific excess of CP to provide "charge-matching" between the relevant basic residues

of the CP and the phosphate backbone of the RNA. Second, the assembly reaction must be carried out by a two-step, pH- and salt-dependent protocol.

In the second part (Chapters 4 and 5) we explore the potential use of self-assembled CCMV virus-like particles (VLPs), both as fluorescent nano-particles with controllable size (Chapter 4), and as gene delivery vectors (Chapter 5). We find that the fluorescent properties of dye-labeled polymers encapsidated by CCMV CP are robust against quenching by external quencher molecules. Additionally, we find that VLPs consisting of RNA genes packaged in CCMV capsids, once delivered to cytoplasm of mammalian cells, are released from their capsids and efficiently expressed, allowing them to be exploited for gene delivery.

In Chapter 6, we have examined the complicated branching networks of ss-RNA secondary structures by direct visualization with cryo-electron microscopy and confirmed a previously reported secondary structure of the entire genome of satellite tobacco mosaic virus. Finally, we discuss current preliminary data and future work in the Summary section.

Throughout this work we have repeatedly exploited the techniques of native agarose gel electrophoresis, velocity sedimentation, fluorescence spectroscopy (and microscopy), and electron microscopy.

The dissertation of Rees Frederik Garmann is approved.

Charles M. Knobler

Margot E. Quinlan

Giovanni Zocchi

William M. Gelbart, Committee Chair

University of California, Los Angeles

2014

Dedications

To my friends.

Table of Contents

Acknowledgements

For technical reasons enforced by the UCLA Graduate Division, my co-advisor, mentor, and friend, Charles M. Knobler, cannot be listed as my committee co-chair. It turns out he was never formally appointed as a co-chair of my committee, despite his complete functional role as such. While I was forced to remove his name from page ii, and his title of co-chair from page iv, I ask the reader to mentally re-insert this missing information. Similarly, Thomas G. Mason should be acknowledged as a member of my committee.

Chapter 1 is a version of Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M. & Gelbart, W. M. (2012). Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J. Virol. 86, 3318-3326.

Chapter 2 is a version of Garmann, R. F., Comas-Garcia, M., Gopal, A., Knobler, C. M., & Gelbart, W. M. (2014). The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. J. Mol. Biol. 426 1050–1060.

Chapter 3 is a version of Garmann, R. F., Comas-Garcia, M., Koay, M. S. T., Cornelissen, J. J. L. M., Knobler, C. M., & Gelbart, W. M. (2014). The Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus. Submitted.

Chapter 4 is a version of Cadena-Nava, R. D., Hu, Y., Garmann, R. F., Ng, B., Zelikin, A. N., Knobler, C. M., & Gelbart, W. M. (2011). Exploiting Fluorescent Polymers to Probe the Self-Assembly of Virus-Like Particles. J. Phys. Chem. B 115: 2386-2391.

Chapter 5 is a version of Azizgolshani, O., Garmann, R. F.; Cadena-Nava, R. D., Knobler, C. M., & Gelbart, W. M. (2013). Reconstituted Plant Viral Capsids Can Release Genes to Mammalian Cells. Virology 441, 12–17

Chapter 6 is a version of Garmann, R. F., Gopal, A., Athavale, S. S., Knobler, C. M., Gelbart, W. M., & Harvey, S. C. (2014) Visualizing the Global Secondary Structure of a Viral Genome with Cryo-Electron Microscopy. Submitted.

Vita

Harbor High School, Santa Cruz, CA, 2003

Undergraduate Researcher with Alexander Pines, UC Berkeley, 2006-2007

William C. Bray Award, 2007

B.S. Chemistry, University of California at Berkeley, 2007

UC Chancellor's Prize 2008-2009

UCLA Department of Chemistry and Biochemistry Excellence in Teaching Award, 2009

UCLA Department of Chemistry and Biochemistry Teaching Assistant Consultant, 2010-2012

George Gregory Research Fellowship 2013

UCLA Dissertation Year Fellowship, 2013-2014

UCLA Distinguished Teaching Award, 2014

UCLA Department of Chemistry and Biochemistry Physical Chemistry Dissertation Prize, 2014

Introduction

If you are ever at a cocktail party and word gets out that you are a virologist, inevitably you will be asked, 'So, are viruses alive?', followed, usually, by a lively debate. This same debate – adapted with a few extra details – extends into the academic setting as well, where the question has remained essentially unresolved. On one hand, like all other living organisms, viruses encode their genetic information using either DNA or RNA and efficiently adapt to environmental pressures through the process of evolution by natural selection. On the other hand, they are inanimate objects, incapable of respiration or locomotion, and often consist of only a few hundred molecules (*i.e.* a few million atoms). Furthermore, some simple viruses can be manufactured from scratch (*i.e.* from non-infectious starting materials) in the laboratory. The precise details of how viruses carry out their complex biological function with such minimal machinery remains a highly active area of fundamental research. So while the question of whether or not a virus is 'alive' is clearly of deep philosophical importance to our understanding of what properties give rise to life, it leads us to an even more profound and interesting question, 'how do viruses work?'.

The structure of many viruses is beautifully simple. For single-stranded RNA-genome plant viruses, the virion often consists of only two structural elements: the RNA carrying the genetic information, and the capsid protein (CP), many copies of which form an icosahedrally-symmetric protective shell (the capsid) around the genetic cargo. During its 'life cycle', the virus particle must be both stable enough to protect its genome from degradation along the treacherous journey between hosts, as well as labile enough to disassemble inside the host cell and release its genetic contents. Additionally (and remarkably), many viruses are also capable of *in vitro* self-

assembly – during which CP and RNA reorganize spontaneously in solution to form a monodisperse population of viruses indistinguishable from those found *in vivo*. Moreover, this *in vitro* assembly can be carried out with non-viral RNAs to form virus-like particles (VLPs) with the same highly symmetric structure as the native virus. These astonishing properties are made possible by a delicate balance of attractive and repulsive forces that has evolved among the viral structural units. The present work (Chapters 1-3) focuses on the interplay between these key interactions and their effect on the *in vitro* assembly pathway of the plant RNA virus, Cowpea Chlorotic Mottle Virus (CCMV).

CCMV (shown in Figure 1) has a multipartite genome encoding four genes contained in three positive-sense single-stranded (ss)-RNA molecules of 3171 nucleotides (nt) – RNA1, 2774 nt – RNA2, and 2173 nt – RNA3. The first two genomic RNAs are packaged into separate capsids and the third genomic RNA is co-packaged with a subgenomic RNA4 of 824 nt, so that each CCMV virion contains about 3000 nt. The CCMV capsid has a 28-nm diameter and is made up of 180 chemically identical copies of a 190-residue CP. The CP molecules are arranged with T=3 Caspar-Klug symmetry (9), where T represents the number of asymmetric positions that the CP adopt in the capsid.

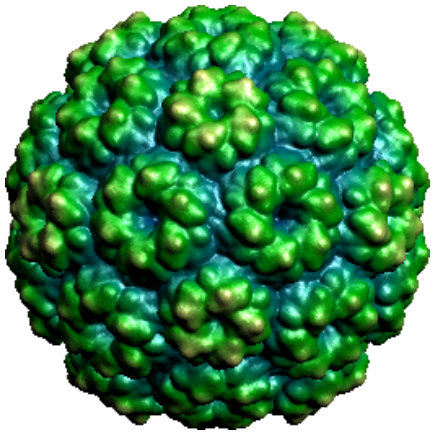

Figure 1. Shown is an exterior view of a 3.2-Å resolution cryo-EM reconstruction of CCMV taken from the VIPERdb2 database (12). The capsid is comprised of 180 chemically equivalent CP molecules that are arranged as pentamers (there are 12 per capsid) or hexamers (there are 20). The capsid diameter is 28nm.

We chose to study CCMV assembly for three reasons. First, CCMV is among the most robust viral self-assembly systems – it was the first icosahedral virus to be reconstituted *in vitro* from purified components (1), and CCMV CP has since been shown to package a wide range of heterologous RNAs (2), in addition to being capable of assembling into empty capsids (3, 4). Second, high-resolution reconstructions of the virion obtained by X-ray crystallography and cryo-electron microscopy (cryo-EM) (5), in conjunction with computational simulations (6), have yielded great insight into the specific molecular interactions that govern the structure, stability, and dynamic properties of the virion. Third, the strength of the CP-CP and CP-RNA interactions can be controlled by adjusting the pH and ionic strength, respectively (7, 8).

While much is known about the fine structure and intermolecular forces within the CCMV virion, the pathway of assembly has remained poorly understood. We investigate the co-assembly of an excess of CCMV CP and ssRNAs of varying lengths (Chapter 1), and find that

any length of RNA (from 140 to 12,000 nucleotides (nt)) can be efficiently packaged so long as the CP:RNA ratio is sufficiently high. In the case of longer RNA molecules (>4,500 nt), multiplet capsids were discovered that consist of a single RNA that is shared by two or more capsids, with the RNA being threaded either through small nm-sized defects or the pores that are naturally present in the capsid. Additionally, we report (Chapter 2) evidence that the efficient assembly of CCMV CP around ssRNA follows a two-step, pH-dependent, mechanism in which CP reversibly binds RNA at neutral pH to form a pre-capsid complex that upon acidification undergoes an irreversible reorganization into the final icosahedral capsid. Lastly, we show (Chapter 3) that the initial pre-capsid is "charge-matched", consisting of RNA decorated by a disordered arrangement of enough CPs to neutralize the negative charges brought by the phosphate backbone of the RNA. This pre-capsid reorganizes into the final, icosahedrally-symmetric, nucleocapsid by displacing the excess CPs from the RNA to the exterior surface of the emerging capsid, through electrostatic attraction between basic residues of the excess CP and the negative charge density of the capsid exterior.

In Chapters 4 and 5, we explore the potential application of self-assembled CCMV VLPs to the fields of materials science and gene delivery. The packaging of 38 kDa poly(styrene sulfonate) (PSS) that has been fluorescently labeled with Rhodamine B gives rise to a bimodal distribution of VLP diameters corresponding to T=1 (18 nm) and T=2 (22 nm) Caspar-Klug structures (Chapter 4); increasing the molar ratio of protein to PSS in the reaction mix shifts the VLP distribution from T=1 to T=2 structures. Additionally, we find that the quenching of Rhodamine B fluorescence by the quencher molecule, methyl viologen, is significantly reduced upon packaging due to the ionic environment inside the capsid. In Chapter 5 we demonstrate that CCMV VLPs encapsidating a heterologous RNA derived from a mammalian virus (Sindbis), are

capable of disassembly in the cytoplasm of mammalian cells. Following this disassembly step and release of the RNA cargo, the RNA undergoes replication and translation similar to that of its naked RNA counterpart transfected under the same conditions. This result establishes the first step in the use of plant viral capsids as vectors for gene delivery and expression in mammalian cells. Furthermore, the CCMV capsid protects the packaged RNA against nuclease degradation and serves as a robust external scaffold with many possibilities for further functionalization and cell targeting.

In Chapter 6, we apply the technique of cryo-EM to directly visualize the global secondary structure of the genomic RNA of satellite tobacco mosaic virus. We compare our results to the predicted secondary structure model generated by modern RNA structure determination methods (10) (selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (11)) and find rough agreement. More specifically, we observe an ensemble of particularly extended molecules by cryo-EM that is collectively consistent with the elongated, relatively unbranched model predicted by SHAPE.

REFERENCES

1. Bancroft, J., and E. Hiebert. 1967. Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. Virology 32:354-356.
2. Cadena-Nava, R. D., M. Comas-Garcia, R. F. Garmann, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart. 2012. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J. Virol. 86:3318-3326.

3. Adolph, K., and P. Butler. 1976. Assembly of a spherical plant virus. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 276:113-122.

4. Lavelle, L., M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, and J. Ruiz-Garcia. 2009. Phase diagram of self-assembled viral capsid protein polymorphs. J. Phys. Chem. B 113:3813-3819.

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

6. Tama, F., and C. L. Brooks. 2002. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. J. Mol. Biol. 318:733-747.

7. Hiebert, E., and J. B. Bancroft. 1969. Factors affecting the assembly of some spherical viruses. Virology 39:296-311.

8. Adolph, K., and P. Butler. 1977. Studies on the assembly of a spherical plant virus III. Reassembly of infectious virus under mild conditions. J. Mol. Biol. 109:345-357.

9. Caspar, D. L. D., A., Klug. 1962. Cold Spring Harbor Symp. Quant. Biol. 27:1–24.

10. Athavale SS*, et al.* 2013. In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One* 8:e54384.

11. Low, J. T. & Weeks, K. M. 2010. SHAPE-directed RNA secondary structure prediction. *Methods* 52:150-158.

12. Carrillo-Tripp, M., Shepherd, C. M., Borelli, I. A., Venkataraman, S., Lander, G., Natarajan, P., Johnson, J. J., Brooks, C. L. III, and Reddy, V.S. 2009. Nucleic Acid Research 37, D436-D442

CHAPTER 1

Self-Assembly of Viral Capsid Protein and RNA Molecules Ranging in Length from 140 to 12,000 Nucleotides: Requirement for a Specific High Protein/RNA Mass Ratio

ABSTRACT Virus-like particles (VLPs) can be formed by self-assembly of capsid protein (CP) with RNA molecules of increasing length. If the protein "insisted" on a single radius of curvature, the capsids would be identical in size, independent of RNA length. But there would be a limit to length of the RNA, and one would not expect RNA much shorter than native viral RNA to be packaged unless multiple copies were packaged. On the other hand, if the protein did not favor pre-determined capsid size, one would expect the capsid diameter to increase with increase in RNA length. Here we examine the self-assembly of CP from cowpea chlorotic mottle virus (CCMV) with RNA molecules, ranging in length from 140 to 12,000 nucleotides (nt). Each of these RNAs is completely packaged if and only if the protein:RNA mass ratio is sufficiently high; this critical value is the same for all the RNAs, and corresponds to equal RNA and N-terminal-protein charges in the assembly mix. For RNAs much shorter in length than the 3000 of the viral RNA, two or more molecules are assembled into 24-nm and 26-nm diameter capsids while for much longer RNAs (>4500nt) a single RNA molecule is shared/packaged by two or more capsids with diameters as large as 30 nm. For intermediate lengths a single RNA is assembled into 26-nm diameter capsids, the size associated with T = 3 wildtype virus. The significance of these assembly results is discussed in relation to likely factors that maintain T=3 symmetry *in vivo*.

INTRODUCTION

Many mature infectious viruses with positive-sense RNA genomes are known to consist of a single copy of the genome inside a one-molecule-thick shell of protein (the capsid protein, CP). In many instances this nucleocapsid structure is believed to arise spontaneously, from self-assembly of the protein around its associated nucleic acid, a process that has been mimicked *in vitro* by direct mixing of the two purified components (i.e., CP and RNA) in an optimized physiological buffer. A particularly well-studied example is that of cowpea chlorotic mottle virus (CCMV), first synthesized in the laboratory in 1967 (8). The native virus has a multipartite genome encoding four genes contained in three sinle-stranded (ss) RNA molecules of 3171 nucleotides (nt) – RNA1, 2774 nt – RNA2, and 2173 nt – RNA3 (2, 18). The first two genomic RNAs are packaged alone and the third genomic RNA is co-packaged with a subgenomic RNA4 of 824 nt, so that each CCMV virion contains about 3000 nt. Just two years later it was demonstrated that CCMV CP was also capable of packaging *heterologous* ssRNA, as well as synthetic anionic polymers (9). A more extensive study of CCMV CP assembly around four heterologous viral RNAs ranging in length from 4000 to 6400 nucleotides (nt) and around short homopolynucleotides (poly(A), poly(C) and poly(U)) was carried out by Adolph and Butler (1).

Similarly, the CP of a closely related virus, brome mosaic virus (BMV), has been shown to package gold nanoparticles of various sizes that have been functionalized with anionic moieties (43) and a negatively-charged near-infrared chromophore (26). Also, we have investigated the packaging by CCMV CP of poly(styrene sulfonate) (PSS) polymers with molecular weights ranging from 38 kDA to 3.4 MDa (12,23). We chose to examine PSS because its physical properties are well established: it has a persistence length and charge density comparable to that of ssRNA, has solution properties (e.g., radius of gyration) that are well

understood, and is readily available in a large range of molecular weights. Carrying out assemblies of virus-like particles (VLPs) for mixes of CCMV CP with PSS of various molecular weights, we determined the effects of polymer length on the formation of VLPs of different sizes (23), and explained the observed behavior in terms of a competition between the polymer-polymer and polymer-capsid interactions and the preferred curvature of the CP, as discussed by Zandi and van der Schoot (49). In particular, we found that a large range of PSS lengths (and hence charges) could be accommodated by the same size protein capsid, up to a certain threshold length at which larger capsids begin to form as the dominant assembly product (12,23).

In the work reported here we extend this study of VLP self-assembly to include a broader range of polyanion lengths and capsid sizes. Also, while we continue to use the same (CCMV) CP, we now feature – instead of PSS – ssRNA molecules with molecular weights varying over a larger range, from about 45 kDa (140 nt) to 4 MDa (12,000 nt).

There are several reasons for working with ssRNA instead of PSS. First, ssRNA is the genome of CCMV and of a majority of viruses pathogenic to plants, animals and humans. Second, from a *physical* point of view, RNA is not a linear polymer like PSS. Self-complementary base-pairing interactions in ssRNA lead to a branched structure containing double-stranded and single-stranded regions that are organized with respect to each other into complex tertiary structures. As a consequence, the radius of gyration and shape of the RNA increase and change with molecular weight in ways that are difficult to predict (48), and it must be expected to interact very differently with CP (10) than does a linear homopolymer like PSS.

The structures of the cDNA templates for the RNAs whose self-assembly with CCMV CP is studied in the present work are shown schematically in Fig. 1-1. The shorter molecules consist of BMV RNA1 (3,200 nt) and RNAs (140, 500, 1000, 1,500 and 2,000 nt) obtained from

it by successively longer truncations at the 3' end. The longer RNAs include: the 5,300 nt ligation product of BMV RNA 1 with the 2117 nt BMV RNA 3; a 4,500 nt fragment lacking the 3' end of this composite; the genome of tobacco mosaic virus (6,400 nt); and two RNAs derived from the RNA genome of the mammalian virus, Sindbis – a 8600-nt RNA made from replacement of its structural protein genes with the enhanced yellow fluorescent protein (EYFP) gene, and the 11,703 nt genome of Sindbis itself. While packaging signals have been identified for the assembly of BMV CP protein around BMV RNA (38), no such sequences have been identified for CCMV (3). In any case, by using CCMV CP to package RNAs derived from BMV, we have eliminated the possible effects of any specific CCMV packaging signals, allowing us to focus on the generic effects of size and charge.

We find that RNAs of *any* length can be completely packaged as long as the protein:RNA mass ratio is sufficiently high. This critical value turns out to be independent of the RNA length, and to involve equal and opposite N-terminus CP and phosphate RNA charges being brought to the assembly mix by this (6:1) mass ratio. Upon increasing the length of the RNA, we find that capsids increase in size, corresponding to three different diameters, determined by EM to be approximately 24, 26 and 30nm.

Further, RNA molecules much shorter than the wildtype (wt) viral length are found to be packaged in multiple copies, whereas much longer RNAs (>4500nt) give rise to assemblies consisting of two or more capsids being formed around them. These results are interpreted in terms of two different physical phenomena. The first involves the preferred capsid-forming curvature of the CCMV CP – in particular the fact that T = 3 capsids are strongly preferred (23, 29, 36), but with slightly smaller and slightly larger diameters also being tolerated. The second physical factor involves the relative magnitudes of the energies of CP-CP interaction and of CP-

RNA interactions. In particular, the large interaction energy of the protein with RNA leads to a saturation of this binding, independent of RNA length, as long as a sufficient excess of protein (here, a 6:1 mass ratio) is present in the assembly mix. For short RNAs this leads to two or more RNAs being packaged in capsids with average size smaller than wildtype; for wildtype lengths (around 3000 nt), a single molecule is packaged in a capsid of wildtype diameter; for longer molecules, a single RNA is packaged by two or more capsids.

MATERIALS AND METHODS

Materials

Restriction enzymes were obtained from New England Biolabs. SP6 RNA polymerase was obtained from AMBION and T7 RNA polymerase was a gift from Prof. Feng Guo (Dept. of Biological Chemistry, UCLA). Enzymes were used as recommended by the manufacturer. The plasmids were grown, purified and analyzed by standard methods (39). Plasmids pT7B1 and pT7B3Sn contain full-length cDNA clones corresponding to BMV RNA 1 (B1) (3234 nt) and BMV RNA 3 (B3) (2117 nt), respectively (17), plasmids p30B, pEYFP-Rep and pTE12 encode the full length of TMV RNA (6395 nt) (15), a Sindbis-derived replicon RNA containing enhanced yellow fluorescent protein (8600 nt) and full-length Sindbis RNA (11,700 nt), respectively (30). The plasmids were grown, purified, and analyzed by standard methods. All other chemicals used were DNase,- RNase- and Protease-free.

Combining B1 and B3 RNAs by directional sub-cloning

Directional sub-cloning of pT7B1 requires the plasmid to have overhangs; these were created by linearization with *Bam*HI and *Eco*RI. The insert that encodes B3 was engineered from

pT7B3Sn by PCR; forward primer d(TTGCGCGGATCCAACTAATTCTCGTTCG) (*Bam*HI site is underlined) and reverse primer d(TGGCCGGAATTCAACACTGTACGGTAC) (*Eco*RI site is underlined). These primers give 8 and 28 fewer nt at the 5' and 3' ends respectively than genomic BMV RNA 3. The resulting PCR product was double-digested with *Eco*RI and *Bam*HI, and subcloned into pT7B1 previously digested with *Bam*HI/*Eco*RI to give the new cloning cDNA , pRDCT7B1B3. The authenticity of ligated product, RNA B1 + B3, was confirmed by restriction analysis followed by sequencing.

PCR amplification of DNA templates to transcribe short RNAs

DNA templates with lengths smaller than full-length BMV RNA 1 were prepared by PCR of pT7B1 plasmid; all truncations were carried out from the 3'-end. We used a universal 5' primer d(TAATACGACTCACTATAGGTAGACCACGGAACGAGGTTC) (T7 promoter is underlined) and as reverse primers: d(GCAATCAACTTCAGCAAATCG), d(CACATCCTCTCCTCATGTC), d(GTCTTCAAACCATACACAGTG), d(CTTGCTCAAATTCTTCAACG), and d(GGATAC AACCAGTTACCGTTG) to obtain the corresponding DNA templates for transcribing RNAs of length 141, 499, 999, 1498 and 1960 nt, respectively.

RNA transcription

Plasmids pT7B3Sn, PT7B1 were linearized with *Bam*HI, p30B and pTE12 with *Xho*I and pEYFP-Rep with *Sac*I. pRDCT7B1B3 was linearized with *Bgl*II and *Eco*RI to produce RNA transcripts of 4452 and 5319 nt, respectively. All the templates were purified by standard procedures (39). All DNA templates and plasmids encoding for RNAs between 140 to 6395 nt

were transcribed with T7 RNA polymerase; plasmids pEYFP and pTE12 were transcribed with SP6 RNA polymerase.

CCMV CP purification

CCMV was purified from infected California cowpea plant (*Vigna ungiculata* cv Black Eye) (7) and capsid protein was isolated largely as described previously (3). Nucleocapsids were disrupted by 24-hr dialysis against disassembly buffer (0.5 M $CaCl_2$, 50 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, and 0.5 mM PMSF) at 4°C and then by ultracentrifugation at 100k rpm for 100 min (4 x 10^5 g) at 4 °C in a Beckman TLA110 rotor. The RNA was pelleted and the CP was extracted from the supernatant in different fractions. Each fraction was immediately dialyzed against protein buffer (1M NaCl, 20mM Tris pH = 7.2, 1 mM EDTA, 1mM DTT, and 1 mM PMSF). The protein concentration and its purity, with respect to RNA contamination, were measured by UV-Vis spectrophotometry; only protein solutions with 280/260 ratios greater than 1.5 were used for assembly. SDS-PAGE gel electrophoresis and MALDI-TOF were employed to ascertain that the purified protein was not cleaved.

In vitro assembly: Preliminary titrations and gel retardation assays

Assembly reactions of CCMV CP with RNAs of different lengths were performed in RNA assembly buffer (henceforth referred to as RAB; 50 mM NaCl, 10 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 50 mM Tris-HCl pH 7.2). CCMV CP was titrated at different weight ratios into a constant RNA concentration (30ng/µl). After overnight assembly at 4 ^oC, a 10 µL aliquot of each ratio was mixed with 2 µl of 100% glycerol (RNase, DNase, and Protease free) and loaded into a 1% agarose gel (EMD) in either virus buffer (0.1 M sodium acetate, 1 mM EDTA, pH 4.8) or

electrophoresis buffer (0.1 M sodium acetate, 1 mM EDTA, pH 6). The samples were electrophoresed for 1.25 hrs at 70 V (electrophoresis buffer) or 4 hrs at 50 V (virus buffer) in a horizontal gel apparatus (Fisher) at 4 °C. They were then stained in a solution of 5 μg/ml of ethidium bromide and visualized with an Alphaimager system.

In vitro assembly of VLPs

Dissociated CCMV CP subunits and desired RNA (30 ng/μl) transcripts were mixed in a ratio (wt/wt) of 6:1 and then dialyzed for 24 hrs at 4 °C against RAB. The samples were then dialyzed against virus suspension buffer (50 mM sodium acetate, 8 mM magnesium acetate pH 4.5) for at least 4 hrs. For assemblies with RNAs longer than BMV RNA1, 100-300 μL of the assembly reactions were subjected to sucrose gradient (10-40% [wt/wt]) centrifugation in virus buffer using an ultracentrifuge SW 50.1 rotor at 33k rpm (1.3×10^5 g) for 2 hr. The fraction corresponding to the VLPs was recovered and dialyzed overnight against virus buffer or virus suspension buffer to remove the sucrose. VLPs from all RNA assemblies were purified and concentrated by washing with virus suspension buffer using a 100 kDa Amicon centrifuge filter (0.5 ml) at 3000g for 5 min; this last step was repeated 4 times. All the procedures were carried out at 4 ^{o}C. Finally, the samples were characterized by UV-Vis spectroscopy with a NanoDrop spectrophotometer (Fisher).

TEM analysis of VLPs

For negative staining, purified VLPs were applied to glow-discharged copper grids (400-mesh) that previously had been coated with Parlodion and carbon. A 6 ml aliquot of VLPs was spread onto the grid for 1 min, blotted with Whatman filter paper and then stained with 6 μL of

1% uranyl acetate for 1 min. Excess stain was removed by blotting with filter paper. The samples were stored overnight in a desiccator and analyzed with a JEM 1200-EX Transmission Electron Microscope equipped with a wide-angle (top mount) BioScan 600 W 1 × 1K pixel digital camera operated at 80 keV. The reported average diameter of VLPs is that of the geometric mean of two orthogonal measurements of the capsids obtained with ImageJ (U.S. National Institutes of Health) software from recorded images.

RESULTS AND DISCUSSION

The packaging experiments were preceded by studies to identify the optimal molar ratio of CP to RNA for packaging. This is the ratio at which there is just sufficient protein to package all of the RNA into RNase-resistant VLPs. It was identified by carrying out, in parallel, assembly reactions in RAB using a variety of protein:RNA wt:wt ratios. Each fixed-ratio protein:RNA solution was allowed to equilibrate for 24 hr, and then dialyzed against virus suspension buffer before being analyzed by agarose gel retardation assays. The results in each case – as illustrated in Fig. 1-2 by the assemblies with BMV RNA1 (3200 nt) and TMV RNA (6400 nt) – were similar. As protein is added to the RNA (i.e., moving from right to left in the gels) the bands, which indicate small protein-RNA aggregates for low CP:RNA wt:wt ratios, first migrate faster and then increasingly more slowly as the ratio increases, finally appearing at the same position as wt CCMV. Note that it has been shown in the case of CCMV that the electrophoretic mobility depends only on the charge on the exterior of the capsid and not on its contents, with empty capsids migrating at precisely the same position as virions containing RNA (25). Hence a band that migrates at the same position as wt CCMV indicates that there is an

intact capsid. (Note, however, that in the case of the 6400nt RNA the band migrates more slowly than that of CCMV; the reason for this will be discussed below.)

The results of the assembly studies lead to an unexpected conclusion: in every case, the optimal assembly weight ratio of protein to RNA is 6:1, *independent* of the length of the RNA. A comparable ratio was found in the *in vitro* packaging of CCMV RNA1 by CCMV CP; in particular, Johnson et al. (24) reported capsid-protein titration gels similar to those shown in Fig. 1-2, suggesting that at least 130 protein dimers per RNA (a weight ratio of 5:1) are required for the complete packaging of RNA1. This is also consistent with early work by Adolph and Butler (1), who reported a significant fraction of free RNA in their reassembly studies of CCMV involving a mass ratio of 5:1 for CP:RNA.

Simple analysis (see below) shows that the "magic ratio" value of 6:1 is associated with a condition of charge equality between the RNA and the N-terminal arginine-rich motif of the CP. Analytical ultracentrifugation results (unpublished data) confirm the gel titration results, establishing the magic ratio as the CP:RNA threshold for assuring complete packaging of the RNA, independent of RNA length. We emphasize that this 6:1 mass ratio *does not depend on any assumptions or models*; we are simply observing that this threshold composition in the assembly mix is required for there to be no free RNA remaining.

Again without any assumptions or models, we can say more, on a microscopic level, about the meaning of this special mass ratio. We can write the total mass of protein in the solution as $M_{CP} = n_{CP} MW_{CP}$ where n_{CP} is the total number of moles of CP and MW_{CP} is the CP molecular weight (about 20,000 Da). Similarly, for the total mass of RNA we write $M_{RNA} = n_{RNA} L_{RNA} MW_{nt}$ where n_{RNA} is the total number of moles of RNA, L_{RNA} is the RNA length in nt and MW_{nt} is the average MW of a nt (about 330 Da). It follows that the ratio of the number

of CP subunits (n_{CP}) to the number of nt ($n_{nt} = n_{RNA} L_{RNA}$) is given by $\frac{n_{CP}}{n_{nt}} = \frac{MW_{nt}}{MW_{CP}} \frac{M_{CP}}{M_{RNA}} = \frac{330}{20{,}000} \frac{6}{1} = \frac{1}{10}$. This implies that we need one CP subunit for every 10 RNA nucleotides in order for there to be no free RNA present in the assembly mix. It does not imply that at this point we have no free protein, nor anything about the amount of cationic charge from protein that is interacting with RNA or otherwise involved in the assembly process.

Nevertheless, the fact that at the magic ratio one CP subunit is present for each 10 nt of RNA implies (since each CP N-terminus has 10 cationic residues) that 10 N-terminal cationic residues are present for each 10 nt. While there are other cationic residues in the CPs that could be involved in RNA binding, several different kinds of experiment (45,14) imply strong interaction of the 10 N-terminal basic residues of the CPs with RNA, with binding energies on the order of 10 k_BT. This is also consistent with charge matching of oppositely-charged colloidal particles – for example, the nonspecific binding of proteins to DNA (16) – being the main driving force for their association/binding in aqueous solution. More explicitly, due to the entropy gain associated with mobile counterion release (31), the free energy of polycation/nucleic acid association is roughly k_BT per charge. For the +10 CP N-termini this suggests binding energies of order 10 k_BT, implying a saturation of binding, from which we conclude that super-stoichiometric numbers of proteins are bound – specifically, 1 per 10 nt, e.g., 300 per wildtype length of RNA rather than the 180 that eventually form the ordered shell of a T=3 capsid.

The need for an amount of protein in excess of the stoichiometric quantity has also been demonstrated in several self-assembly studies involving CCMV CP and *heterologous* RNAs. For example, Adolph and Butler (1) found unencapsidated RNA following reaction of CCMV and RNAs from Turnip Crinkle Virus (4051 nt), Bushy Stunt Virus (4776 nt), Turnip Yellow Mosaic

Virus (6318 nt) and Tobacco Mosaic Virus (6400 nt) at CP:RNA weight ratios of 4:1, exceeding in all cases the stoichiometric ratios expected for T=3 capsids. Even if the longer RNAs assembled into T=4 capsids, the stoichiometric ratios would still be well below the 4:1 employed.

The 6:1 weight ratio for complete packaging of RNA was also found by Kobayashi and Ehara (27) in their *in vitro* encapsidation studies of the closely related plant virus, cucumber mosaic virus (CMV), whose CP N-termini also carry a charge of +10. Porterfield, et al. (35) have used gel assays to examine the *in vitro* assembly of hepatitis B virus (HBV) core protein (molecular weight 21,000) around CCMV RNA 1 and HBV pregenomic RNA, both 3200 nt long, and around xenopus elongation factor RNA, 1900 nt in length. The HBV protein has 17 arginines at its N-terminus. Taking this into account, we expect the optimal weight ratio in these experiments to be 3.6. The observed ratios are indeed about 4.

VLP products as a function of RNA length

Fig. 1-3 shows a panel of typical negative stain electron micrographs of the assembly mixes corresponding to 6:1 CP:RNA mass ratio, for all 11 of the increasing lengths of RNA studied. The lower right-hand picture is wt CCMV. The numbers of capsids sharing a single RNA molecule (in the case of the large nt lengths), the sizes of VLPs formed in each instance (see Fig. 1-4A), and the numbers of RNAs packaged per VLP (in the case of the shorter RNA lengths), are discussed below.

When Verduin and Bancroft (47) examined the assembly of CCMV CP around TMV RNA they found some difficulty in assigning a T number to the VLPs that formed. The TEM data agreed with T=4 but sedimentation analyses were more consistent with T=7. We have already noted that the gels for the assembly of CCMV CP around TMV RNA seem to show a

product at the magic ratio that migrates more slowly than wt CCMV, suggesting too that the capsid may be larger than T=3. A careful examination of the EMs for the TMV VLPs makes clear the reason for this ambiguity. The majority of the capsids are paired – see, for example, the 6400 nt frame in Fig. 1-3 and the upper right-hand image ("B") in Fig. 1-4B. Such doublets, which are most evident when the density of capsids in the image is not so high that the particles are all closely packed, could also be seen clearly in the early studies of TMV RNA packaging by CCMV CP (22), but went unremarked. The doublets also account for the appearance of two sedimentation bands in Adolph and Butler's packaging studies of TMV RNA by CCMV CP (1).

Formation of multiplets

As shown in Fig. 1-4B, the multiplets first appear as a small fraction of doublets for 3200 nt RNA. The fraction of doublets observed in the micrographs increases markedly at 4500 nt and at the same time a few triplets can be seen. The number of singlet capsids decreases markedly with increasing RNA length and, successively, the fractions of doublets and triplets rise and go through maxima, cf. Fig. 1-4B. Assemblies with the 11,700-nt Sindbis RNA lead primarily to quadruplets and higher-order multiplets. (Note that each of the average diameters plotted in Fig. 1-4A contains contributions only from the capsids associated with the predominant multiplet, i.e. singlet, doublet, triplet or quadruplet, for the corresponding RNA length.)

Doublet and triplet capsids have been observed in *in vivo* studies of both animal and plant viruses. Field Emission Scanning Electron Microscope images of a recombinant adenovirus show mixtures of singlet, doublet, triplet and some higher multiplet capsids (33). After isolation, the virus was stored at -80 °C and incubated at 37 °C after thawing. The fraction of doublets and triplets remained relatively constant at about 18 % over a period of 4 hr, but the number of

singlets decreased from 80 to 60% due to the appearance of higher-order multiplets. Unlike the doublets and triplets, the higher-order multiplets in this study appeared to be aggregates of virus and proteinacious material. The doublet structures have the superficial appearance of geminiviruses (50), which often form triplets and quadruplets (11). In the case of the African cassava mosaic geminiviruses containing DI ssDNA, the number of multiplets increases with DNA length (21).

The multiplet VLPs that we observe in CCMV VLPs are unlikely to have structures akin to those of gemini viruses, in which capsids share missing pentamer faces. This is evident when the multiplets are treated with RNase, which has the effect of separating them into single shells, cf. Fig. 1-S1**,** for the case of quadruplets forming from Sindbis RNA. Clearly, then, these multiplet capsids do not form a continuous shell; they are more likely two or more capsids that share a single molecule of RNA that is threaded through nm-sized holes in the capsids. Indeed, in some EM images, cf. Fig. 1-S2, a bit of RNA can be seen poking out between some of them.

Mechanism for the formation of multiplets

The regular appearance of increasingly higher-order multiplets with increasing RNA length is evidence that the multiplets do not simply arise from physical aggregation of assembled capsids. Moreover, the multiplets often have highly regular structures, such as tetrahedral for the quadruplets (see "D" in upper part of Fig. 1-4B), which do not appear in large randomly organized aggregates. Also, analyses of electron micrographs of multiplets, upon successive dilutions by factors of 2 and 4 of an assembly mix involving CP and Sindbis RNA show (see Fig. 1-S3) that the fraction of particles of each multiplicity is independent of dilution, indicating further that the multiplets are not simply physical aggregates.

Recent coarse-grained simulations of capsid self-assembly around a linear polymer indicate (19) that the above situation corresponds to the excess protein being bound to the polymer, saturating its adsorption sites. Indeed, the fact that all of the N-terminal charge on the CP is compensated by the RNA charge is consistent with saturation of binding, i.e., the absence of free protein at the magic ratio. This situation in turn corresponds to capsid protein-polymer interactions being dominant, compared to protein-protein interactions. More explicitly, Elrad and Hagan (19) report computational results for this case in which the first step in nucleocapsid assembly involves the disordered binding of excess protein, followed by *de*sorption of protein with reorganization of the remaining protein as a closed, icosahedral, capsid including the polymer.

This scenario is consistent with measurements of the differences in diameter between partners in a multiplet, Fig. 1-S4. While in the majority of the cases the capsid diameters are identical within the experimental uncertainty of ± 1 nm, there is a significant fraction in which the size differences are larger, suggesting that the capsids share different fractions of the RNA. The decrease in the fraction of singlets and the successive rise and fall of the fractions of doublets and triplets and the rise of the fraction of quadruplets is qualitatively consistent with a model in which protein adsorption onto RNA is saturated, with fluctuations leading to unbinding of excess protein and simultaneous formation of two or more capsids once the RNA is close to twice the wildtype length or longer.

Distributions of diameters determined from electron micrographs are shown by the histograms in Fig. 1-5 for the eleven RNAs studied. Also included in the figure for comparison is a measurement for wt CCMV, for which the average number of packaged nucleotides is 3000. Note that it is difficult to determine the intrinsic sizes of the VLPs in solution from the electron

micrographs because, as is well known, there is shrinkage of capsids associated with the drying that occurs when the sample is deposited and stained on the grid. For example, as seen in the figure, wt CCMV, known from cryo-EM to be 28 nm in diameter in solution (20), is found by TEM to be 26 nm in diameter.

The observed capsid diameters appear to fall into three populations with diameters of 24, 26 and 30 nm. To assign these sizes unequivocally to T numbers (13) requires structural determinations. Nevertheless it is possible to make tentative assignments on the basis of size alone and the knowledge that the RNAs packaged in single capsids are resistant to nuclease, indicating that they are contained in closed structures. Clearly those capsids with diameters clustered around 26 nm are similar to those of wt CCMV and therefore can be assumed to have T = 3 triangulation numbers. Those around 24 nm in diameter fall into a size range that has been identified with the pseudo Caspar-Klug "T = 2" structure composed of pentamers of dimers in assemblies of BMV (28). None of the VLPs formed have diameters smaller than 20 nm, the value consistent with a T = 1 triangulation number as has been observed in assemblies of BMV (44) and of virus-like particles formed from CCMV protein lacking the N-terminus (40).

While there have been reports of CCMV capsids with T > 3, this has not been demonstrated by structural determinations. As previously noted, we believe that the assignment, based on estimates of the sedimentation coefficient, of a triangulation number of 7 to capsids formed in the assembly of CCMV protein around TMV RNA [47] was faulty because of the presence of doublets. Sun, et al. [43], in their study of BMV protein assembly around functionalized gold nanoparticles, have observed a population of particles with diameters distinctly larger than those that have been identified as T = 3. But the distribution of sizes was very heterogeneous and fewer than half of the gold particles had complete capsids; no attempt

was made to assign a T number to them. The larger diameter capsids that we have observed, however, are homogeneous in their distribution, well-formed and closed, as evidenced by the protection they afford against RNase to the packaged RNA. If one assumes that the average area occupied by a capsid protein does not change with T number, it follows that the capsid diameter should be proportional to $T^{1/2}$ [23]. Thus if the measured diameter of a T = 3 capsid is 26 nm, the diameter of a T = 4 capsid would be expected to be 26nm $(4/3)^{1/2}$ = 30 nm, in accord with the size in the distributions that we have observed. Assignment of T = 4 to this population therefore seems reasonable. (Note that the predicted diameter of a T = 7 particle is about 40 nm, a size inconsistent with the observed diameters.)

The appearance of multiplets rather than increasingly larger diameter capsids is attributable to the preference of CCMV CPs to form T = 3 shells, which is a consequence of the protein's preference for a particular curvature. It is possible to form capsids with diameters that differ from those with the preferred curvature, but only at a cost of free energy. We have previously shown (36) how this energy cost sets a limit on the number of shells that can be formed in the multishell particles that arise in the self-assembly of CCMV CP in the absence of RNA (29). Thus, as can be seen from Fig. 1-3 and the accompanying size distributions shown in Fig. 1-5, RNAs longer than the 3000 nt that is packaged in the wt virus can be accommodated in a T = 4 capsid (and RNAs smaller than the wt in a T = 2 capsid). But still larger molecules, if they were packaged in a single particle, requiring T = 7 capsids, demand too high a curvature-energy cost. The packaging of the RNA in two or more capsids of smaller diameter avoids this energy expenditure. (We attribute the appearance of a very small fraction of singlet capsids in the packaging of longer RNAs to a small amount of RNA degradation.)

For RNAs shorter than BMV RNA 1 (3200 nt) the scenario is different. VLPs containing 2000, 1500, 1000 and 140 nt RNAs are a mixture of T = 2 and T = 3 capsids, while the 500 nt RNA is predominantly packaged into T = 2. The EM data for almost all of these samples (Fig. 1-5) show a broad VLP size distribution with a mean diameter that is intermediate between what we would expect for T = 2 and 3 capsids, indicating that both populations are significantly represented. While it is difficult to fit these size distributions into two populations, it is clear that not all VLPs assembled from shorter RNAs have the same T = 2 and 3 populations; in particular, the observations suggest that for the 2000 and 140 nt RNAs the T = 3 capsids are preferred over T = 2, with the opposite holding for 1500 nt and 1000 nt RNA. Only the 500 nt RNA produced a symmetric and relatively narrow distribution, with an average diameter of 23 nm, as expected for T = 2 capsids (see Fig. 1-5). From the gel retardation assays, we know that at the magic ratio all the RNA gets packaged, suggesting that for RNAs between 140 and 1500 nt more than one molecule is packaged.

To examine this possibility we measured the UV-Vis absorbance of the purified VLPs. Proteins and nucleic acids exhibit a strong UV absorption at 280 and 260 nm, respectively, so the ratio of the absorbance at 260 over 280 nm in a pure VLP solution tells us the relative amounts of RNA and CP within a capsid. Since CCMV and BMV have the same capsid size and average number of nucleotides per capsid, they have the same 260/280 absorbance ratio. By comparing this ratio with that of the VLPs and by knowing the relative capsid size populations we obtain an initial estimate of the number of small RNAs packaged in a capsid. The 260/280 ratios for VLPs containing 3200, 2000, 1000 and 500-nt RNAs are indistinguishable from that of CCMV, from which we conclude that there is only one 3200 and 2000 nt RNA molecule per T = 3 and T = 2 capsid, respectively. (The number of CPs is proportional to the T number.) We can also conclude

that *on average* there are two 1000 nt RNA molecules per T = 2 capsid. Similarly, since the capsids containing 500 nt RNA are T = 2, there are on average four 500 nt RNA molecules per capsid. The only sample that had a ratio lower than CCMV was the one containing 1500 nt RNA (1.5, vs 1.7 for CCMV), strongly suggesting – since T = 2 is preferred over T = 3 for this length – that there is a single molecule in T = 2.

It is interesting that no T=1 VLPs are observed, even for the shortest (140-nt) RNAs. In early work Pfeiffer et al. (34) reported the formation of T = 1 VLPs involving *BMV* CP, but no information is provided on the lengths of the ssRNAs. Also, in our studies of the packaging of short (38 kDa), fluorescently (Rhodamine)-labeled, PSS molecules by CCMV CP, we found T = 1 VLPs as the predominant product (12). And yet rough estimates of the size (radius of gyration, R_g) of the 38 kDA-PSS and of the 140 nt RNA suggest that the RNAs are smaller. It may be that the several-nm, rigid-ring, rhodamine label introduces a very different scenario for interaction of the anionic polymer with the capsid interior and hence for the preferred curvature of the VLP. Also, the branching of the RNA due to its significant secondary structure results in a different distribution of charge than for PSS. For these reasons we expect that the preferred diameter of VLP will depend not only on the size and charge of the polymer but also on its structural nature and resulting interaction with protein.

CONCLUSIONS

In summary, we have shown that RNA of any length – varying from 140 to 12,000 nt – can be packaged completely by CCMV CP as long as the protein:RNA mass ratio is as large as 6:1. For sufficiently short lengths (<2000 nt) the RNA is packaged into T = 2 and T=3 capsids with 1-4 molecules of RNA in the capsid. For large enough RNAs (>4500nt) single molecules

of RNA are packaged by two or more T = 3 or T = 4 capsids. And for intermediate lengths, single RNAs are packaged, as with wt length, into T = 3 capsids. These scenarios arise from the capsid curvature preferred by the CP, and from the relative magnitudes of the CP-RNA and the CP-CP interactions.

In vitro self-assembly of single capsids of CCMV having RNAs of ~3000 nt is reminiscent of the *in vivo* scenario. More explicitly, CCMV virions purified from infected leaf tissue display a remarkable homogeneity in size and physical appearance: RNA1 (3171nt) and RNA2 (2774 nt) are each packaged independently into T = 3 capsids whereas RNA3 (2173 nt) and RNA 4 (824 nt) are co-packaged into a third virion. It has been shown that *in vivo* assembly of nucleocapsids in positive-strand RNA viruses pathogenic to humans (32), animals (4,46) and plants (3) is functionally coupled to replication-dependent transcription and translation. Consequently, macromolecular interactions, such as the interaction between CP and viral replicase (RNA1 and RNA2 gene products), resulting from commonly shared subcellular localization sites (6), might play an important role ensuring that only one kind of stable virion population with T=3 symmetry is assembled and maintained.

While the existence of the operationally-defined magic ratio is an experimental fact, we can speculate on its implications for understanding the mechanism of viral assembly. We note that the relation of the ratio to a charge balance involving only the N-terminus basic residues seems to be at odds with the structural studies by Speir, et al. (42). They found that 13 residues other than those on the protein tail interact with the RNA, seven of them basic, and that the RNA "forms three distinct bulges directly under the shell near the quasi three-fold axes where it interacts with Lys87 residues." It is not evident, however, that these contacts exist *before* a capsid is assembled. Indeed, the quasi 3-fold axes lie between 2 hexamers and a pentamer and at

the stage in assembly that we have identified in which there is an excess of protein associated with the RNA with respect to the number of proteins necessary to form a capsid, it is unlikely that distinct capsid pentamers exist. Moreover, the structure of the protein dimers in solution has never been determined – images showing the protein conformations are simply those abstracted from the assembled capsids.

Although the simulations of assembly by Elrad and Hagan (19) are for a simplified model in which capsomer subunits adsorb onto a linear polymer represented as a freely jointed chain of spherical monomers, they can provide a basis for beginning to understand the mechanism underlying the magic ratio. They demonstrate that the RNA plays a significant role in assembly, so that the mechanism for the assembly of *empty* capsids, which has been examined in detail by Zlotnick, et al. (51) is not likely to be relevant to assembly around RNA. Two limiting scenarios have been distinguished in Elrad and Hagan's work: nucleation of a capsid "embryo" involving a small number of proteins, followed by growth; and an "en masse" adsorption of protein subunits onto the polymer that approaches or exceeds the stoichiometric numbers required for the assembly, followed by a cooperative rearrangement to form a capsid. The controlling factors are the protein subunit-subunit interaction energy and the protein-RNA binding energy. The en-masse mode of assembly is favored when the subunit interaction energy is much weaker than the driving force for subunit adsorption onto the polymer.

A nucleation and growth mechanism has been shown to apply to the *in vitro* assembly of turnip crinkle virus (41). In this case both a specific nucleation center (37) and a nucleating event (the binding of a small protein complex) have been identified. Here a magic ratio would clearly not be expected to be observed.

In contrast, in the case of the assembly of the CCMV capsid protein around RNAs that we have studied, neither a specific site nor a local nucleating event can be identified. More likely, the entire RNA molecule is saturated with bound capsid protein that is present in a 70% excess over that eventually involved in the formation of capsids. Closed capsids then arise by fluctuations that lead to capsid formation and the shedding of excess protein. If the RNA is significantly shorter than 3000 nt – the length of wildtype molecules packaged separately (RNA1 and RNA2) – two or more of them are packaged in capsids whose sizes are consistent with T = 2 and T = 3 structures, i.e., with 24 and 26 nm diameters. For RNAs with lengths around 3000 nt, a single molecule is packaged in a 26 nm capsid. And for RNA lengths greater than 4500 nt, two or more capsids are involved in the packaging of a single molecule.

ACKNOWLEDGMENTS

We thank Dr. Feng Guo for providing the T7 polymerase and Drs. Robijn Bruinsma, Paul van der Schoot and Roya Zandi for helpful discussions. This work was supported by NSF Grant CHE 0714411 (to CMK and WMG) and NIH Grant 1R21AI82301 (to ALNR). RDC-N and MC-G received partial support from CONACyT and UC-Mexus.

FIGURES

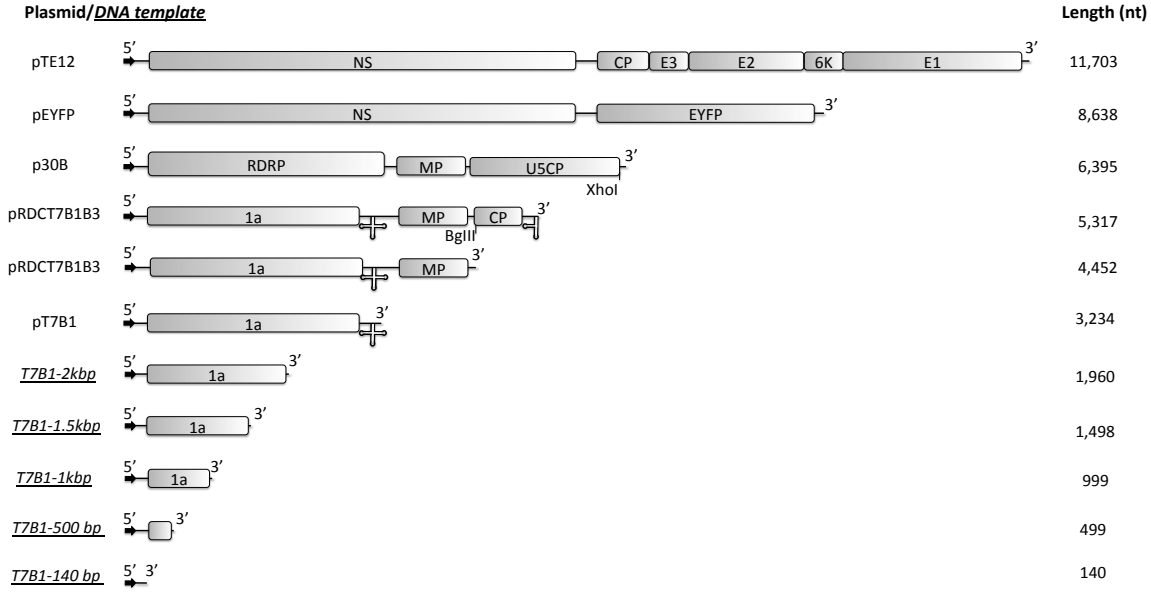

Figure 1-1. Schematic representation of the plasmids and DNA templates used for RNA syntheses. The arrows denote the transcription promoter and the clover leaf at the 3' end represents the highly conserved t-RNA-like structure (3'TLS) in the corresponding RNA transcripts. The boxes represent the open reading frames of the RNAs; 1a is a viral replicase, MP and CP are the movement protein and capsid protein, respectively, E1, E2, E3 and 6K are structural proteins and NS represents all non-structural viral proteins for Sindbis Virus. EYFP is the sequence that encodes for the enhanced yellow fluorescent protein, RDRP denotes the RNA dependent RNA polymerase genes, and U5CP is the CP gene of strain U5 of tobacco mild green mosaic virus. The representations of the plasmid templates are scaled to their relative lengths.

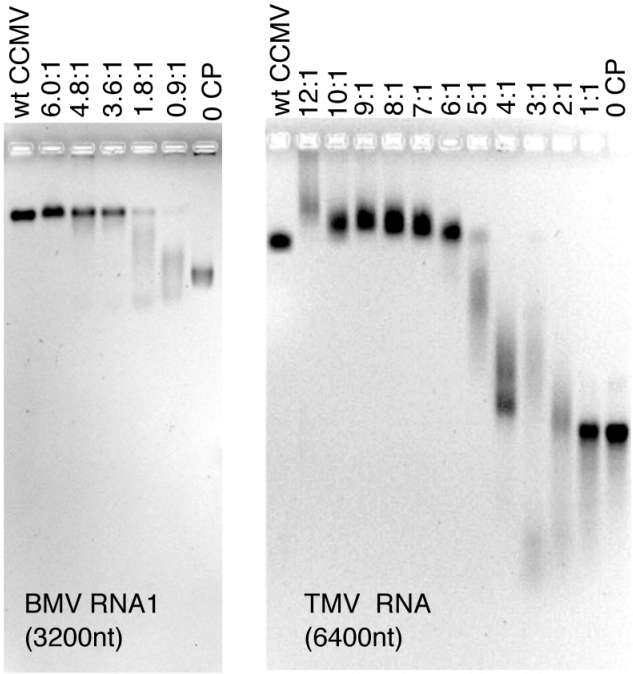

Figure 1-2. CP-RNA assembly titrations: Gel retardation assays. Shown are 1% agarose gels stained with EtBr. At the left is a titration of 3217nt of BMV RNA1 with varying amounts of CCMV CP ranging from 0 (right-most lane, RNA) to the "magic" ratio, 6:1 (lane second from left); CP:RNA ratios are wt:wt ratios of CP to RNA. This gel was run in electrophoresis buffer. Right is a similar gel retardation assay carried out with 6395nt of TMV RNA and run in virus buffer. In both gels the left-most lane shows the position of wt CCMV. Note, in each gel, the appearance of a smear of intensity accompanying every assembly carried out below the magic ratio.

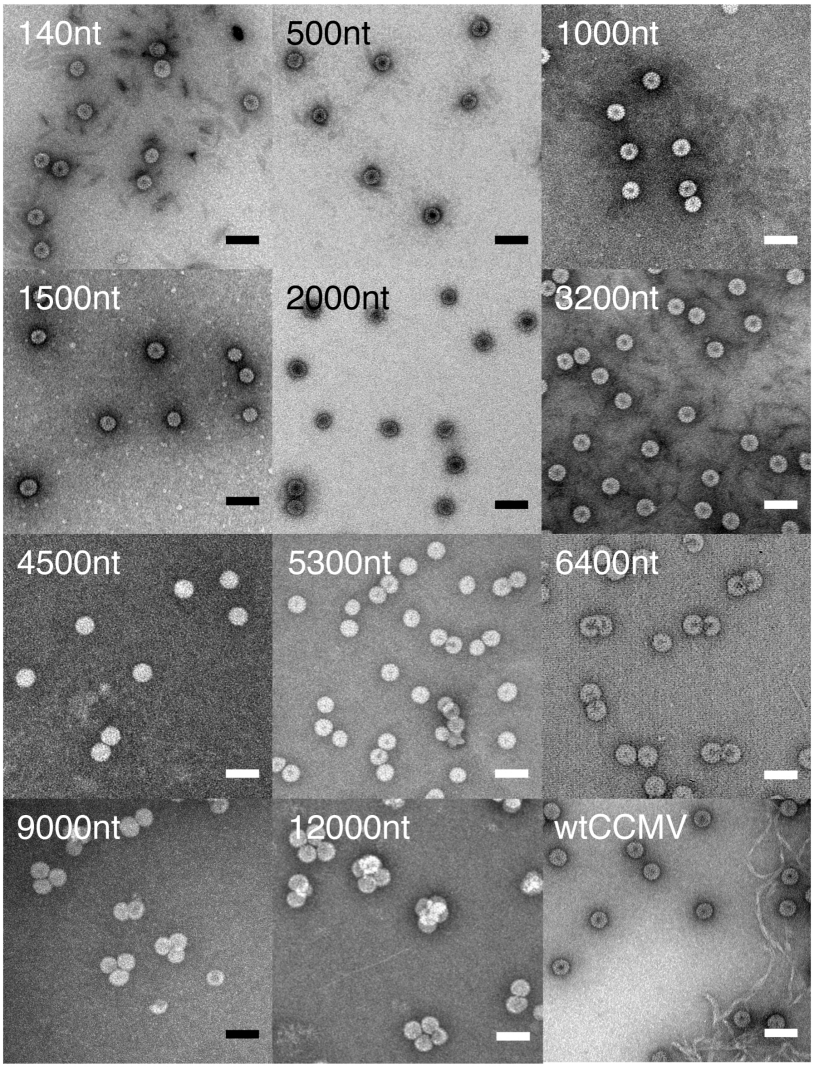

Figure 1-3. VLP products as a function of RNA length. Negative-stain TEM images of the assembly products of varying lengths of ssRNA mixed with CCMV CP at the magic ratio (6:1 wt:wt) in RNA assembly buffer. The length of the RNA in each assembly is labeled in the upper left corner of each image. The lower right-most image shows wt CCMV capsids. Grids were stained with uranyl acetate. Scale bar shows 50 nm.

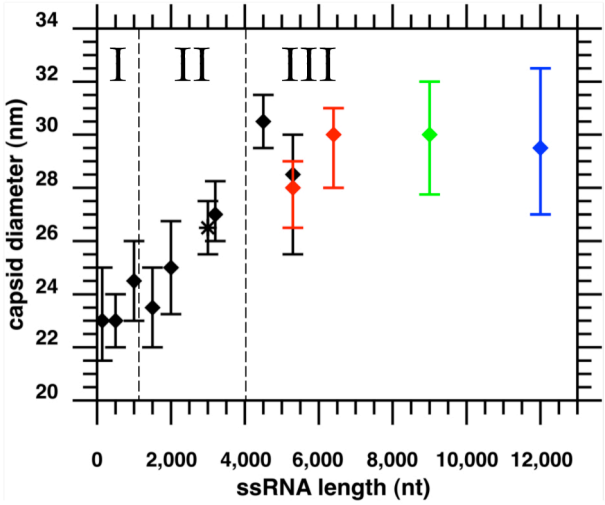

Figure 1-4A. Average capsid diameter as a function of RNA length. Plot of the average capsid diameter as a function of the length of ssRNA packaged, with each assembly carried out at the magic ratio in RAB. Each average contains contributions only from the capsids associated with the predominant multiplet (i.e. singlet, doublet, triplet or quadruplet) for the corresponding RNA length (except in the case of 5300nt, where equal amounts of singlets and doublets were observed). Singlets are plotted as black diamonds, doublets as red, triplets as green and quadruplets as blue. The plot is divided along the abscissa into three regions: (I) denotes the region between 140nt and ~1000nt where we find predominantly singlet capsids that contain multiple ssRNA molecules; (II) indicates the region between ~1500nt and ~4000nt where each singlet capsid contains only one RNA molecule – the point denoted with an asterisk corresponds to the wt CCMV virion; (III) is the region above ~4000nt corresponding to the appearance of multiplet capsids sharing an RNA molecule.

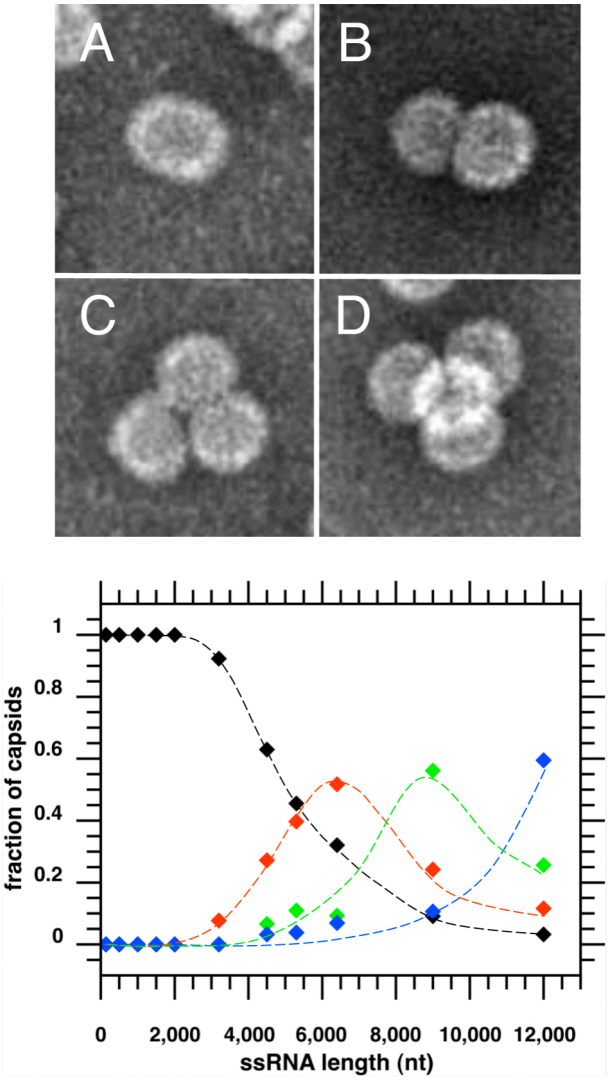

Figure 1-4B. Fraction of multiplets as a function of RNA length. The frequency of multiplet capsids – defined as the fraction of capsids present in the TEM as doublets (red), triplets (green) and quadruplets (blue) – is seen to increase with ssRNA length after 3200nt. Singlets (black) predominate for smaller lengths. Hand-drawn best-fit lines have been added to aid the eye in following the appearance and disappearance of the different multiplet populations as RNA length increases. The panel of negative-stain TEM images above shows typical structures observed for singlets (A), doublets (B), triplets (C) and quadruplets (D).

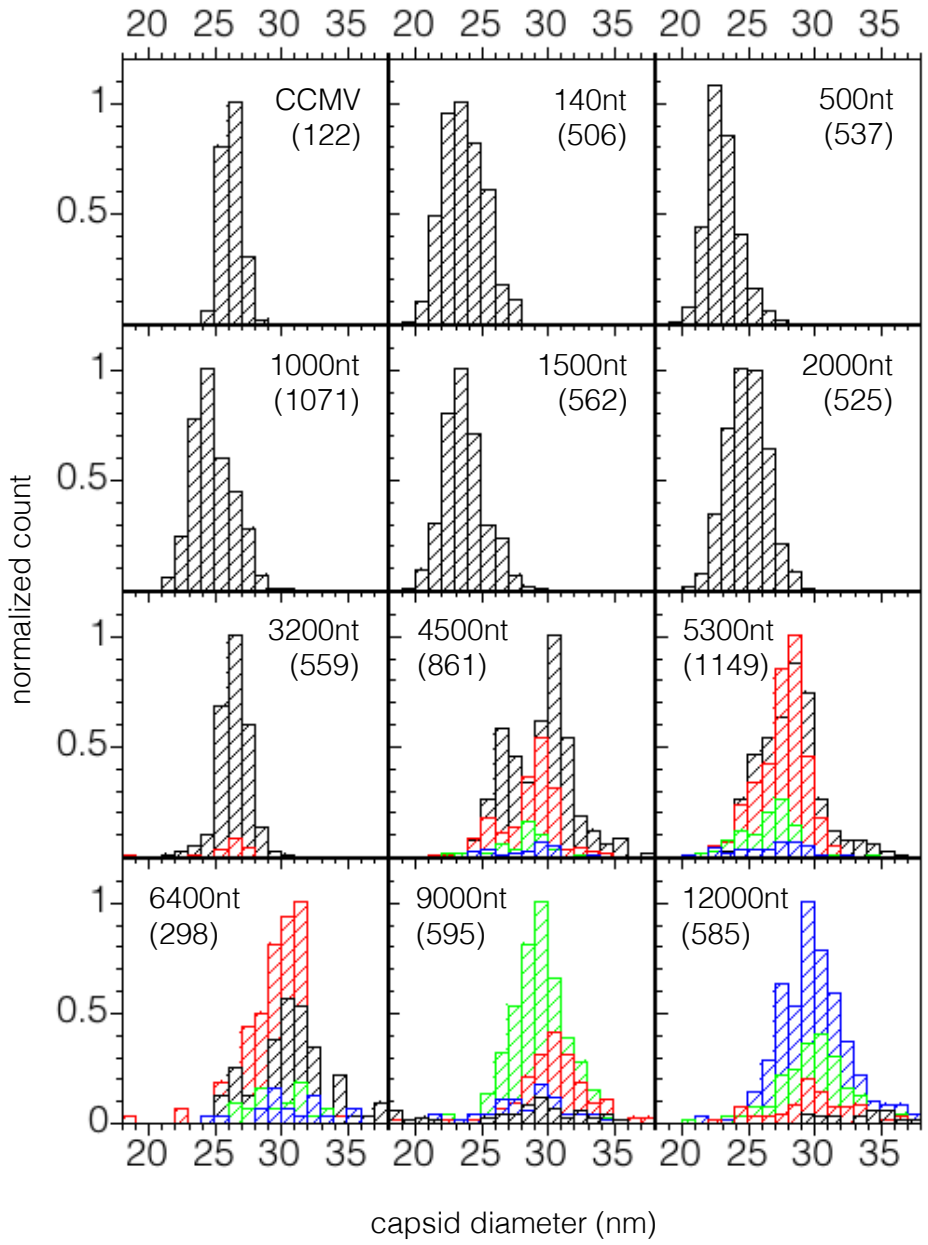

Figure 1-5. Distribution of capsid diameters appearing in singlets and multiplets, for each length of RNA packaged. Contributions from singlets, doublets, triplets and quadruplets are shown in black, red, green and blue, respectively. The distribution at the upper left is for wt CCMV. In

each distribution the counts have been normalized to that of the most abundant size. Total numbers of particles measured are given in parentheses.

SUPPLEMENTARY FIGURES

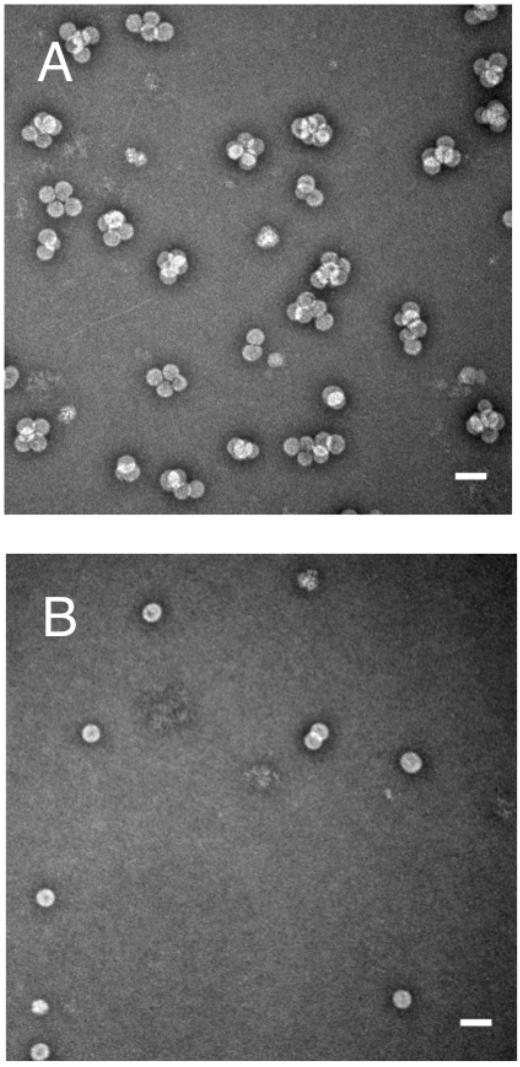

Figure 1-S1. Effect of RNase on Multiplets. (A) shows a typical TEM image of 12000nt RNA assembled with CCMV CP. (B) shows the same assembly mixture after exposure to RNase A in Virus Buffer for 1 hr. Note the reduction in the number of multiplets as well as an overall reduction in the total number of capsids present.

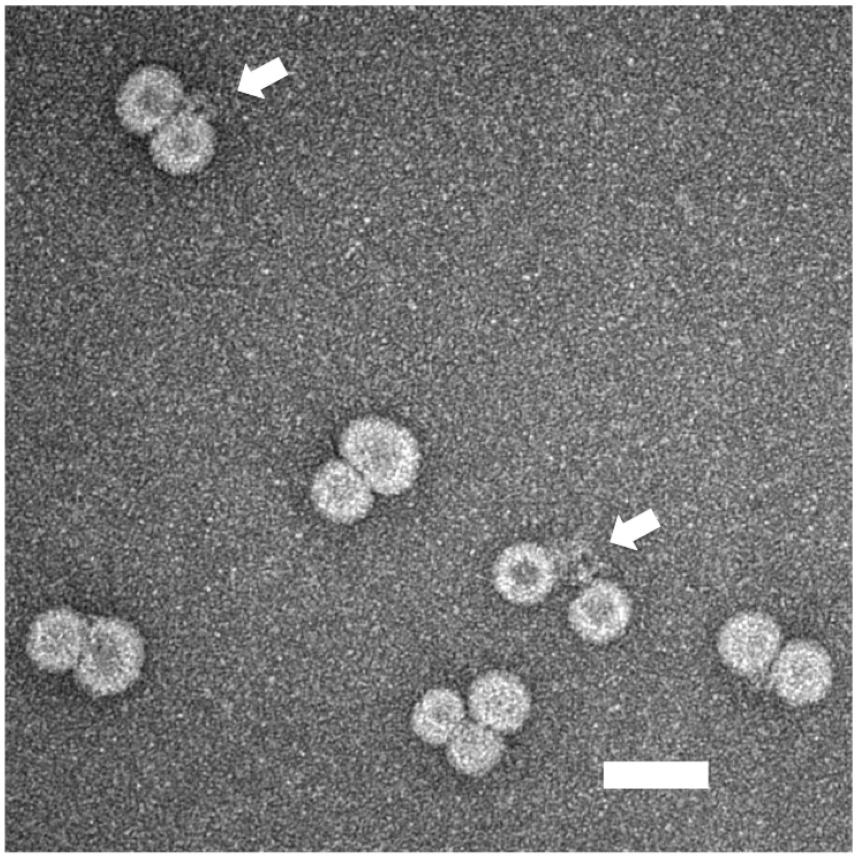

Figure 1-S2. Shared RNA between capsids in a multiplet. The arrows point to RNA seen linking two capsids.

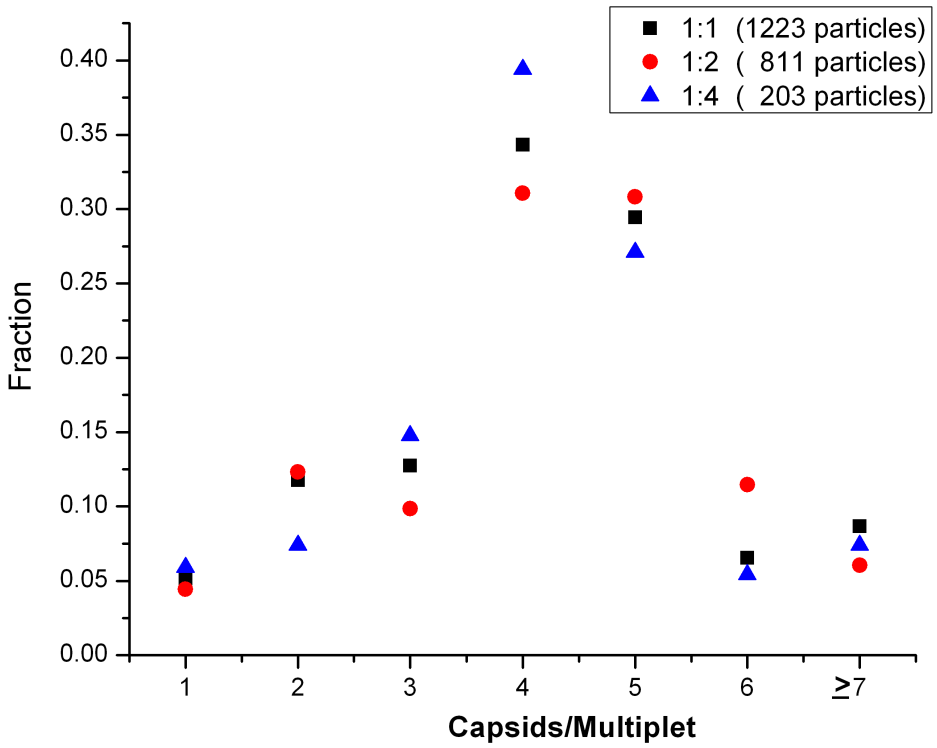

Figure 1-S3. Absence of effect of 2-fold and 4-fold dilution on frequency of multiplets. The different symbols show the feffect of sample dilution on fractions of mulitplets of each type. The sample represented by the black squares has been diluted by a factor of 2 (red circles) and a factor of 4 (blue triangles). There is no systematic effect on the populations of multiplets with dilution.

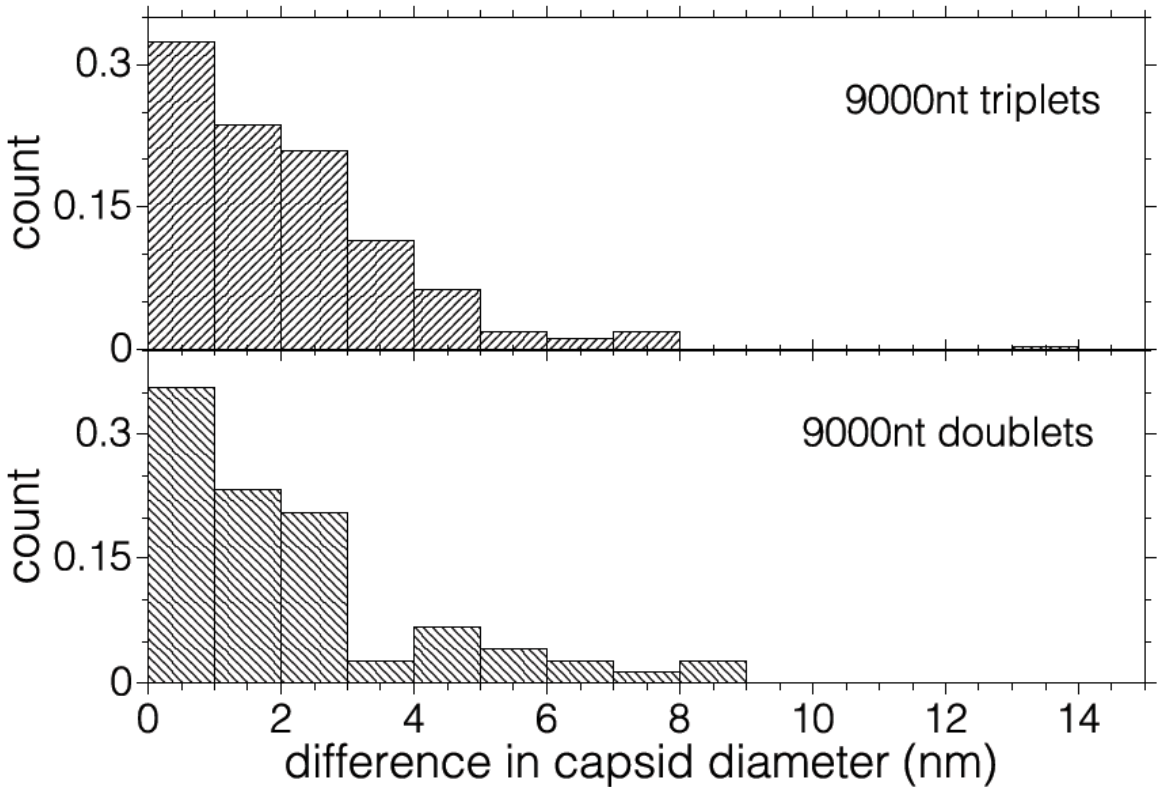

Figure S4. Size asymmetries of capsids within doublets and triplets formed from assembly of 9000nt RNA and CCMV capsid protein.

REFERENCES

1. Adolph, K. W., and P. J. G. Butler. 1977. Studies on the Assembly of a Spherical Plant Virus III. Reassembly of Infectious Virions Under Mild Conditions. J. Mol. Biol. 109: 345 – 357.

2. Allison, R. F., M. Janda, and P. Ahlquist. 1988. Infectious in vitro transcripts from cowpea chlorotic mottle virus cDNA clones and exchange of individual RNA components with brome mosaic virus. 1988. J. Virol. 62:3581-3588.

3. Annamalai, P., and A.L.N. Rao. 2005. Dispensibility of 3' tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. Virology. 332: 650-658.

4. Annamalai, P., and A. L. N. Rao. 2006. Packaging of Brome Mosaic Virus Subgenomic RNA Is Functionally Coupled to Replication-Dependent Transcription and Translation of Coat Protein. J. Virol. 80:10096 – 10108.

5. Annamalai, P., F. Rofail, D. A. DeMason, and A. L. N. Rao. 2008. Replication-Coupled Packaging Mechanism in Positive-Strand RNA Viruses: Synchronized Coexpression of Functional Multigenome RNA Components of an Animal and a Plant Virus in Nicotiana benthamiana Cells by Agroinfiltration. J. Virol. 82:1484-1495

6. Bamunusinghe, D., J-K. Seo, and A. L. N. Rao. 2011. Subcellular Localization and Rearrangement of Endoplasmic Reticulum by Brome Mosaic Virus Capsid Protein. J. Virology. 85:2953-2963.

7. Bancroft, J. B. 1970. The Self-Assembly of Spherical Plant Viruses. Adv. Virus Res. 16: 99-134.

8. Bancroft, J. B., and E. Hiebert. 1967. Formation of an Infectious Nucleoprotein from Protein and Nucleic Acid Isolated from a Small Spherical Virus. Virology 32: 354-356.

9. Bancroft, J. B., E. Hiebert, and C. E. Bracker. 1969. The effects of various polyanions on shell formation of some spherical viruses. Virology 39: 924-930.

10. Basnak, G., V. L. Morton, Ó. Rolfsson, N. J. Stonehouse, A. E. Ashcroft, and P. G. Stockley. 2010. Viral Genomic Single-Stranded RNA Directs the Pathway Towards a T = 3 Capsid. J. Mol. Biol. 395:924-936.

11. Briddon, R. W., and P. G. Markham. 1995. Family Geminiviridae. In: F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo and M. D. Summers, Eds, Virus Taxonomy Arch. Virol., Springer-Verlag, Wien/New York pp. 158–165.

12. Cadena-Nava, R. D., Y. Hu, R. F. Garmann, B. Ng, A. N. Zelikin, C. M. Knobler, and W. M. Gelbart. Exploiting Fluorescent Polymers to Probe the Self-Assembly of Virus-Like Particles. 2011. J. Phys. Chem. B 115: 2386-2391.

13. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol. 27: 1-24.

14. Choi, Y. G. and A. L. N. Rao. 2000. Molecular Studies on Bromovirus Capsid Protein VII. Selective Packaging of BMV RNA4 by Specific N-Terminal Arginine Residues. Virology 275: 207-217.

15. Choi, Y. G. , and A. L. N. Rao. 2000. Packaging of tobacco mosaic virus subgenomic RNAs by brome mosaic virus coat protein exhibits RNA controlled polymorphism. Virology 275: 249-257.

16. deHaseth, P.L., Lohman, T.M. and Record, M.T. Jr. 1977. Nonspecific interactions of lac repressor with DNA: An association reaction driven by counterion release. Biochem. 16: 4783-4790.

17. Dreher, T. W., A. L. N. Rao, and T. C. Hall.1989. Replication *in vivo* of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. 206: 425–438.

18. Dzianott, A. and J. Bujarski. 1991. The nucleotide sequence and genome organization of the RNA-1 segment in two bromoviruses: broad bean mottle virus and cowpea chlorotic mottle virus. Virology 185: 553-562.

19. Elrad, O. M., and M. F. Hagan. 2010. Encapsulation of a polymer by an icosahedral virus. Phys. Biol. 7: 045003 (2010).

20. Fox, J. M., G. Wang, J. A. Speir, N. H. Olson, J. E. Johnson, T. S. Baker, and M. J. Young. 1998. Comparison of the Native CCMV Virion with *in Vitro* Assembled Virions by Cryoelectron Microscopy and Image Reconstruction. Virology 244: 212-218.

21. Frischmuth, T., M. Ringel, and C. Kocher. 2001. The size of encapsidated single-stranded DNA determines the multiplicity of *African cassava mosaic virus* particles. J. Gen Virol. 82: 673-676.

22. Hiebert, E., J. B. Bancroft, and C. F. Bracker. 1968. The Assembly *in Vitro* of Some Small Spherical Viruses, and Other Nucleoproteins. Virology 34: 492-508.

23. Hu, Y., R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart. 2008. Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size. Biophys. J. 94: 1428- 1436.

24. Johnson, J. M., D. A. Willits, M. J. Young, and A. Zlotnick. 2004. Interaction with Capsid Protein Alters RNA Structure and the Pathway for *In Vitro* Assembly of Cowpea Chlorotic Mottle Virus. J. Mol. Biol. 335: 455-464.

25. Johnson, M. W., G. W. Wagner, and J. B. Bancroft. 1973. A Titrimetric and Electrophoretic Study of Cowpea Chlorotic Mottle Virus and its Protein. J. Gen. Virol. 19:263-273.

26. Jung, B., A. L. N. Rao, and B. Anvari. 2011. Optical Nano-Constructs Composed of Genome-Depleted *Brome Mosaic Virus* Doped with a Near Infrared Chromophore for Potential Biomedical Applications.ACS Nano. 5:1243-1252.

27. Kobayashi, A., and Y. Ehara. 1995. *In vitro* Encapsidation of Cucumber Mosaic Virus RNA Species. Ann. Phytopathol. Soc. Jpn. 61: 99-102.

28. Kroll, M. A., N. H. Olson, J. Tate, J. E. Johnson, T. S. Baker, P. Ahlquist. RNA-controlled polymorphism in the *in vivo* assembly of a 180-subunit and 120-subunit virions from a single capsid protein. 1999. Proc. Natl. Acad. Sci. USA. 96: 13650-13655.

29. Lavelle, L., M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, and J. Ruiz-Garcia. 2009. Phase Diagram of Self-Assembled Viral Capsid Protein Polymorphs. J. Phys. Chem. B 113:3813-3820.

30. Lustig, S., A. C. Jackson, C. S. Hahn, D. E. Griffin, E. G. Strauss, and J. H. Strauss. 1988. Molecular basis of Sindbis virus neurovirulence in mice. J. Virol. 62: 2329-2336.

31. Mascotti, D. P. and T. M. Lohman. 1990. Thermodynamic Extent of Counterion Release Upon Binding Oligolysines to Single Stranded Nucleic Acids. Proc. Natl. Acad. Sci. USA 87: 3142- 3146.

32. Nugent, C. I., K. L. Johnson, P. Sarnow and K. Kirkegaard. 1997. Functional Coupling between Replication and Packaging of Poliovirus Replicon RNA. J. Virology 73: 427-435.

33. Obenauer-Kutner, L. J., P. M. Ihnat, T.-Y. Yang, B. J. Dovey-Hartman, A. Balu, C. Cullen. R. W. Bordens, and M. J. Grace. 2002. The Use of Field Emission Scanning Electron Microscopy to Assess Recombinant Adenovirus Stability. Hum. Gene Ther. 13: 1687-1696.

34. Pfeifer, P., M. Herzog, and L. Hirth. 1976. Stabilization of Brome Mosaic Virus. Phil. Trans. R. Soc. Lond. B 276:99-107.

35. Porterfield, J. Z., M. S. Dhason, D. D. Loeb, M. Nassal, S. J. Stray, and A. Zlotnick. 2010. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. J. Virol. 84: 7174-7184.

36. Prinsen, P. , P. van der Schoot, W. M. Gelbart, and C. M. Knobler. 2010. Shell Structures of Virus Capsid Proteins. J. Phys. Chem. B 114: 5522-5533.

37. Qu, F., and T.J. Morris. Encapsidation of Turnip Crinkle Virus is Defined by a Specific Packaging Signal and RNA Size. 1997. J. Virol. 71: 1428-1435.

38. Rao, A. L. N. 2006. Genome Packaging by Spherical Plant RNA Viruses. Ann. Rev. Phytopathol. 44: 61-87.

39. Sambrook J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

40. Sikkema, F. D., M. Comellas-Aragonès , R. G. Fokkink, B. J. M. Verduin, J. J. L. M. Cornelissen, and R. J. M. Nolte. 2007. Monodisperse polymer–virus hybrid nanoparticles. Org. Biomol. Chem. 5: 54-57.

41. Sorger, P. K., P. G. Stockley, and S. C. Harrison. Structure and Assembly of Turnip Crinkle Virus II. Mechanism of Reassembly *in Vitro.* 1986. J. Mol. Biol. 191: 639-658.

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

43. Sun, J., C. DuFirt, M. C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. A. Rotello, C. C. Holzenberg, B. Kao, and B. Dragnea. 2007. Core-controlled polymorphism in virus-like particles. Proc. Natl. Acad. Sci. USA 104*:* 1354-1359.

44. Tang, J., K. A. Johnson, M. J. Dryden, M. J. Young., A. Zlotnick, and J. E. Johnson. 2006. The role of subunit hinges and molecular "switches" in the control of viral capsid polymorphism. J. Struct. Biol. 154: 59-67.

45. van der Graaf , M., R.M. Scheek, C. C. van der Linden and M. A. Hemminga. 1992. Conformation of a pentacosapeptide representing the RNA-binding N-terminus of cowpea chlorotic mottle virus coat protein in the presence of oligophosphates: a two-dimensional proton

nuclear magnetic resonance and distance geometry study. Biochem. 3: 9177-9182.

46. Venter, P. A., N. K. Krishna, and A. Schneemann. 2005. Capsid Protein Synthesis from Replicating RNA Directs Specific Packaging of the Genome of a Multipartite, Positive-Strand RNA Virus. J. Virol. 79:6239-6248.

47. Verduin, B. J. M., and J. B. Bancroft. 1969. The Infectivity of Tobacco Mosaic Virus RNA in Coat Proteins from Spherical Viruses. Virology 37: 501-506.

48. Yoffe, A. M., P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul. 2008. Predicting the sizes of large RNA molecules. Proc. Natl. Acad. Sci. USA 105: 16153-16158.

49. Zandi, R., and P. van der Schoot. 2009. Size Regulation of ss-RNA Viruses. Biophys. J. 96: 9-20.

50. Zhang, W., N. H. Olson, T. S. Baker, L. Faulkner, M. Agbandje-McKenna, M. I. Boulton, J. W. Davies, and R. McKenna. 2001. Structure of the Maize Streak Virus Geminate Particle. Virology 279: 471-477.

51. Zlotnick, A., R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young. 2000. Mechanism of Capsid Assembly for an Icosahedral Plant Virus. Virology 277:450-456.

CHAPTER 2

The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions

ABSTRACT The strength of attraction between capsid proteins (CPs) of Cowpea Chlorotic Mottle Virus (CCMV) is controlled by the solution pH. Additionally, the strength of attraction between CP and the single-stranded (ss) RNA viral genome is controlled by ionic strength. By exploiting these properties, we are able to control and monitor the *in vitro* co-assembly of CCMV CP and ssRNA as a function of the strength of CP-CP and CP-RNA attraction. Using the techniques of velocity sedimentation and electron microscopy, we find that the successful assembly of nuclease-resistant virus-like particles (VLPs) depends delicately on the strength of CP-CP attraction relative to CP-RNA attraction. If the attractions are too weak, the capsid cannot form; if they are too strong, the assembly suffers from kinetic traps. Separating the process into two steps – by first turning on CP-RNA attraction, and then turning on CP-CP attraction – allows for the assembly of well-formed VLPs under a wide range of attraction strengths. These observations establish a protocol for the efficient *in vitro* assembly of CCMV VLPs, and suggest potential strategies that the virus may employ *in vivo*.

INTRODUCTION

The primary interactions within the virus particle are those between neighboring protein molecules forming the capsid (CP-CP) and those between capsid protein and nucleic acid (CP-RNA). The present work focuses on the interplay between these key interactions and their effect on the *in vitro* assembly pathway of the plant RNA virus, Cowpea Chlorotic Mottle Virus (CCMV).

The CCMV virion consists of ~3,000 nts of single-stranded RNA packaged within a 28-nm-diameter capsid made up of 180 chemically identical copies of a 190-residue capsid protein. We chose to study CCMV assembly for three reasons. First, CCMV is among the most robust viral self-assembly systems – it was the first icosahedral virus to be reconstituted *in vitro* from purified components[1], and CCMV CP has since been shown to package a wide range of heterologous RNAs[2], in addition to being capable of assembling into empty capsids[3,4]. Second, high-resolution reconstructions of the virion obtained by X-ray crystallography and cryo-electron microscopy (cryo-EM)[5], in conjunction with computational simulations[6], have yielded great insight into the specific molecular interactions that govern the structure, stability, and dynamic properties of the virion. Third, and most importantly, the strength of the CP-CP and CP-RNA interactions can be controlled by adjusting the pH and ionic strength (I), respectively[7,8]. This last point, which we exploit heavily in the present work, is described in detail in the following sections.

While much is known about the structure and intermolecular forces within the CCMV virion, the pathway of assembly remains poorly understood. In an influential paper by Zlotnick and coworkers[9], empty capsids of CCMV CP were shown to assemble, under acidic conditions and in the absence of RNA, by first generating a pentamer of CP dimers (the nucleation phase) followed by the stepwise association of additional CP dimers (the growth phase) until capsid completion. To provide insight into the assembly pathway of RNA-containing CCMV virus-like particles (VLPs) Zlotnick et al. carried out further studies, involving RNA and a limiting amount of CP under physiological (neutral) pH conditions, in which they demonstrated that CP initially binds RNA with low cooperativity due to the relatively weak lateral attraction between CPs[10].

Expanding on this work, we previously investigated the co-assembly of an excess of CCMV CP and ssRNAs of varying lengths[2,11]. We reported evidence that the assembly of CCMV CP around ssRNA followed a two-step, pH-dependent, mechanism in which CP reversibly binds RNA at neutral pH to form a pre-capsid complex that upon acidification undergoes an irreversible reorganization into the final icosahedral capsid[11].

In the present work, we study the pathway of assembly of CCMV CP and BMV (Brome Mosaic Virus) RNA 1, a 3,200-nt heterologous RNA shown to be efficiently packaged by CCMV CP[2,11]. We choose to work with heterologous RNA in order to focus on generic features of viral assembly and avoid any specific interactions that may exist between CCMV CP and CCMV RNA. We monitor assembly by a combination of EM and velocity sedimentation. By controlling the relative strength of the CP-CP and CP-RNA interactions through the pH and ionic strength of the assembly buffer, we are able to characterize the structure of disordered intermediates and kinetically trapped assembly products, as well as elucidate key features of the final capsid reorganization.

RESULTS AND DISCUSSION

CP-CP interactions

Knowledge of CP-CP interactions in CCMV has come from decades of biophysical theory and experiment[8,12,13], and more recently from a high-resolution structural determination of the CCMV virion[5] and subsequent computational studies[6].

Disassembled CP exists in solution as a dimer[8] stabilized by a strong non-covalent clamp between the amino- and carboxy-terminal arms of each CP monomer[5]. The dimer serves as the fundamental assembling unit[8]. Lateral interactions between dimers within the capsd arise from a

combination of hydrophobic, electrostatic, and divalent metal-mediated forces between amino acid residues that line the dimer-dimer interfaces. The average strength of attraction between CP dimers has been shown to be pH dependent, decreasing sharply with increasing pH[13]. This is primarily due to electrostatic repulsions that are generated by the deprotonation of adjacent acidic residues at neutral pH, in the absence of divalent metal ions[6]. This effective decrease in lateral attraction is illustrated by the decrease in thermal stability of wild-type CCMV at elevated pH values (Fig. 2-1). Here we see that the virion is stable to 75º C at pH 4.5 and I = 0.1M, but a sharp drop of 30º C in melting temperature occurs upon raising the pH above 6.

CP-RNA interaction

CP interacts with RNA primarily through a disordered, highly-basic, N-terminal arginine-rich motif (ARM). Studies exploiting truncated CP lacking this highly basic region have shown that RNA packaging requires the ARM[14,15]. While sequence-specific interactions have been shown to play a role in the binding of CP to RNA – both for CCMV[16] as well as other icosahedral plant viruses[17-20] – the predominant interaction in the case of CCMV is non-specific electrostatic attraction. Residues 1-26 contribute +10 charges that interact strongly with the negatively-charged phosphate backbone of the RNA. Physical arguments based on the entropy gain associated with mobile counterion release upon polycation/nucleic acid binding estimate the free energy benefit to be roughly kT per charge. Since there are +10 charges on the CP N-termini, we expect binding energies on the order of 10 kT per CP. Binding assays – carried out between BMV, which contains 9 basic residues – confirm this estimate[21].

The electrostatic nature of CP-RNA interactions suggests that the binding should be turned off at sufficiently high salt concentrations due to a reduction in the entropic benefit of

counterion release as well as charge screening – indeed, increasing ionic strength is a key step in the protocol for disassembling CCMV for the purpose of isolating CP[22]. To assay how ionic strength controls assembly, CP and BMV RNA 1 were separately labeled with different fluorescent probes and subjected to sucrose gradient velocity sedimentation to test for binding (Fig. 2-2). We see that CP and RNA co-sediment as a high-density complex in low-salt (0.1M) buffer (Fig. 2-2d), but migrate as separate species at high (1M) ionic strength (Fig. 2-2c) – confirming that CP-RNA interactions can be turned off by high ionic strength. On the other hand, because the positive charges on the ARM derive from arginine and lysine residues that do not titrate over the pH ranges explored in this work, the energy of binding CP to RNA is effectively pH-independent.

General assembly procedure

The present work focuses on monitoring the assembly pathway of CCMV CP and BMV RNA1 as a function of the relative strength of CP-CP and CP-RNA interactions, as well as on the order in which these interactions are "turned on". Experimentally, this is accomplished by directing mixtures of CP and RNA through the composition space defined by the pH and ionic strength of the buffer solution. To keep track of the location in this space we label each buffer by its pH and ionic strength, in parenthesis, according to the following format, "*(pH,I)-buffer"*. For example, the canonical assembly buffer described by Bancroft (50 mM NaCl, 50 mM Tris-HCl [pH 7.2], 10 mM KCl, 5 mM $MgCl_2$) will be referred to as *(7.2,0.1)-buffer* – because it has a pH of 7.2 and an ionic strength of 0.1M. Additionally, any dialysis step that goes from one point in composition space to another will be denoted by a lower case *"d"*. For example, dialyzing a

solution of CP and RNA in Bancroft's assembly buffer against the canonical storage buffer (50 mM sodium acetate [pH 4.5], 8 mM magnesium acetate) will be written *(7.2,0.1)d(4.5,0.1)*.

The general assembly protocol consists of direct mixing of CP and RNA at neutral pH and high ionic strength, *(7.2,1.0)-buffer*, followed by subsequent dialysis and equilibration steps. *(7.2,1.0)-buffer* is chosen as the initial buffer for all assembly reactions because it allows for mixing of the reactants in the absence of attractive interactions that drive assembly; more explicitly, both CP-CP and CP-RNA interactions are turned off. After mixing, we are able to selectively turn on CP-RNA and/or CP-CP attraction by dialysis against buffers of appropriate pH and ionic strength. In the following sections, it will be useful to refer to assembly reactions as either "one-step" or "two-step" depending on the number of dialysis steps that follow the initial mixing of CP and RNA.

One-step Assemblies

To test the extent of assembly under conditions of weak CP-CP attraction and strong CP-RNA attraction, we dialyzed a solution of fluorescently labeled CP and RNA in *(7.2,1.0)-buffer* against *(7.2,0.1)-buffer* and analyzed the products by a combination of velocity sedimentation and transmission electron microscopy (Fig. 2-3).

The fluorescence intensity profiles of fractions collected from sucrose density gradient assays show that the reaction products co-sediment as a broad band of complexed CP and RNA with a density similar to that of the CCMV virion (Fig. 2-3a, black curve). Analytical ultracentrifugation measurements confirm this result (data not shown). Unlike CCMV virions produced in a natural infection, or in a two-step assembly (see following section), these nucleoprotein complexes are not stable against RNase – the sedimentation profile after RNase

treatment shows the complexes were completely digested into low-molecular weight species consistent with free CP dimers and short fragments of RNA (Fig. 2-3a, grey curve).

Direct visualization of the reaction products by cryo-EM clarifies the basis for their sedimentation behavior and susceptibility to RNase digestion. As seen in Fig. 2-3b, the complexes imaged in vitrified *(7.2,0.1)-buffer* appear as disordered clumps with diameters (~30 nm) similar to that of the CCMV virion, accounting for their similar sedimentation profiles. However, unlike the native virion, the CPs in these complexes lack the regular structure that is known to protect CCMV against digestion by RNase. Closer investigation of their structure (Fig. 2-4a) shows that a fraction of the population exhibits regions of curvature and topography that suggest partial capsid formation. Whether these small patches of ordered CP are long-lived or transient structures could not be determined, though long-lived partial capsid shells seem unlikely given the reversible nature of CP-RNA binding at neutral pH[11].

Conditions of neutral pH and low ionic strength are therefore insufficient for nucleocapsid formation. While the strength of CP-RNA attraction drives all CP in solution into nucleoprotein complexes, CP-CP attractions are too weak to produce closed capsids. The attractions appear to be sufficient for the nucleation of transient partial capsid structures, but these ordered embryos are incapable of developing into closed shells. These observations are consistent with the initial stages of a generalized assembly pathway first described by McPherson[23] and later by Devkota et al.[24], in which CP initially binds the nucleic acid, forming a disordered complex that later rearranges to form the ordered capsid. This scenario, named the "en-masse" pathway by Elrad and Hagan[25], has been observed in coarse-grained simulations of the co-assembly of idealized capsid protein oligomers around linear[25] and branched[26] polymers

representing ssRNA under conditions when CP-RNA attraction dominates over CP-CP attraction.

The more traditional (and also less direct) imaging technique of negative-stain electron microscopy – in which sample is adsorbed onto a hydrophilic carbon-coated substrate, dried, and then treated with an electron-dense stain solution before visualization – yields higher contrast images (Figs. 2-3c and 2-4c) but suffers from a susceptibility to artifacts incurred during sample preparation. The most commonly used stain in EM, aqueous 1-2% uranyl acetate (UA), is acidic (pH ≈ 4.5) and great care must therefore be taken when using it to stain pH-dependent samples. Complete removal of excess stain before storing the grid is essential to reducing artifacts associated with its low pH. Comparison of CP-RNA complexes stained briefly with UA (Fig. 2-3c) against those from the same sample subjected to prolonged exposure due to incomplete stain removal (Fig. 2-3d) shows one dramatic and previously unreported artifact – *prolonged staining with uranyl acetate drives CCMV capsid formation*.

Following the failure of the one-step neutral-pH assembly procedure to generate well-formed VLPs, we analyzed the products of one-step assembly under acidic conditions – i.e., conditions of strong CP-CP attraction – by dialyzing a mixture of CP and RNA in *(7.2,1.0)-buffer* against *(4.5,0.1)-buffer*. Sucrose density gradients showed evidence of high-molecular-weight aggregates of CP and RNA (Fig. 2-5a) – all of the CP and RNA was found as a pellet at the bottom of the gradient. Direct visualization by cryo-EM showed micron-sized aggregates of capsid-like structures (Fig. 2-5b). The increased contrast of negative-stain micrographs revealed that the capsids comprising the aggregate were often irregular in structure (i.e. non-spherical; Fig. 2-5c). Because the sample had been acidified before staining, potential artifacts arising from the low pH of uranyl acetate were not a concern. Treatment of these aggregates with RNase

resulted in the generation of CP-RNA complexes with a sedimentation profile similar to that of closed VLPs (Fig. 2-5a), suggesting that the aggregated capsids are held together by RNA.

Previous studies[2] have shown that RNA can be shared between two or more CCMV capsids, the RNA serving as a tether between particles and forming small bunches of VLPs (multiplet capsids). Addition of RNase to these multiplet structures results in the digestion of the RNA tethers and liberation of the constituent capsids. The aggregated structures resulting from the low-pH assembly reaction appear to be an extreme case of such multiplet formation, with the majority of capsids sharing portions of RNA with at least two other capsids. Multiplet structures are not unique to CCMV and have been reported to occur in the *in vitro* assembly of Turnip Crinkle Virus[20] under "conditions that favor assembly so strongly that they 'lock in' mistakes", as well as for bacteriophage fr[27] when the RNA packaged was significantly longer than the viral genome. While the mechanism of multiplet formation has not been established experimentally, coarse-grained simulations[25,28] suggest that it is a form of kinetic trapping resulting from strong CP-CP attraction – consistent with our experimental observations as well as with other kinetic–thermodynamic models[29,30].

Two-step assemblies

Individually, both the neutral-pH and low-pH one-step assembly reactions fail to generate well-formed spherical VLPs. However, a two-step assembly procedure consisting of an initial equilibration at neutral pH and subsequent acidification results in nuclease-resistant particles with closed CP capsids. This two-step assembly was achieved by mixing CP and RNA in *(7.2,1.0)-buffer*, dialysis against *(7.2,0.1)-buffer* overnight, and a second dialysis step of 6 h against *(4.5,0.1)-buffer*.

As seen by the sucrose gradient profile (Fig. 2-6a), the products of this two-step pH-dependent assembly procedure have an average size and density comparable to those of the complexes before acidification, with the addition of a slight high-density shoulder. However, these complexes are not converted to low-molecular-weight species upon treatment with RNase – as seen in Fig. 2-6a, the CP sedimentation profile is unchanged. And both negative-stain- and cryo-EM show that acidification leads to the formation of spherical capsids that are indistinguishable from wild-type virions (Figs. 2-6b and c). The VLPs that form upon acidification remain after dialysis back to neutral pH (Fig. 2-7c) for at least 1 week at 4ºC (longer times were not tested).

As seen in Figs. 2-7a and b, the two-step procedure directs the assembly through both stages of the en-masse pathway[23-25], disordered complex formation followed by capsid organization. The disordered nucleoprotein complexes act as ideal pre-capsid intermediates. The initial neutral-pH equilibration period allows the system to anneal under conditions of weak CP-CP attraction into particles consisting of a single RNA molecule decorated by an excess of CP. Upon lowering the pH, the lateral interactions between nearby CP dimers bound in the pre-capsid complex are increased and the closed capsid is formed. Comparing the products of the two-step assembly (Fig. 2-6) with those of the one-step neutral-pH assembly (Fig. 2-3), we see that acidification is necessary for generating CP-CP attractions strong enough to drive the formation of complete capsids. Further comparison against the products of the one-step acidic assembly that are shown in Fig. 2-5 demonstrates that the neutral-pH annealing step is crucial for avoiding kinetic traps associated with strong CP-CP attraction.

To test if capsid formation can be triggered by intermediate pH values ($4.5 < pH < 7.2$), we prepared samples of CP-RNA complexes at pH 7.2 and, in parallel, dialyzed each against

assembly buffers with successively lower pH while keeping the ionic strength constant at 0.1M. Cryo- and negative-stain-EMs (Fig. 2-8) of the products of each two-step, intermediate-pH reaction show that closed spherical capsids are absent above pH 6.5. At pH 6.5, a mixture of VLPs and disordered complexes is present, whereas below pH 6.0 the only structures observed are closed spherical VLPs.

CCMV undergoes numerous changes between pH 6 and 6.5. As shown in Fig. 2-1, there exists a sharp decrease in thermal stability of CCMV between pH 6 and 6.5. Additionally, the well-known swelling transition of CCMV is triggered by raising the pH above 6.5 in the absence of divalent metal ions[31]. Thus, the observation that the transition between disordered nucleoprotein complex and VLP occurs between pH 6.5 and 6 suggests that the lateral interactions responsible for driving capsid assembly are the same as those that are involved in maintaining capsid stability and triggering swelling. Computational simulations[6] have linked the swelling transition with the deprotonation of specific acidic residues that interact at the dimer-dimer interfaces within the capsid. We propose that the same residues play a key role in the assembly of VLPs.

It still remained to be shown whether a one-step reaction involving moderate pH is capable of generating well-formed nucleocapsids. To test this we directly dialyzed a mixture of CP and RNA in *(7.2,1.0)-buffer* against *(6.0,0.1)-buffer* (Fig. 2-9). The sedimentation profile generated was significantly broader and shifted toward higher density than that of the two-step reaction of Fig. 2-6. RNase treatment sharpened the band and, as in the case of the one-step low-pH assembly reaction, gave rise to a peak position consistent with that of CCMV. Both cryo- and negative-stain-EM show that the reaction products are a mixture of well-formed single capsids

and doublet capsid structures. We interpret the doublet particles as kinetically-trapped multiplet structures similar to those produced in the one-step low-pH assembly reaction.

SUMMARY AND CONCLUSIONS

The successful self-assembly of CCMV CP and ssRNA into VLPs depends delicately on the strength of CP-CP attraction relative to CP-RNA attraction. If too weak, closed capsids do not form (Fig. 2-3); if too strong, the assembly suffers from kinetic traps (Fig. 2-5). We can control the strength of CP-CP and CP-RNA attractions *in vitro* by adjusting the pH and ionic strength, respectively: the strength of CP-CP attraction varies inversely with pH (Fig. 2-1) and CP-RNA attraction varies inversely with ionic strength (Fig. 2-2). We have determined that CP-CP attraction is too weak at neutral pH to drive VLP formation and that the pH must be decreased below 6 in order to form closed RNase-resistant nucleocapsid structures (Fig. 2-8). However, even under conditions that provide the weakest CP-CP attraction capable of generating closed nucleocapsids (pH = 6 and I = 0.1M), assembly results in kinetically trapped structures (Fig. 2-9).

Dividing the assembly procedure into two steps – by first lowering the ionic strength (turning on CP-RNA attraction), and later reducing the pH (turning on CP-CP attraction) – avoids the kinetic traps that are produced in the one-step reaction under conditions of moderate-to-strong CP-CP attraction. The equilibration step at neutral pH allows the system to anneal, under conditions of weak CP-CP interactions, into a homogeneous population of individual RNA molecules decorated by a high density of CP (Fig. 2-3). These pre-capsid intermediates resist kinetic trapping upon acidification and generate the best possible yields of spherical VLPs (Fig. 2-6). This two-step assembly procedure, similar to that used in packaging functionalized gold

nanoparticles by Brome Mosaic Virus CP[32], should help inform the development of new virus-based nanomaterials involving self-assembled structures of CCMV CP around anionic polymer cargos.

The observations that (i) the products of a one-step assembly reaction are structurally different from those of a two-step reaction carried out between identical initial and final conditions (Fig. 2-5c vs. 2-6c), and (ii) nucleocapsids that are incapable of assembling at neutral pH, once acidified, are metastable against dialysis back to neutral pH (Fig. 2-7c), demonstrate that the assembly and disassembly of RNA-containing CCMV VLPs are path-dependent processes that occur out of equilibrium. Unlike the assembly of pure CCMV CP[4], the assembled structures found in a mixture of CCMV CP and RNA cannot be predicted by knowledge of the pH and ionic strength alone – the buffer history is important.

The relevance of these findings *in vivo* has yet to be demonstrated. It is generally accepted that the cytoplasm of plant cells is maintained near neutral pH with ionic strength of approximately 0.1M. Our *in vitro* results show that these conditions are insufficient for nucleocapsid formation in the absence of cellular host factors. While it has been observed that significant structural rearrangement of subcellular membranes occurs during CCMV infection[33], no measurements of variations in the solution conditions (e.g. pH or other ionic concentrations) within these structures have been reported. It is therefore premature to implicate acidification as a maturation pathway for CCMV *in vivo*. Likewise, while the metastability of closed capsids at neutral pH is an important feature of CCMV *in vitro*, *in vivo* processes exploiting these metastable states have not been directly observed.

MATERIALS AND METHODS

Buffers

The following is a list of the assembly buffers used, their common name (if appropriate), and their composition: *(pH 7.2, ionic strength 1.0)-buffer* – Buffer B or Protein buffer or Dimer buffer[8] – 1 M NaCl, 20 mM Tris [pH 7.2], 1 mM EDTA, 1 mM DTT, 1 mM PMSF; *(7.2,0.1)-buffer* – Bancroft's assembly buffer[7] – 50 mM NaCl, 50 mM Tris-HCl [pH 7.2 or 7.0], 10 mM KCl, 5 mM $MgCl_2$; *(6.5,0.1)-buffer* – 50 mM NaCl, 50 mM Citric Acid-NaOH [pH 6.5 or 6], 10 mM KCl, 5 mM $MgCl_2$; *(4.5,0.1)-buffer* – Virus storage buffer, VSB[34] – 50 mM sodium acetate [pH 4.5], 8 mM magnesium acetate.

CCMV purification and labeling

CCMV was purified from infected California cowpea plants (Vigna unguiculata cv. Black Eye) as previously described by Bancroft[12]. Alexa Fluor-647 NHS ester (AF647-NHS; Molecular Probes, U.S.A.) was covalently linked to solvent-exposed lysines on the exterior surface of CCMV by direct mixing of 5 mg/mL CCMV and AF647-NHS at a mass ratio of 200:1, in 0.1 M HEPES (pH 7.2), 5 mM $MgCl_2$, for 2 h at 4ºC[35], followed by overnight dialysis against VSB. Unreacted dye was removed by collecting the capsids in a sucrose cushion, and resuspending them in VSB. The density of labeling (DOL), defined as the fraction of CPs labeled by AF647-NHS, was determined by UV-Vis spectroscopy according to the manufacturer.

Fluorescent labeling of CP and RNA

CCMV CP was isolated from fluorescently labeled CCMV with a DOL of ~0.05, as well as from unlabeled CCMV, as previously described by Annamalai et al.[22]. The CP concentration

and the degree of RNA contamination were measured by UV-Vis spectrophotometry; only CP solutions with 280/260 ratios greater than 1.5 (corresponding to <0.5% RNA contamination) were used in assembly reactions. SDS-PAGE and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry were used to make certain that the purified protein was not cleaved.

Fluorescently-labeled BMV RNA 1 was generated by *in-vitro* transcription of pT7B1 plasmid[36] with T7 RNA polymerase, rATP, rGTP, rCTP, rUTP, and *a fluorescently labeled rUTP-AF488 (ChromaTide® Alexa Fluor® 488-5-UTP;* Molecular Probes, U.S.A*.)*. A *rATP:rGTP:rCTP:rUTP:rUTP-AF488 molar ratio of 600:600:600:5.32:1 was used and resulted in a DOL of 0.5 rUTP-AF488s per RNA*. RNA was purified in TE buffer by washing the transcription products using a 100 kDa Amicon Filter device.

Measuring the apparent melting curve of CCMV by differential scanning fluorimetry

Procedure adapted from Hema et al.[37] 20 mL aliquots of WT CCMV were prepared at a final concentration of 0.2 mg/mL and 2.5X SYPRO orange fluorescent dye (Molecular Probes, U.S.A.) in 50 mM Phosphate Buffer at pH values of 5.5, 6.5, 7.5 and 8.0. An additional aliquot of identical concentrations of CCMV and dye was prepared by dilution in VSB. Differential scanning fluorimetry was performed in a 96-well plate 3200 Opticon2 (MJ Research) real-time PCR machine. All samples were heated from 25 to 95ºC, in 1ºC increments with a 15 second stabilization period at each temperature. During the stabilization period, the fluorescence intensity was measured against a blank containing the appropriate buffer and 2.5X SYPRO orange. Ex/Em wavelengths of 470/550 nm were used. Increased fluorescence emission derives from the binding of SYPRO orange to hydrophobic regions of the CP that are exposed upon

capsid melting. The apparent melting temperature is defined as the temperature at which the derivative of the fluorescence emission signal (-dI/dT) is maximal.

In vitro assembly reactions

Assembly reactions were carried out by direct mixing of CP and BMV RNA1 at neutral pH and high ionic strength, *(7.2,1.0)-buffer*, followed by subsequent dialysis and equilibration steps. All dialyses were equilibrated for at least 6h at 4°C.

Electron microscopy

Negative stain: 6 µL of assembly reaction was deposited on glow-discharged copper grids (400-mesh) that previously had been coated with Parlodion and carbon. After 1 min, the grids were blotted and stained with 6 µL of 2% uranyl acetate for 1 min followed by complete stain removal and storage in a desiccator overnight.

Cryo: 3 µL of assembly reaction was deposited on a Quantifoil holey carbon grid (200 mesh; R2/1) that had been previously glow-discharged. The grids were then blotted and flash-frozen by rapid plunging into liquid ethane cooled to liquid nitrogen temperature. Micrographs were acquired using an FEI Tecnai G2 TF20 microscope operated at an accelerating voltage of 200 kV. Images were recorded at 3-4 microns underfocus with a TIETZ F415MP 4k X 4k pixel CCD camera. Total beam exposure was maintained between 20-40 e/Å^2 for cryo-EM samples.

Velocity sedimentation by sucrose density gradient centrifugation

10-40% (w/v%) sucrose gradients were made by subjecting 3.5 mL of 25% sucrose in the desired buffer to repeated (3X) freeze-thaw cycles. The gradients were loaded with 0.2 mL of

sample and centrifuged for 3 h at 30,000 rpm at 4 ºC in a SW50.1 rotor (Beckman-Coulter) and manually fractionated by pipetting 0.2-mL aliquots from top to bottom. The 19 extracted fractions were loaded into a 96-well flat-black flat-bottom plate. The fluorescence of the labeled CP (AF647; ex/em 650/668 nm) and RNA (AF488; 495/519 nm) were measured for each fraction using a Tecan M1000 plate reader.

ACKNOWLEDGEMENTS

We thank Feng Guo for providing the T7 polymerase; A. L. N. Rao for providing the cDNA used in generating BMV RNA 1 and for helpful discussions; and Ivo Atanasov for advice on TEM measurements. TEM images were obtained in the California NanoSystems Institute (CNSI) Electron Imaging Center for Nano-Machines supported by NIH (1S10RR23057). This work was supported by the U. S. National Science Foundation, in the form of grant CHE 1051507 to W.M.G. and C.M.K.

FIGURES

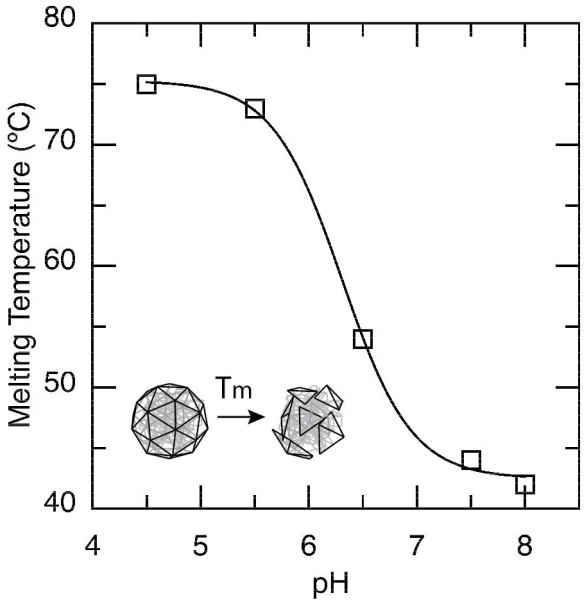

Figure 2-1. Strength of CP-CP attraction is pH dependent. The apparent melting temperature of wild-type CCMV decreases sharply as a function of pH at I = 0.1M.

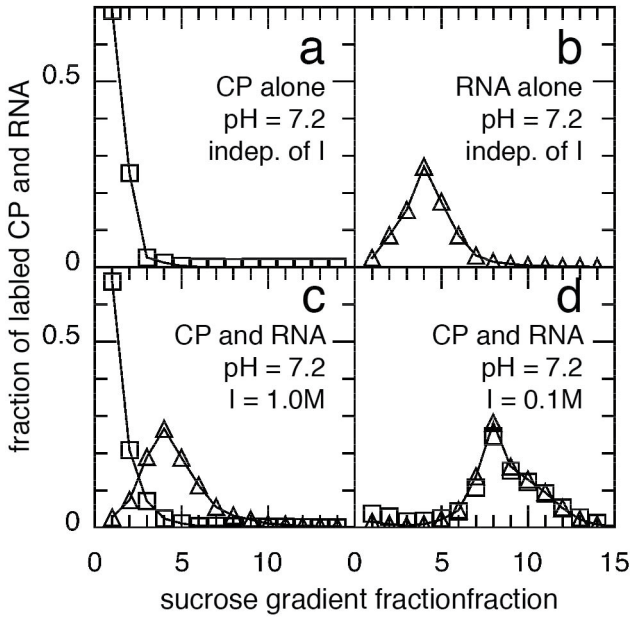

Figure 2-2. Strength of CP-RNA attraction depends on ionic strength. Solutions of fluorescently labeled CP and RNA in either high-salt assembly buffer (pH=7.2, I=1M) or low-salt assembly buffer (pH=7.2, I=0.1M) were analyzed by sucrose density gradient centrifugation. Fractions were collected from top (fraction 1) to bottom (fraction 19) and the fluorescence intensity was measured. (a) Purified CP has a salt-independent sedimentation profile that peaks at the top of the gradient. (b) Purified RNA also showed a salt-independent profile. (c) CP and RNA equilibrated in high-salt assembly buffer 4 h at 4º C do not interact. (d) CP and RNA equilibrated in low-salt assembly buffer for 4 h at 4º C co-sediment as a high-density complex of CP and RNA.

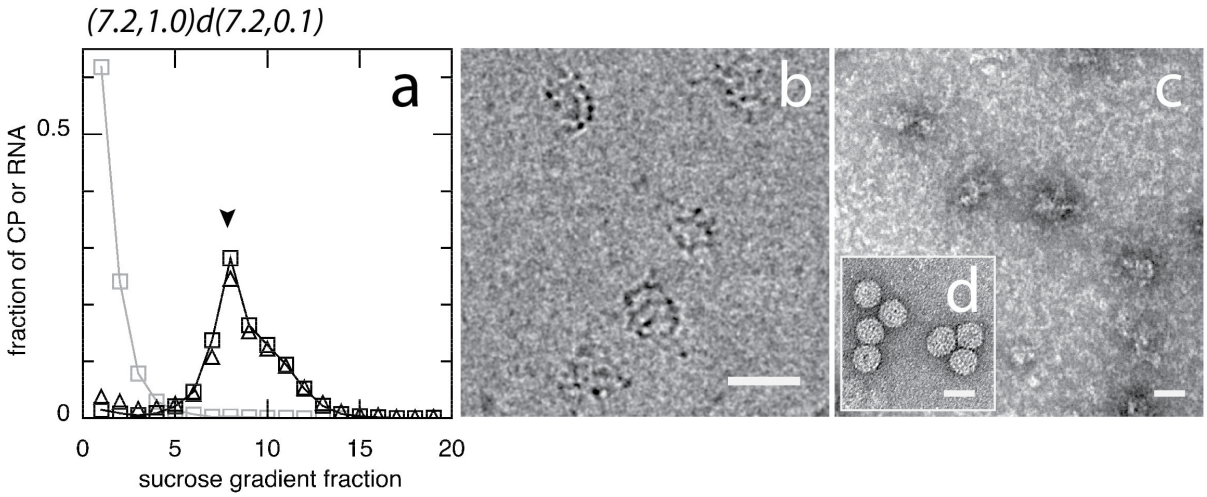

Figure 2-3. Assembly under conditions of weak CP-CP attraction results in disordered complexes of CP and RNA. Fluorescently labeled CP and RNA were mixed in *(7.2,1.0)-buffer*, dialyzed against *(7.2,0.1)-buffer* and equilibrated overnight at 4ºC. (a) The sedimentation profile of the assembly products with CP is plotted as black squares and that of RNA as black triangles. An aliquot of product that was treated with RNase A was analyzed in parallel and the CP profile is plotted as grey squares. The arrow shows the peak fraction of CCMV virions when analyzed under identical conditions. Fractions were collected from top (fraction 1) to bottom (fraction 19). (b) Cryo-EM micrograph of the reaction products. (c) and (d) Uranyl acetate negative-stain-EM micrographs – in (c) the stain was carefully and completely removed from the grid before storage, in (d) excess stain was not completely removed. Scale bars show 28 nm.

(7.2,1.0)d(7.2,0.1)

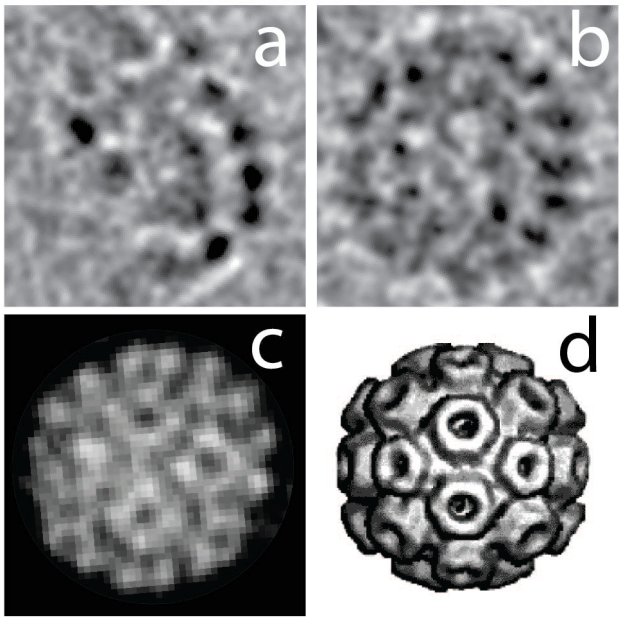

Figure 2-4. Disordered CP-RNA complexes display some capsid features. (a)-(c) Regions of electron micrographs that have been enlarged to enable the examination of single particles. (a) Cryo-EM image of a CP-RNA complex visible in the upper left corner of Fig. 2-3b. (b) Cryo-EM image of wtCCMV. (c) Negative-stain-EM image of wtCCMV. (d) Cryo-EM reconstruction of the CCMV virion taken with permission from Fox et al.[38] Note that the right-most edge of the largely disordered complex shown in (a) has curvature and "knobbiness" that are consistent with the CCMV capsid and its morphological units.

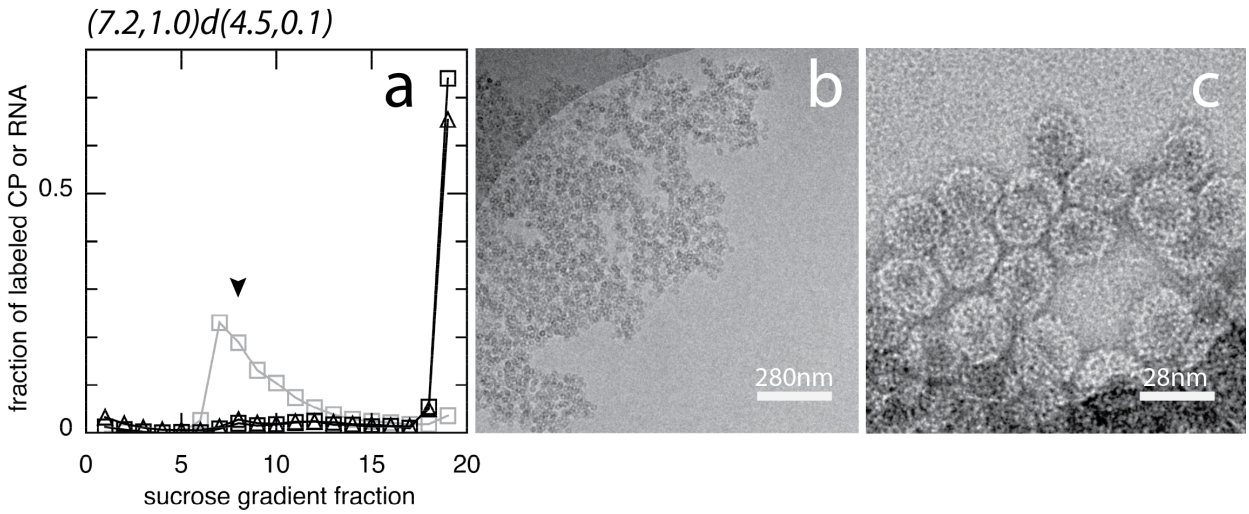

Figure 2-5. Aggregated structures of irregular capsids result from one-step low-pH assembly. Fluorescently labeled CP and RNA were mixed in (7.2,1.0)-buffer, dialyzed against (4.5,0.1)-buffer and equilibrated overnight at 4ºC. (a) The sedimentation profile of the assembly products with CP is plotted as black squares and RNA as black triangles. An aliquot of product that was treated with RNase A was analyzed in parallel and the CP profile is plotted as grey squares. The arrow shows the peak fraction of CCMV virions when analyzed under identical conditions. Fractions were collected from top (fraction 1) to bottom (fraction 19). (b) Cryo-EM micrograph of the reaction products at low magnification; note the aggregates are microns in size. (c) Negative-stain-EM micrograph. Scale bar shows 280 nm in (b) and 28 nm in (c).

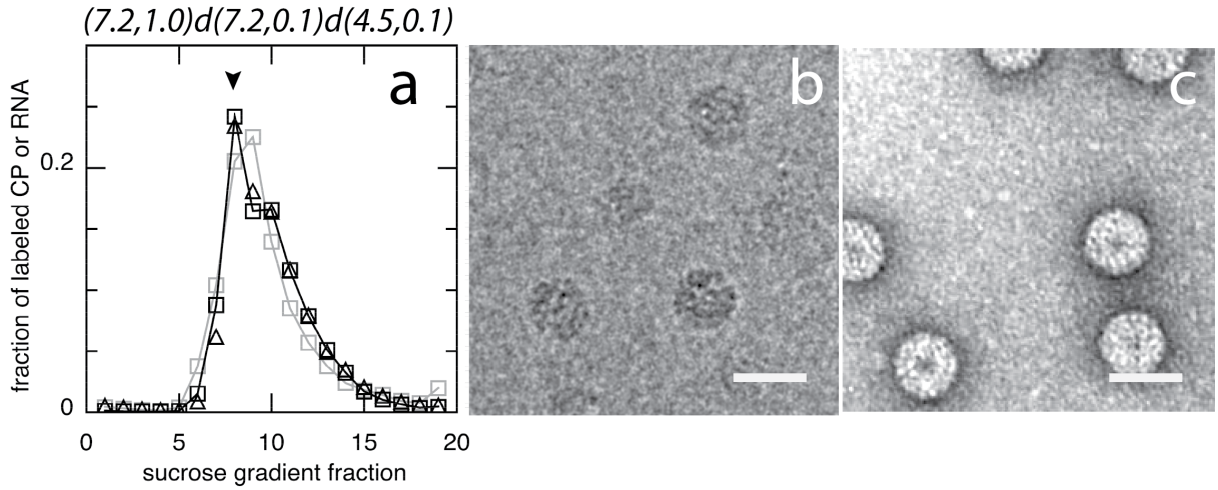

Figure 2-6. Two-step pH-dependent assembly generates capsid structures from disordered complexes by turning on CP-CP attraction. Fluorescently labeled CP and RNA assembled into disordered complexes as in Fig. 2-3, was further dialyzed against low pH *(4.5,0.1)-buffer* and equilibrated for 6h at 4ºC. (a) The sedimentation profile of the assembly products with CP is plotted as black squares and RNA as black triangles. An aliquot of product that was treated with RNase A was analyzed in parallel and the CP profile is plotted as grey squares. The arrow shows the peak fraction of CCMV virions when analyzed under identical conditions. Fractions were collected from top (fraction 1) to bottom (fraction 19). (b) Cryo-EM micrograph. (c) Negative-stain-EM micrograph. Scale bars show 28 nm.

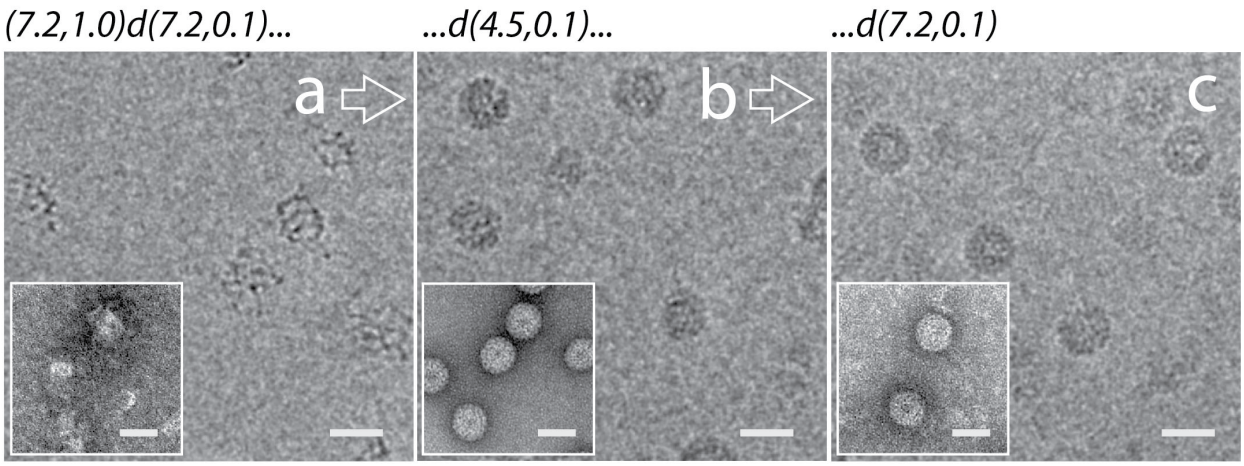

Figure 2-7. Nucleocapsid formation shows strong hysteresis. Cryo-EM micrographs (negative–stain images are inlaid) of assembly reactions with different buffer histories. (a) Images of disordered CP-RNA complexes, prepared as in Fig. 2-3, in neutral pH *(7.2,0.1)-buffer*. (b) Spherical VLPs formed by dialyzing the complexes in (a) against low pH *(4.5,0.1)-buffer*. (c) The spherical particles formed in (b) maintain their structure after dialysis back to neutral pH *(7.2,0.1)-buffer*. Note, the samples imaged in (a) and (c) are composed of chemically equivalent mixtures of CP and RNA suspended in identical buffer solution; they differ only in their buffer history – (c) has been previously acidified, while (a) has not. Scale bars show 28 nm.

(7.2,1.0)d(7.2,0.1)d(π,0.1)

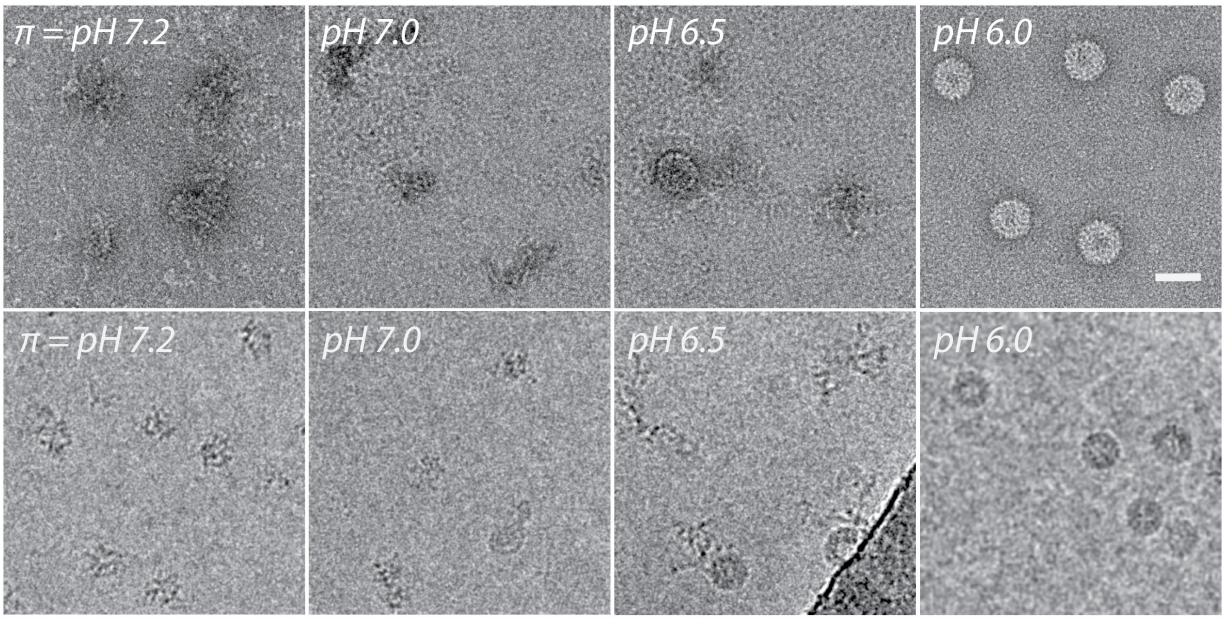

Figure 2-8. The transition between disordered complexes and spherical particles occurs between pH 6.5 and 6. CP and RNA were assembled into disordered complexes, as in Fig. 2-3, and aliquots were dialyzed, in parallel, against buffered solutions of constant ionic strength (I=0.1M) and successively lower pH. (Negative-stain- (top) and cryo-EM (bottom) micrographs are shown for final pH values π = 7.2, 7.0, 6.5, and 6.0. Disordered complexes dominate at pH values above 6. A sharp transition occurs between pH 6 and 6.5 in which all complexes are converted to spherical VLPs. Scale bar shows 28 nm.

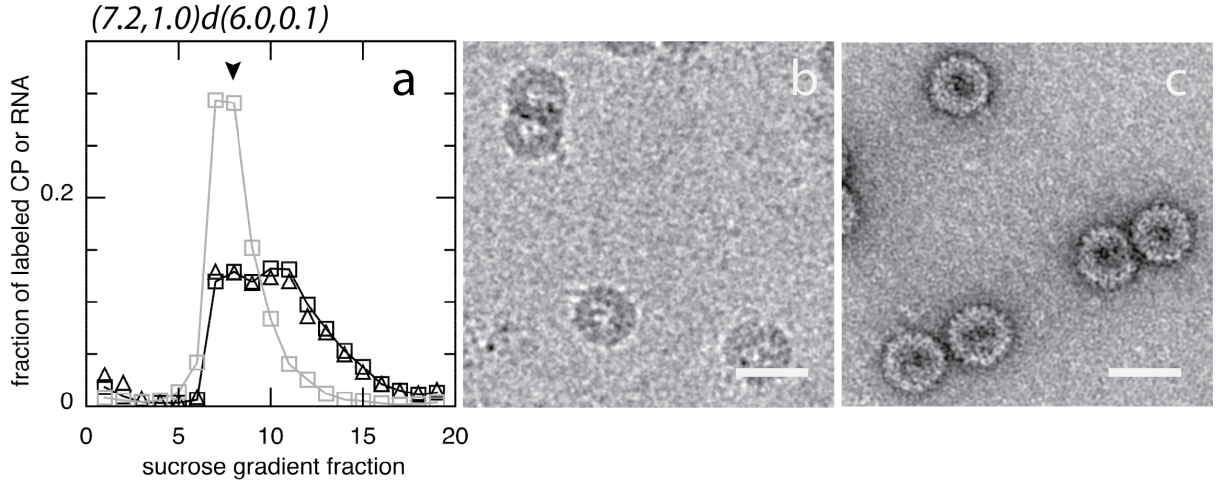

Figure 2-9. Singlet and doublet capsid structures result from one-step moderate-pH assembly. Fluorescently labeled CP and RNA were mixed in *(7.2,1.0)-buffer*, dialyzed against *(6.0,0.1)-buffer* and equilibrated overnight at 4ºC. (a) The sedimentation profile of the assembly products with CP is plotted as black squares and that of RNA as black triangles. An aliquot of product that was treated with RNase A was analyzed in parallel and the CP profile is plotted as grey squares. The arrow shows the peak fraction of CCMV virions when analyzed under identical conditions. Fractions were collected from top (fraction 1) to bottom (fraction 19). (b) Representative cryo-EM micrograph of the reaction products. (c) Representative negative-stain-EM micrograph. Note the presence of doublet particles. Scale bars show 28 nm.

REFERENCES

1. Bancroft, J. & Hiebert, E. (1967). Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. Virology 32, 354-356.

2. Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M. & Gelbart, W. M. (2012). Self-assembly of viral capsid protein and RNA molecules of

different sizes: requirement for a specific high protein/RNA mass ratio. J. Virol. 86, 3318-3326.

3. Adolph, K. & Butler, P. (1976). Assembly of a spherical plant virus. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 276, 113-122.

4. Lavelle, L., Gingery, M., Phillips, M., Gelbart, W. M., Knobler, C. M., Cadena-Nava, R. D., Vega-Acosta, J. R., Pinedo-Torres, L. A. & Ruiz-Garcia, J. (2009). Phase diagram of self-assembled viral capsid protein polymorphs. J. Phys. Chem. B 113, 3813-3819.

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

6. Tama, F. & Brooks, C. L. (2002). The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. J. Mol. Biol. 318, 733-747.

7. Hiebert, E. & Bancroft, J. B. (1969). Factors affecting the assembly of some spherical viruses. Virology 39, 296-311.

8. Adolph, K. & Butler, P. (1977). Studies on the assembly of a spherical plant virus III. Reassembly of infectious virus under mild conditions. J. Mol. Biol. 109, 345-357.

9. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. (2000). Mechanism of capsid assembly for an icosahedral plant virus. Virology 277, 450-456.

10. Johnson, J. M., Willits, D. A., Young, M. J. & Zlotnick, A. (2004). Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. J. Mol. Biol. 335, 455-464.

11. Comas-Garcia, M., Cadena-Nava, R. D., Rao, A. L. N., Knobler, C. M. & Gelbart, W. M. (2012). In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. J. Virol. 86, 12271-12282.

12. Bancroft, J. B. (1970). The self-assembly of spherical plant viruses. Adv. Virus Res. 16, 99-134.

13. Johnson, J. M., Tang, J., Nyame, Y., Willits, D., Young, M. J. & Zlotnick, A. (2005). Regulating self-assembly of spherical oligomers. Nano Lett. 5, 765-770.

14. Vriend, G., Hemminga, M. A., Verduin, B., De Wit, J. L. & Schaafsma, T. J. (1981). Segmental mobility involved in protein-RNA interaction in cowpea chlorotic mottle virus. FEBS lett 134, 167-171.

15. Zhao, X., Fox, J. M., Olson, N. H., Baker, T. S. & Young, M. J. (1995). In vitro assembly of Cowpea chlorotic mottle virus from coat protein expressed in Escherichia coli and in vitro transcribed viral cDNA. Virology 207, 486-494.

16. Bayer, T. S., Booth, L. N., Knudsen, S. M. & Ellington, A. D. (2005). Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA 11, 1848-1857.

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

18. Choi, Y. G., Dreher, T. W. & Rao, A. L. N. (2002). tRNA elements mediate the assembly of an icosahedral RNA virus. Proc. Natl. Acad. Sci. USA 99, 655-660.

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

20. Sorger, P., Stockley, P. & Harrison, S. (1986). Structure and assembly of turnip crinkle virus II. Mechanism of reassembly in vitro. J. Mol. Biol. 191, 639-658.

21. Choi, Y. G. & Rao, A. L. N. (2000). Molecular studies on bromovirus capsid protein. VII. Selective packaging on BMV RNA4 by specific N-terminal arginine residuals. Virology 275, 207-217.

22. Annamalai, P. & Rao, A. (2005). Dispensability of 3′ tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. Virology 332, 650-658.

23. McPherson, A. (2005). Micelle formation and crystallization as paradigms for virus assembly. Bioessays 27, 447-458.

24. Devkota, B., Petrov, A. S., Lemieux, S., Boz, M. B., Tang, L., Schneemann, A., Johnson, J. E. & Harvey, S. C. (2009). Structural and Electrostatic Characterization of Pariacoto Virus: Implications for Viral Assembly. Biopolymers 91, 530-538.

25. Elrad, O. M. & Hagan, M. F. (2010). Encapsulation of a polymer by an icosahedral virus. Physical Biology 7, 045003.

26. Perlmutter, J. D., Qiao, C. & Hagan, M. F. (2013). Viral genome structures are optimal for capsid assembly. eLife, http://arxiv.org/abs/1305.3556.

27. Hohn, T. (1969). Role of RNA in the assembly process of bacteriophage fr. J. Mol. Biol. 43, 191-200.

28. Zhang, R. & Linse, P. (2013). Icosahedral capsid formation by capsomers and short polyions. J. Chem. Phys. 138, 154901/1-14.

29. Zlotnick, A. (2007). Distinguishing Reversible from Irreversible Virus Capsid Assembly. J. Mol. Biol. 366, 14-18.

30. Zlotnick, A., Porterfield, J. Z. & Wang, J. C.-Y. (2013). To build a virus on a nucleic Acid substrate. Biophys. J. 104, 1595-1604.

31. Jacrot, B. (1975). Studies on the assembly of a spherical plant virus. J. Mol. Biol. 95, 433-446.

32. Tsvetkova, I., Chen, C., Rana, S., Kao, C. C., Rotello, V. M. & Dragnea, B. (2012). Pathway switching in templated virus-like particle assembly. Soft Matter 8, 4571-4577.

33. Kim, K. S. (1977). An Ultrastructural Study of Inclusions and Disease Development in Plant Cells Infected by Cowpea Chlorotic Mottle Virus. J. Gen. Virol. 35, 535-543.

34. Rao, A. L. N., Duggal, R., Lahser, F. C. & Hall, T. C. (1994). Analysis of RNA replication in plant viruses. Methods in molecular genetics 4, 216-236.

35. Gillitzer, E., Willits, D., Young, M. & Douglas, T. (2002). Chemical modification of a viral cage for multivalent presentation. Chem. Commun., 2390-2391.

36. Dreher, T. W., Rao, A. L. & Hall, T. C. (1989). Replication in vivo of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. 206, 425-438.

37. Hema, M., Murali, A., Ni, P., Vaughan, R. C., Fujisaki, K., Tsvetkova, I., Dragnea, B. & Kao, C. C. (2010). Effects of amino-acid substitutions in the Brome mosaic virus capsid protein on RNA encapsidation. Mol. Plant-Microbe Interact. 23, 1433-1447.

38. Fox, J. M., Wang, G., Speir, J. A., Olson, N. H., Johnson, J. E., Baker, T. S. & Young, M. J. (1998). Comparison of the native CCMV virion with in vitro assembled CCMV virions by cryoelectron microscopy and image reconstruction. Virology 244, 212-218.

CHAPTER 3

The Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus

ABSTRACT We have recently discovered (Cadena-Nava, et al. *J. Virol.* 2012, *86*, 3318) that the *in vitro* packaging of RNA by the capsid protein (CP) of Cowpea Chlorotic Mottle Virus is optimal when there is a significant excess of CP, specifically that complete packaging of all the RNA in solution requires sufficient CP to provide charge matching of the N-terminal positively-charged arginine-rich motifs (ARMS) of the CPs with the negatively-charged phosphate backbone of the RNA. We show here that packaging results from the initial formation of a "charge-matched" protocapsid consisting of RNA decorated by a disordered arrangement of CPs. This protocapsid reorganizes into the final, icosahedrally-symmetric, nucleocapsid by displacing the excess CPs from the RNA to the exterior surface of the emerging capsid, through electrostatic attraction between the ARMs of the excess CP and the negative charge density of the capsid exterior. As a test of this scenario we prepare CP mutants with "extra" and "missing" (relative to wildtype) cationic residues and show that a correspondingly smaller and larger excess of CP is needed, respectively, for complete packaging of RNA.

IMPORTANCE Cowpea Chlorotic Mottle Virus (CCMV) has long been studied as a model system for the assembly of single-stranded RNA viruses. While much is known about the electrostatic interactions within the CCMV virion, relatively little is known about these interactions *during* assembly – i.e., within intermediate states preceding the final nucleocapsid structure. Theoretical models and coarse-grained molecular dynamics simulations suggest that

viruses like CCMV assemble by the bulk adsorption of CPs onto the RNA – driven by electrostatic attraction – followed by structural reorganization into the final capsid. Such a mechanism facilitates assembly by condensing the RNA for packaging while simultaneously concentrating the local density of CP for capsid nucleation. We provide experimental evidence of such a mechanism by demonstrating that efficient assembly is initiated by the formation of a disordered protocapsid complex whose stoichiometry is governed by electrostatics (charge matching of the anionic RNA and the cationic N-termini of the CP).

INTRODUCTION

Unlike double-stranded DNA viruses, which generally assemble by first forming an empty capsid into which their stiff, extended, genome is subsequently packaged by strong molecular motors powered by ATP hydrolysis, viruses with positive-sense single-stranded (ss)-RNA genomes *spontaneously* assemble by coupling the packaging of their flexible, compact, genome to the formation of their capsid through concerted CP-RNA and CP-CP attractive interactions (1).

Strong electrostatic attraction between the CP and the ss-RNA genome provides the initial driving force for assembly of a number of such viruses. Consistent with this fact, their CPs typically contain extended, positively-charged, arginine-rich motifs (ARMs) that bind strongly to the negatively-charged phosphate backbone of RNA (2). Removing these ARMs, either by mutation or proteolytic cleavage, destroys the ability of the CP to package the RNA (3-5). Furthermore, CP alone – i.e., in the absence of the viral genome or an appropriate anionic scaffold – often is unable to assemble into capsid-like structures under physiological conditions (6-8).

Equilibrium thermodynamics and statistical mechanics have been used to understand how electrostatics influences the stability of fully formed RNA viruses and the optimum lengths of their packaged genomes. Specifically, the observation that the interiors of most RNA viruses are (negatively) *overcharged* (2) – meaning that the negative charge brought by the nucleic acid is greater than the positive charge brought by the basic residues lining the interior of the capsid – has been treated in detail (9-12). While recent coarse-grained molecular dynamics simulations (11, 13-17) have begun to elucidate the role of electrostatics *during* assembly, experimental data remain scarce due to the difficulty of resolving the inherently short-lived assembly intermediates.

In a previous study on the *in vitro* self-assembly of Cowpea Chlorotic Mottle Virus (CCMV) CP and RNAs of various lengths (18), we reported that optimal packaging conditions required that the assembly mixture contain a relative concentration of CP/RNA in excess of what is found in the final capsid structure. More precisely, complete packaging of all the RNA in solution, irrespective of its length (from 140 nt to 12,000 nt), called for the total positive charge brought by the ARMs of the CP to be approximately equal to the total negative charge brought by the RNA. For lengths like those of CCMV RNA1 and RNA2, each of which is packaged separately *in vivo* and involves about 3000 nt, as many as 300 CPs (150 CP dimers) are required for complete packaging (6), consistent with the fact that the ARM of CCMV CP carries a net charge of +10. The role of this specific amount of excess CP within the pathway of assembly remained unclear, e.g., does it interact directly with the RNA or with other CPs within the intermediate states or is it simply required by the law of mass action for driving one capsid-worth of CP onto the RNA? The inability to probe any of the assembly intermediates left us without answers to these questions.

Recent progress in our ability to control the attractive forces that drive assembly has allowed us to trap stable (i.e., long-lived) assembly intermediates and characterize their coarse-grained structure by cryo-EM (19). These intermediates – formed under conditions of weak CP-CP attraction found at neutral pH and low ionic strength – appear as disordered complexes consisting of an undetermined number of CP molecules bound to single RNAs through strong electrostatic attraction. The complexes, which we refer to as protocapsids, were irreversibly converted to well-formed nucleocapsids upon increasing CP-CP attraction by acidifying the assembly buffer – a "two-step assembly" protocol. Interestingly, we observed that one-step protocols that failed to generate these particular intermediates also failed to produce well-formed nucleocapsids and instead gave rise to kinetically-trapped aggregated structures.

In the work presented here we consider CPs carrying different amounts of cationic charge and systematically quantify the average number of CPs that must bind the RNA within each protocapsid in order to drive nucleocapsid formation. From this, we verify that the intermediate electrostatic conditions required for packaging involve approximate *charge matching* of the ARMs of the bound CP and the phosphate backbone of the RNA – distinct from the *overcharged* state found for the final assembly products. More explicitly, we show that the formation of a charge-neutral complex is the key requirement for efficient assembly under the *in vitro* conditions described. Lastly, we show that the structural reorganization that leads to the final nucleocapsid involves the translocation of the excess CP, from the RNA to the exterior surface of the forming capsid, in a process that is facilitated by the electrostatic attraction between the ARMs of the displaced CP and the negative surface charge density of the capsid.

METHODS AND MATERIALS

CCMV purification and labeling

CCMV was purified from infected California cowpea plants (*Vigna unguiculata* cv. Black Eye) as previously described by Bancroft (20). Alexa Fluor-647 NHS ester (AF647-NHS; Molecular Probes, U.S.A.) was covalently linked to solvent-exposed lysines on the exterior surface of CCMV by direct mixing of 5 mg/mL CCMV and AF647-NHS at a mass ratio of 200:1, in 0.1 M HEPES (pH 7.2), 5 mM $MgCl_2$, for 2 h at 4ºC, followed by overnight dialysis against low pH, low ionic strength buffer. Unreacted dye was removed by collecting the capsids in a sucrose cushion, followed by resuspension. The density of labeling (DOL), defined as the fraction of CPs labeled by AF647-NHS, was determined by UV-Vis spectroscopy according to the manufacturer.

Purification of CP

CCMV CP was isolated from fluorescently labeled CCMV with a DOL of ~0.05, as well as from unlabeled CCMV, as previously described by Annamalai, et al. (21). The CP concentration and the degree of RNA contamination were measured by UV-Vis spectrophotometry; only CP solutions with 280/260 ratios greater than 1.5 (corresponding to <0.5% RNA contamination) were used in assembly reactions. SDS-PAGE and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry were used to make certain that the purified protein was not cleaved.

Synthesis of RNA

BMV RNA 1 was generated by *in-vitro* transcription of pT7B1 plasmid (22) with T7 RNA polymerase, rATP, rGTP, rCTP, and rUTP. RNA was purified in TE buffer by digesting the template DNA with DNase I followed by repeated washing of the digestion products using a 100 kDa Amicon Filter device.

A 500-nt RNA consisting of the 5' end of BMV RNA1 was prepared by PCR amplification of the first 500 bp of the pT7B1 plasmid, followed by *in-vitro* transcription and purification according to the protocol mentioned above.

Synthesis and purification of mutant CP

Two mutant variants, R10P and R13P/R14G, were genetically engineered by introducing specific mutations into the wild-type pET15b-CCMV plasmid template using standard QuikChange site-directed mutagenesis techniques (Stratagene). Here, R13P/R14G refers to the double mutant variant to which two arginine residues of the N-terminal ARM were replaced with single proline and glycine residues, resulting in an overall +8 charged ARM. In the case of R10P, in addition to the replacement of a single arginine of the N-terminal ARM into a proline, a 10-amino acid residue extension was also inserted into the ARM, resulting in an overall +12 charged ARM. In both cases, the mutations were confirmed by DNA sequencing (MWG Eurofins, Germany) and the plasmid DNA was transformed into *E. coli* BL21(DE3)pLysS (bearing an N-terminal hexahistidine-tag, Novagen) for protein expression. Starting cultures were grown overnight at 37°C from glycerol stock cells in 7 mL of LB medium containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Overnight cultures were used to inoculate 0.5 L LB medium containing ampicillin (100 µg/mL) and chloramphenicol (34 µg/mL) and grown to an

optical density of OD_{600} 0.6. Protein expression was induced following addition of IPTG to a final concentration of 0.1 mM at 30°C for 5 h. The cells were harvested by centrifugation (10,000 g for 15 min) and the cells were lysed using BugBuster according to standard procedures (Novagen). R10P and R13P/R14G were purified using nickel affinity column chromatography with a modified version of the supplier protocol (Novagen). Weakly bound and other unwanted proteins were washed with 0.1 M phosphate buffer (pH 8.0), 0.3 M NaCl and 12.5 mM imidazole, before washing with 0.1 M phosphate buffer (pH 8.0), 1.5 M NaCl, 12.5 mM imidazole to remove bound RNA. Purified protein was eluted with 0.1 M phosphate buffer (pH 8.0), 1.5 M NaCl, 0.25 M imidazole before dialyzing overnight to 50 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 10 mM $MgCl_2$, 1 mM EDTA, to remove excess imidazole. The purified protein was dialyzed to 50 mM sodium acetate buffer (pH 5.0), 1 M NaCl and, 1 mM NaN_3 to induce empty capsid formation and frozen to -20ºC.

Prior to removal of the His-tag, the samples were thawed and dialyzed against 1 M NaCl, 20 mM Tris pH 7.2 buffer for 6 h at 4ºC. After dilution with 10X cleavage buffer (Novagen), 1 unit of biotinylated thrombin (Novagen) was added per mg of CP, and the reaction was carried out at 4ºC for 18 h. After digestion agarose/streptavidine beads (Novagen) were added to the sample and incubated for 30 minutes at room temperature and were removed by filtering through a 1 mm spin filter. The buffer was exchanged with 20 mM sodium phosphate pH 7.4, 0.5 M NaCl and 10 mM imidazole using a 10 kDa MWCO Amicon spin filter (Millipore). The sample was purified from the cleaved His-tag as well as the uncleaved CP by using a 10 mL His-tag affinity column (GE Healthcare). Finally it was concentrated and buffer exchanged to buffer B (plus 1 mM EDTA, 1 mM PSMF and 1 mM DTT) with a 10 kDa Amicon filter. The integrity of the samples was double-checked by 15% SDS-PAGE and MALDI-TOF.

Two-step *in vitro* assembly reactions

Assembly reactions were performed following the previously optimized two-step protocol (19). Specifically, 30 nM BMV RNA1 and varying concentrations of CP were mixed under conditions that prevent their interaction (neutral pH and high ionic strength; 1 M NaCl, 20 mM Tris [pH 7.2], 1 mM EDTA, 1 mM DTT, 1 mM PMSF). This non-interacting mixture of RNA and CP was subjected to two subsequent dialysis and equilibration steps, each carried out for at least 6h at 4°C. Here we have assumed that 6h is sufficient for the reactions to reach equilibrium. The first step turns on CP-RNA attraction while maintaining weak CP-CP attraction (neutral pH and low ionic strength; 50 mM NaCl, 50 mM Tris-HCl [pH 7.2 or 7.0], 10 mM KCl, 5 mM $MgCl_2$) and results in the formation of disordered CP-RNA complexes that serve as the intermediate templates (protocapsids) to proper assembly. The second step turns on CP-CP attraction (low pH and low ionic strength; 50 mM sodium acetate [pH 4.5], 8 mM magnesium acetate) and triggers the irreversible formation of the final nucleocapsid structure.

For assembly reactions involving 500-nt RNA, 500 nM RNA was mixed with 12.5 μM CP and subjected to the two-step assembly protocol described above.

Electron microscopy

Cryo-EM (Figure 3-2): 3 μL of assembly reaction was deposited on a Quantifoil holey carbon grid (200 mesh; R2/1) that had been previously glow-discharged. The grids were then blotted and flash-frozen by rapid plunging into liquid ethane cooled to liquid nitrogen temperature. Micrographs were acquired using an FEI Tecnai G2 TF20 microscope operated at an accelerating voltage of 200 kV. Images were recorded at 3-4 microns underfocus with a

TIETZ F415MP 4k × 4k pixel CCD camera. Total beam exposure was maintained between 20-40 e/Å^2 for cryo-EM samples.

Negative stain (Figure 3-4): 6 µL of assembly reaction was deposited on glow-discharged copper grids (400-mesh) that previously had been coated with Parlodion and carbon. After 1 min, the grids were blotted and stained with 6 µL of 2% uranyl acetate for 1 min followed by complete stain removal and storage in a desiccator overnight.

Gel-shift titration assay

A 10-µL aliquot of each assembly reaction was mixed with 3 µL of glycerol and loaded into a 1% agarose gel in virus electrophoresis buffer (0.1 M sodium acetate, 1 mM EDTA, pH 4.5 or 5.5). The samples were electrophoresed at 4°C at 50 V and stained with a solution of 5 µg/mL ethidium bromide. The gels were visualized with a FX Pro Plus Fluorimager/PhosphorImager (Bio-Rad) by exciting the RNA-intercalated ethidium bromide and measuring the emitted fluorescence intensity.

Velocity sedimentation by sucrose density gradient centrifugation

10-40% (w/v) sucrose gradients were made by subjecting 3.5 mL of 25% sucrose in the desired buffer to repeated (3X) freeze-thaw cycles. The gradients were loaded with 0.2 mL of sample and centrifuged for 3 h at 30,000 rpm at 4ºC in a SW50.1 rotor (Beckman-Coulter) and manually fractionated by pipetting 0.2-mL aliquots from top to bottom. The 19 extracted fractions were loaded into a 96-well flat-black flat-bottom plate. The fluorescence of the labeled CP (AF647; ex/em 650/668 nm) was measured for each fraction using a Tecan M1000 plate reader.

RNase A digestion assay

Assembly reactions containing 150 and 90 CP_2:RNA (mol CP dimer : mol RNA) that were subjected to the first step of the two-step assembly protocol, as well as wild-type (wt) CCMV and naked BMV RNA1 were equilibrated in neutral pH, low-ionic-strength assembly buffer with various concentrations of RNase A for 90 min on ice. 10 µL aliquots were then mixed with 3 µL glycerol and loaded into a 1% agarose gel in virus electrophoresis buffer (0.1 M sodium acetate, 1 mM EDTA, pH 5.5) – constituting the second step (pH-lowering) of assembly.

RESULTS AND DISCUSSION

Assembly of RNA and wild-type CP

Wild-type (wt) CP purified from CCMV virions harvested from infected California cowpea plants (*Vigna ungiculata* cv. Black Eye) form stable dimers (CP_2) in solution that serve as the fundamental assembling units during packaging (23). The RNA used is a 3,200-nt *in vitro* transcript of genomic RNA1 from brome mosaic virus (BMV RNA1). We choose to work with *heterologous* BMV RNA1 instead of the equal-length genomic RNA1 from CCMV in order to reduce the role of sequence specific interactions that may exist between CP and RNA that have coevolved; specific interactions cannot be completely ruled out, however, due to the high degree of sequence homology between CCMV and BMV RNA.

To quantify the extent of packaging as a function of the relative concentration of CP_2 and RNA, mixtures of 30 nM RNA and varying concentrations of CP_2 (from 0 to 5.4 µM) were subjected to the previously optimized two-step, pH- and ionic-strength-dependent assembly protocol described in Methods. Gel shift titration assays (Figure 3-1) of the final assembly

products, stained with ethidium bromide (EtBr), show that the amount of packaged RNA – defined experimentally as the integrated EtBr stain intensity migrating with the same electrophoretic mobility as CCMV virions – depends strongly on the relative amount of CP_2 present in the assembly mixture. Three regions of the assembly titration are evident. Region I, spanning CP_2:RNA ratios of 0 to 50, shows no packaging – instead, all of the RNA signal is present as broad smears with electrophoretic mobilities consistent with a distribution of incomplete CP-RNA complexes. Region II, spanning 50 to 140 CP_2:RNA, shows a coexistence between packaged RNA and incomplete complexes, with the extent of packaging increasing at higher concentrations of CP_2. Region III, corresponding to ratios above 140 CP_2:RNA, exhibits complete packaging – though it should be noted that higher ratios begin to migrate slightly more slowly than CCMV for reasons that will be described later. Quantitative measurements of the total packaged RNA signal are plotted in Figure 3-3C in red.

Consistent with previous work (18, 24, 25), we have established that only 50% of the RNA in solution is packaged upon the equilibration of 90 CP_2:RNA – the relative concentration corresponding to the stoichiometry of the final nucleocapsid structure – and that upwards of 140 CP_2:RNA must be added to the assembly mixture in order to drive complete packaging (Figure 3-3C, red curve). This observation is by itself not surprising – chemical reactions often require an excess of one or more reagents to reach completion – and leaves ambiguous the role of the excess CP_2 within the pathway of assembly. To elucidate its function in assembly, CP_2 was labeled with a fluorescent probe (Alexa Fluor-647 NHS ester) and subjected to the normal two-step RNA-packaging protocol. The reaction was assayed at each step of the assembly by fluorescently-detected sucrose density-gradient velocity sedimentation (Figure 3-2), allowing for the accurate quantification of CP_2 binding.

The green curve of Figure 3-2a is the sedimentation profile for the assembly intermediates generated by equilibrating 150 fluorescently labeled capsid protein dimers (CP_2*) per BMV RNA1 at neutral pH and low ionic strength. The position and width of the sedimentation peak, compared to that of fluorescently labeled CCMV virions (CCMV*; Figure 3-2A, gray curve), are consistent with the distribution of disordered, capsid-sized, CP-RNA complexes observed by cryo-EM (19). Figure 3-2B shows one such protocapsid. Additionally, we see that essentially all of the CP_2* present in the assembly mixture sediments between fractions 6 and 12, signifying that it is bound to RNA – in contrast to free (i.e. unbound) CP_2* that has been shown to sediment between fractions 1 and 2 (19, 23). This observation demonstrates that the excess CP_2 required for complete RNA packaging is tightly bound within the pre-nucleocapsid complex and thus actively involved in the reorganization process that leads to the final virion.

Implicit in this observation is the fact that each RNA is capable of binding an amount of CP_2 in significant excess of the 90 CP_2 found in the final capsid. This is not surprising provided that CP-RNA binding is non-specific and predominantly electrostatic in nature. Consider that BMV RNA1 carries a charge of -3,200 and that extending from each CP_2 is a pair of flexible 26-residue ARMs carrying a combined charge of +20. A purely electrostatic binding model would suggest that an upper-limit of 160 CP_2 (corresponding to a total charge of +3,200) should be able to bind a single RNA before saturation – our observation that an average of 150 CP_2 bind each RNA falls close to this limit.

Cryo-EM (Figure 3-2D) confirms that acidifying the assembly reaction reorganizes the disordered protocapsids containing an excess of bound CP_2 (on average 150 CP_2 per RNA) into well-formed nucleocapsids (19). This process logically requires the unbinding of 60 CP_2 from

each RNA to allow for the remaining 90 CP_2 to assume their place within the capsid lattice. The predicted free energy cost of disrupting this number of strong electrostatic interactions is quite large – roughly 1,200 kT (26) – and has been confirmed by equilibrium binding assays (27). It is a remarkable fact that, despite this penalty, nucleocapsid reorganization proceeds spontaneously. The predominant driving force for the rearrangement comes from the formation of a large number of acidic pH-stabilized contacts between the 90 CP_2 that make up the capsid (28, 29). In addition to these contacts, velocity sedimentation measurements of the reaction mixture after acidification suggest that extra interactions involving the 60 displaced CP_2 also play a role.

The sedimentation profile after acidification (Figure 3-2C, green curve) shows no CP_2* in fractions 1 and 2 – demonstrating that the excess CP_2* has not been released from the virion as free dimers into solution. Instead, we observe an increase in the fast-sedimenting shoulder that is suggestive of nucleocapsids with excess CP_2* bound to their exterior. This interpretation is based on the observed interaction of a stoichiometrically equivalent mixture of CP_2* and unlabeled CCMV virions (60 CP_2*:CCMV). At neutral pH and low ionic strength, we find that 60% of the added CP_2* co-sediments with CCMV (Figure 3-2A, red curve) and shifts the sedimentation peak from fraction 8 (naked CCMV*) to fraction 9 (60 CP_2*:CCMV). This shift, in combination with analytical ultracentrifugation measurements (data not shown), confirms previous observations that excess CP_2* binds the exterior surface of CCMV (30) – as expected from the electrostatic attraction between the ARMs of the CP_2* and the negative surface charge density of CCMV (31). An identical binding assay carried out under acidic conditions (Figure 3-2C, red curve) shows that >95% of the CP_2* binds the exterior surface of CCMV and generates a broader sedimentation peak with a faster-sedimenting shoulder. The increase in total binding at low pH (the reduction in free CP* signal) reflects the strengthening of lateral CP_2-CP_2 interactions

caused by the decrease in electrostatic repulsions between proximal dimers. Thus, we conclude that the pH-triggered reorganization of the protocapsid involves the displacement of excess CP_2 from the RNA to the newly-forming capsid exterior. Electrostatic and lateral interactions involving the displaced CP_2 – in addition to the formation of lateral contacts between the 90 CP_2 that constitute the capsid – help to overcome the energetic cost of unbinding the excess CP_2 from RNA.

While the binding of excess CP_2 to the exterior surface of the assembled nucleocapsids is consistent with the retarded electrophoretic mobilities observed at high CP_2:RNA ratios in Figure 3-1, resolving the excess CP_2 by cryo-EM has proven difficult. Similar assembly reactions carried out using 500-nt truncations of BMV RNA1 that package into "pseudo T=2" capsids produce particles with additional shells (multishells) of CP_2 that can be resolved by cryo-EM (Figure 3-3). The assembly products of full-length BMV RNA1, however, do not show multishell formation – neither do equivalent mixtures of wtCCMV and excess CP_2. In the later cases, the excess surface-bound CP_2 seems to be more disordered and thus harder to visualize. This may be due to the fact that the smaller "T=2" core particles support the growth of multishells with curvatures closer to the preferred T=3 capsid, whereas the larger T=3 cores are less likely to favor the generation of ordered multishells due to the smaller curvatures they would have to adopt (31).

Assembly of RNA and mutant CP with varying charge on the ARM

Two mutant variants (R13P/R14G and R10P) differing only in their N-terminal ARMs were recombinantly expressed and used in these studies – specifically, their ARMs carry different total charges (Table 1). For the following discussion, we refer to the mutants by the

total charge of their ARMs. Explicitly, because the dimer of R13P/R14G carries two ARMs with a combined charge of +16 (vs the wt 20), we refer to it as “mCP16”. Similarly, because R10P carries a charge of +24 (vs 20), we refer to it as “mCP24”.

To test how the charge of the ARM affects the extent of packaging, the usual two-step assembly protocol was carried out using mCP16 and mCP24, in parallel. Gel shift titration assays (Figure 3-4a and 3-4b) of the assembly products again show three distinct regions. Region I shows no packaging, and spans mCP_2:RNA ratios of 0-45 and 0-22 for mCP16 and mCP24, respectively. The incomplete assembly products in Region I appear as relatively sharp bands of stain intensity with electrophoretic mobilities that increase and eventually plateau with increasing concentration of mCP_2. This is distinct from the case of $wtCP_2$ where the smeared bands continuously retard with increasing protein concentration. The cause of this difference remains unclear, though it should be noted that previous studies on RNA packaging by $wtCP_2$ have reported similar increases in electrophoretic mobility at low CP_2:RNA ratios using slightly different assembly protocols (25). Region II shows a two-state coexistence between incomplete complexes and packaged RNA, and spans mCP_2:RNA ratios of 60-165 (mCP16) and 33-121 (mCP24). This two-state behavior is suggestive of cooperative assembly, and is not observed in the reactions containing wtCP. Region III shows complete packaging above mCP_2:RNA ratios of 180 (mCP16) and 132 (mCP24). Quantitative measurements of the total packaged RNA signal measured by gel shift titration assay are plotted in Figure 3-3c. Negative stained electron micrographs (Figure 3-5) show that the assembly products obtained using mCP and wtCP are indistinguishable.

Despite the qualitative differences in the gel shift titration assays, Figure 3-4c demonstrates that the fraction of RNA packaged at suboptimal ratios of CP_2:RNA increases with

the charge of the ARM. Consistent with this observation, the CP_2:RNA ratio required for complete packaging varies inversely with the total charge on the N-termini of the CP_2. The product of the number of charges on the ARM and the CP_2:RNA ratio required for complete packaging gives +2,900 for mCP16, +2,800 for wtCP, and +3,200 for mCP24, i.e., the product is constant to within about 10%, even as the charge of the ARM varies by as much as 40%. Considering the large (of order 10%) experimental error associated with agarose gel densitometry, these measurements establish that approximate charge neutralization of the RNA (which carries a charge of -3,200) is a general requirement for complete packaging under the *in vitro* conditions described. These observations should be contrasted with experiments by Ni et al. (32) on the *in vivo* packaging of RNA by mutant BMV CP with varying arrangements of basic residues within the ARM. They were able to show that the length and types of RNA encapsidated depended not only on the total charge of the ARM but also on the primary sequence of the N-terminal residues. A complete understanding of packaging in terms of electrostatic arguments could not be reached by Ni et al., and their experiments serve to remind us of the relative complexity of RNA packaging *in vivo*.

The effect of RNase on packaging – an unexpected result

The RNase A digestion assay has long been used in *in vitro* assembly experiments to report on packaging – RNA that survives RNase treatment is considered to be packaged. While employing the standard RNase A digestion assay as a complementary technique for quantifying the extent of packaging, we noticed an unexpected result: RNA packaging was enhanced upon the addition of RNase during the first step of assembly to reactions containing an insufficient amount of CP_2.

The products of controlled digestion reactions carried out by equilibrating, at neutral pH, varying concentrations of RNase A with naked RNA and wtCCMV, were electrophoresed at low pH and stained with EtBr (Figures 3-6a and 3-6b, respectively). Comparison of the digestion profiles confirms that the capsid of CCMV protects RNA against degradation – RNase has no measurable effect on wt virions. On the other hand, naked RNA is shown to be incompletely digested by low concentrations of RNase A into a distribution of fragments with lengths of a few hundred nucleotides. The decrease in total RNA signal within each lane containing moderate concentrations of RNase suggests that short fragments (<10s of nt) are also generated that cannot be resolved due to their high electrophoretic mobility and lack of double-stranded segments required for EtBr intercalation.

Assembly reactions containing 150 CP_2:RNA and 90 CP_2:RNA in the first stage of assembly (neutral pH and low ionic strength) were identically treated and analyzed (Figures 3-6c and 6d, respectively). By gel shift assays run at low pH, we measured no effect of RNase on assembly mixtures containing the required excess of CP_2. However, RNase did affect reactions containing only the stoichiometric amount of CP_2 (90 CP_2:RNA) – the fast moving smear of stain intensity (corresponding to incomplete complexes) was removed at moderate enzyme concentrations. Additionally (and remarkably), moderate RNase treatment resulted in an increase in the packaged RNA signal as well as the CP_2 signal associated with the packaged RNA (which was independently quantified by coomassie stain densitometry and plotted in Figure 3-6e). Taken together, these observations suggest that the addition of RNase A to an assembly reaction containing an insufficient amount of CP_2 increases the number of assembled nucleocapsids.

This unexpected increase in assembly products can be understood in terms of our previous observations on the requirements for efficient packaging. Consider that assembly

reactions containing 90 CP_2:RNA package only 50% of the RNA in solution upon acidification and that the remaining RNA is found as incomplete complexes decorated by an insufficient amount of CP_2 to trigger capsid formation. Addition of RNase initiates the digestion of a fraction of these incompletely packaged RNAs and allows for the redistribution of their CP_2 to other unpackaged RNAs. The redistribution is made possible by the reversible binding equilibrium between CP_2 and RNA during the first step of assembly (neutral pH and low ionic strength) and is driven by the tendency for full length RNAs to outcompete shorter molecules at binding CP_2 (24). By digesting a fraction of the RNA available for packaging, RNase treatment effectively increases the CP_2:RNA ratio found in the assembly mixture – resulting in more efficient packaging of the remaining RNA upon acidification.

CONCLUSIONS

While the precise mechanisms underlying the *in vitro* assembly of ss-RNA viruses remain unclear, coarse-grained features have begun to emerge. It has been proposed in previous theoretical studies of the assembly of RNA viruses whose CPs contain highly basic ARMs that the initial formation of a disordered CP-RNA complex facilitates packaging by simultaneously generating a high local density of CP as well as compacting the RNA (13, 14, 33, 34). Our observation that the efficient *in vitro* packaging of RNA by CCMV CP is initiated by the formation of such a complex – consisting of a single RNA molecule decorated by enough CP to match the negative charge of the phosphate backbone – supports these claims. Our results exploiting mutant CPs demonstrate that the total charge of the bound ARMs – rather than the total number of bound CP – must be sufficiently high for efficient packaging. This suggests that the rearrangement of the RNA through interaction with the polycationic ARMs of the CP is a key

step in assembly. Furthermore, we observe that the structural reorganization that gives rise to the final virion involves the transfer of excess CP_2 from the RNA to the exterior surface of the forming capsid through the electrostatic attraction of the ARMs of the displaced CP_2 and the outward-facing negative charge density of the remaining RNA-bound CP_2. This pathway is represented, qualitatively, in Figure 3-7.

ACKNOWLEGEMENTS

We thank Feng Guo for providing the T7 polymerase; A. L. N. Rao for providing the cDNA used in generating BMV RNA 1 and for helpful discussions. TEM images were obtained in the California NanoSystems Institute (CNSI) Electron Imaging Center for Nano-Machines supported by NIH (1S10RR23057). This work was supported by the U. S. National Science Foundation, in the form of grant CHE 1051507 to W.M.G. and C.M.K. Additional funding for R.F.G. was provided by the University of California, Los Angeles Dissertation Year Fellowship.

FIGURES

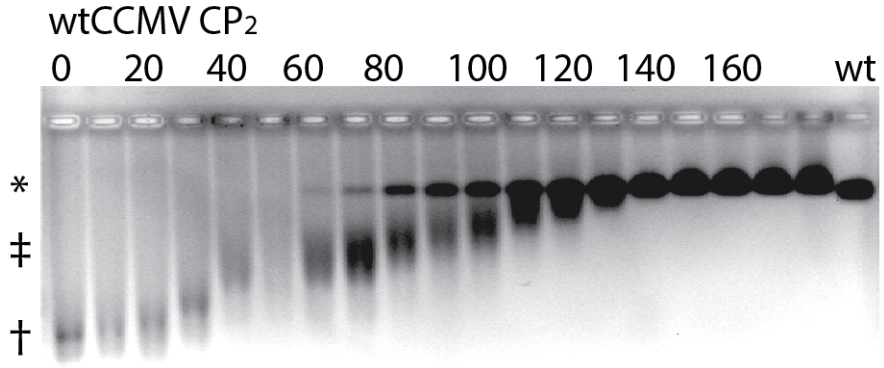

Figure 3-1. CP-RNA assembly titration: Gel shift assay. 1% agarose gel run in low-pH electrophoresis buffer and stained with EtBr. The titration of a fixed concentration of BMV RNA1 with varying amounts of wt capsid protein dimer (CP_2) ranging from 0 (left-most lane) and increasing to the right; molar CP_2:RNA ratios are reported above each lane. The right-most lane contains wtCCMV. The band position of assembled capsids (*), naked RNA (†), and incomplete CP-RNA complexes (‡) have been denoted to the left of the gel. We define "complete packaging" to occur at the point in the titration that corresponds to the disappearance of incomplete CP-RNA complexes (140 CP:RNA). Regions I, II, and III (see text) – i.e., lanes 1-6, 7-14, and 15-19, respectively – correspond to smears of increasingly-CP-bound RNAs (I), coexisting incomplete complexes and nucleocapsids (II), and nucleocapsids with increasing amounts of CP bound on their exteriors (III).

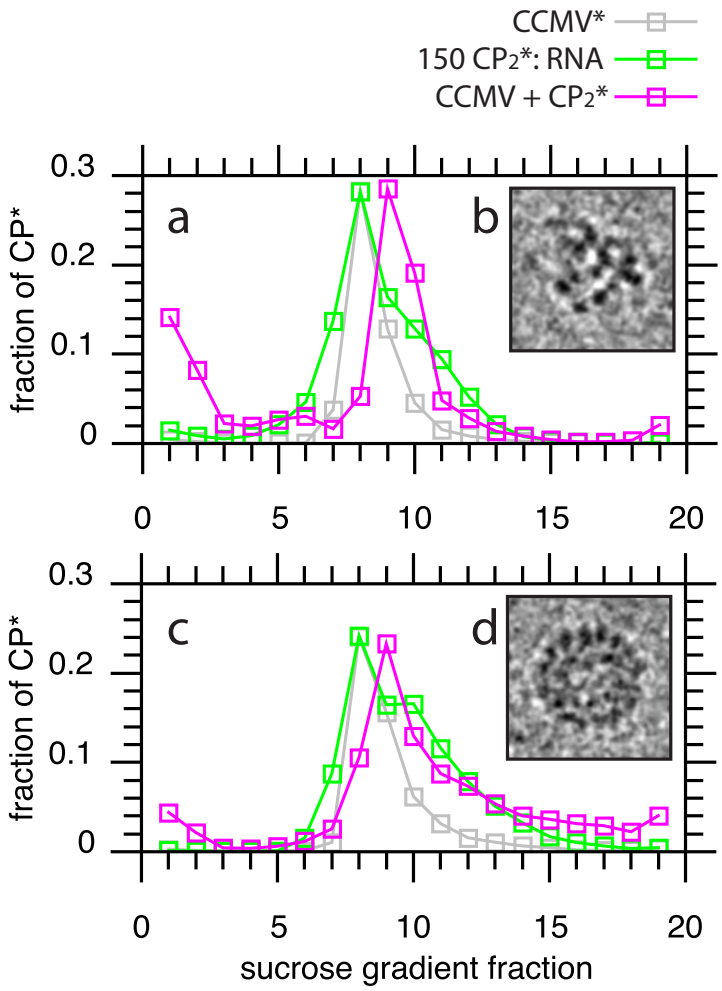

Figure 3-2. Sucrose density gradient velocity sedimentation. Different mixtures containing fluorescently labeled CP dimers (CP_2*) were analyzed by sucrose density gradient centrifugation before (a) and after acidification (c). After centrifugation the 10-40% gradients were fractionated from top (fraction 1) to bottom (fraction 19) and the fluorescence intensity was measured. The sedimentation profile for an assembly reaction consisting of 150 CP_2* per RNA is shown in green; a stoichiometrically equivalent mixture of 60 CP_2* added per CCMV capsid is shown in red; and CCMV virions that were fluorescently labeled (CCMV*) are shown in gray. The peak intensity for CCMV* has been renormalized to match the peak intensity for the assembly

reaction. Inlays (b) and (d) show representative cryo-EM images of the assembly complexes before and after acidification, respectively. Note that none of the CP_2* present in the assembly mixtures sediments as free dimer (which is found in the top fractions of the gradient) – i.e. all of the CP_2* is bound in CP-RNA complexes.

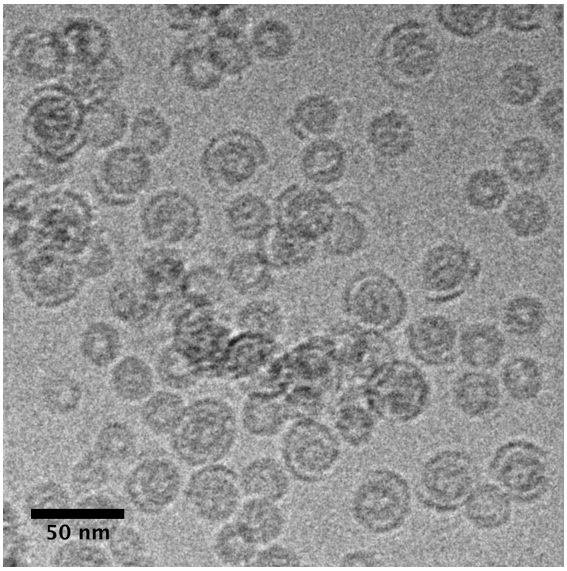

Figure 3-3. The products of an assembly reaction containing a 500-nt truncation of BMV RNA1 and enough CP to completely package the RNA (25 CP_2:RNA, which corresponds to charge-matching between the CP-ARM and the RNA) were imaged by cryo-EM after acidification. A fraction of the "T=2" capsids show a partial second shell consisting of the excess of CP required for charge-matching. Scale bar shows 50 nm.

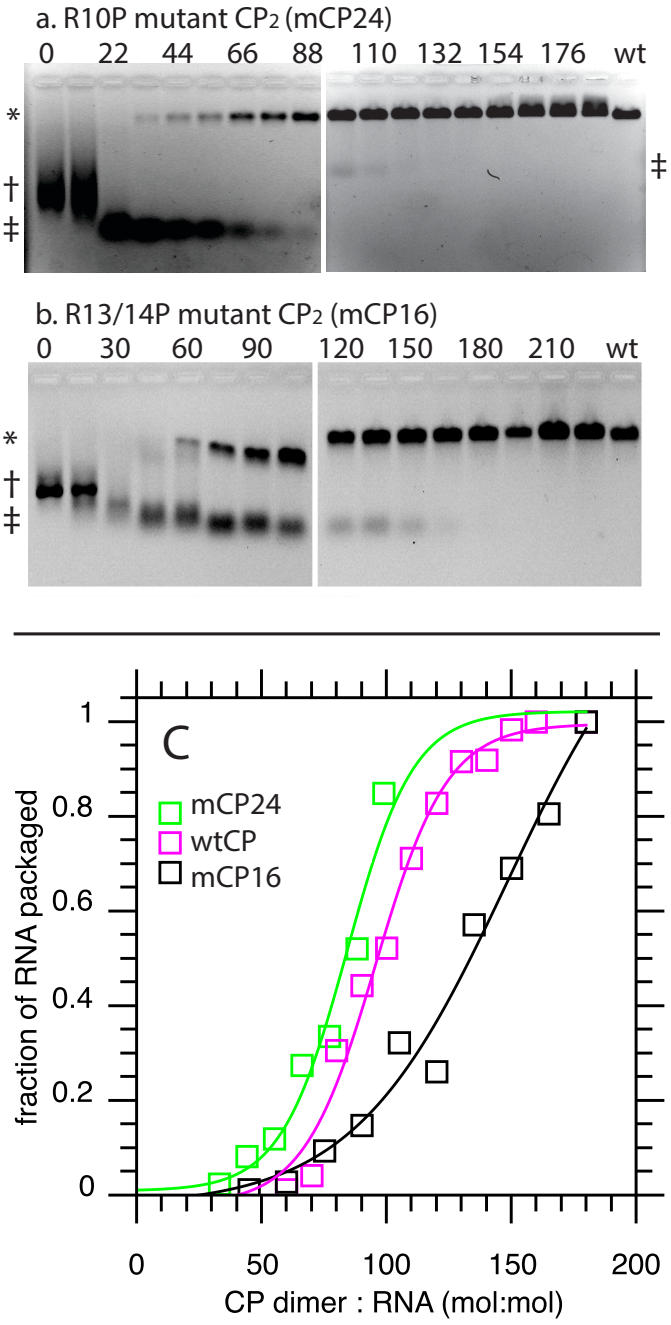

Figure 3-4. CP-RNA assembly titrations and densitometry of gel shift assays. 1% agarose gels run in low-pH electrophoresis buffer and stained with EtBr. As in Figure 3-1, all gels show the titration of BMV RNA1 with varying amounts of mutant capsid protein dimer (mCP_2) ranging from 0 (left-most lane) and increasing to the right; molar mCP_2:RNA ratios are reported above each lane. The right-most lane contains wtCCMV. The band position of assembled capsids (*),

naked RNA (†), and incomplete CP-RNA complexes (‡) have been denoted to the left of the gels. (a) an assembly titration involving mutant CP_2 which carries a total N-terminal charge of +24 per dimer. (b) a titration using mutant CP_2 carrying a N-terminal charge of +16. Note that the relative position of (‡) in the right gel of (a) is different from that in the left gel due to pH changes in the running buffer that occur during electrophoresis. (c) densitometry profiles generated by plotting the integrated EtBr signal from the capsid band of each lane as a function of mCP_2:RNA. The plot has been normalized so that the band intensity corresponding to the point of complete packaging is equal to one. The lines show best-fit sigmoid curves that are meant to aid the eye in following the extent of packaging as a function of CP:RNA.

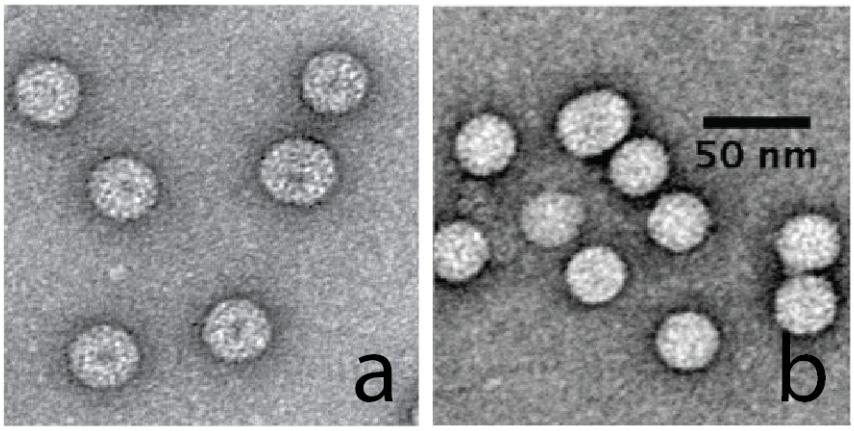

Figure 3-5. Negative-stained-EM images of BMV RNA1 completely packaged within VLPs assembled from wt CP (a) and mutant R10P CP containing +24 N-terminal charges (b) are indistinguishable. Scale bar shows 50 nm.

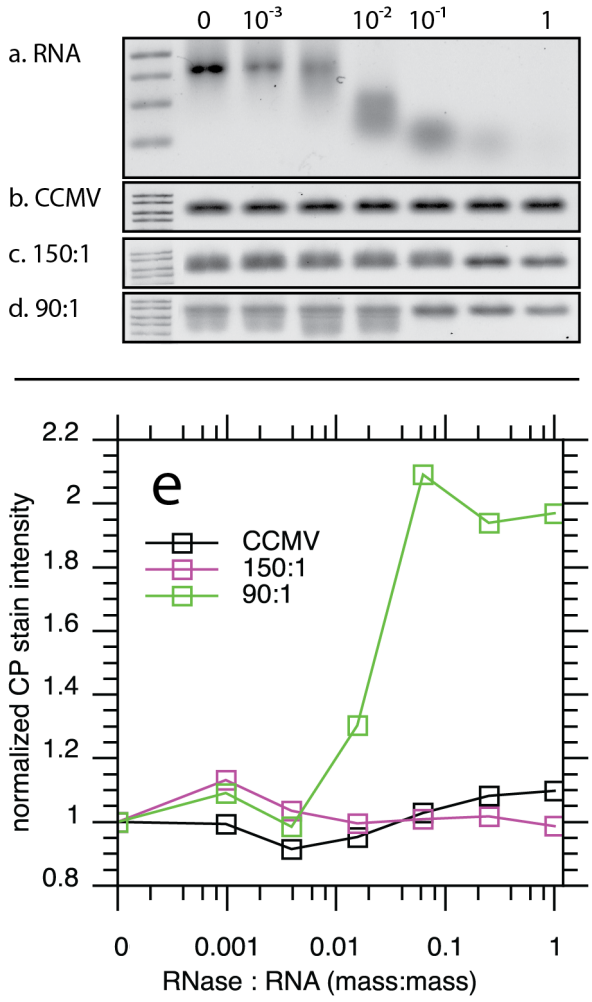

Figure 3-6. RNase A digestion of RNA, wtCCMV, and CP-RNA assembly reactions monitored by native gel electrophoresis and densitometry. Increasing amounts of RNase A were added to samples pre-equilibrated in neutral pH assembly buffer and left on ice for 30 min before electrophoresis. Electrophoresis was carried out in 1% native agarose gels cast in low-pH gel electrophoresis buffer. After electrophoresis, gels were stained with both EtBr and Coomassie Instant Blue protein stain. The mass ratio of RNase:RNA increases from left to right and is denoted above each lane. The left-most lane in each gel contains a dsDNA ladder. (a) RNase digestions of BMV RNA1, stained with EtBr. The dsDNA ladder shows (from top to bottom) 2.0, 1.5, 1.0, 0.5 kbp. (b), (c), and (d) RNase digestions of CCMV, an assembly reaction consisting of

150 CP_2 per RNA , and a similarly prepared assembly consisting of 90 CP_2 per RNA, respectively, also stained with EtBr. The dsDNA ladder shows (from top to bottom) 10.0, 8.0, 6.0, 5.0 kbp. (e) densitometry profiles were generated by plotting the integrated protein stain signal from the capsid band of each lane as a function of RNase:RNA ratio. The plot has been normalized so that the band intensity corresponding to the point containing no RNase is equal to one.

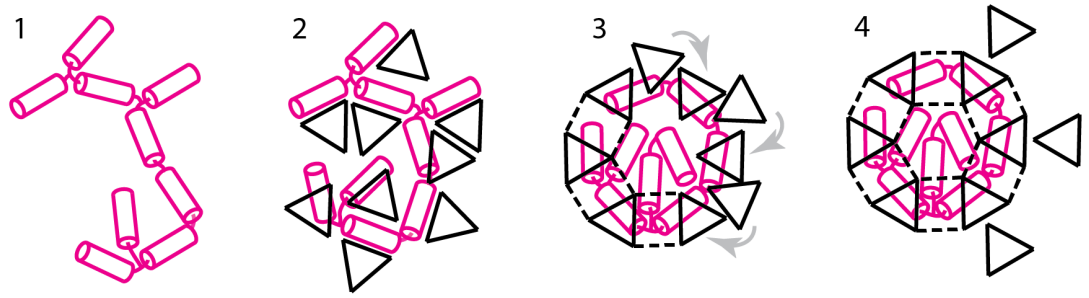

Figure 3-7. RNA is symbolized as a collection of jointed cylinders (1). Each cylinder represents an amount of (negative) charge equal to the amount of (positive) charge brought by the ARMs of each CP dimer, and thus represents a single binding site. CP dimers are represented by black triangles, which saturate the RNA binding sites during the first stage of assembly, slightly compacting the RNA (2). Upon acidification, lateral interactions between CP dimers increase in strength and drive nucleation of the capsid (represented by dashed lines) as well as further reorganization of the RNA (3). Growth of the capsid displaces the excess of bound CP to the exterior surface where it is stabilized by electrostatic attraction between its ARMs and the negative surface charge density of the capsid (4).

	Charge (dimer)	Abbr. Name	N-terminal Sequence
wild-type	+20	CP	**M**STVGTG**K**L TR**A**Q**RR**AAA**R** **K**N**KR**NT**R**VVQ
R13P/R14G	+16	mCP16	**M**STVGTG**K**L TR**A**QPGAA**R** **K**N**KR**NT**R**VVQ
R10P	+24	mCP24	**M**STVGTG**K**L TPAQ**RR**AG**K**L TPAQ**RR**AAA**R K …**

Table 1. The N-terminal sequences of the three CP variants are shown. Cationic residues are labeled in bold face and the 10 residue extension that has been inserted to generate the mCP24 variant is underlined.

REFERENCES

1. Hohn, T. 1976. Packaging of Genomes in Bacteriophages: A Comparison of ssRNA Bacteriophages and dsDNA Bactgeriophages. Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 276:143-150.
2. Belyi, V. A., and M. Muthukumar. 2006. Electrostatic origin of the genome packing in viruses. Proc. Natl. Acad. Sci. USA 103:17174-8.
3. Larson, S. B., R. W. Lucas, and A. McPherson. 2005. Crystallographic structure of the T=1 particle of brome mosaic virus. J. Mol. Biol. 346:815-31.
4. Rao, A. L., and G. L. Grantham. 1995. Biological significance of the seven amino-terminal basic residues of brome mosaic virus coat protein. Virology 211:42-52.
5. Schmitz, I., and A. L. Rao. 1998. Deletions in the conserved amino-terminal basic arm of cucumber mosaic virus coat protein disrupt virion assembly but do not abolish infectivity and cell-to-cell movement. Virology 248:323-31.
6. Lavelle, L., M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres, and J. Ruiz-Garcia. 2009. Phase diagram of self-assembled viral capsid protein polymorphs. J. Phys. Chem. B 113:3813-3819.
7. Hohn, T. 1969. Role of RNA in the assembly process of bacteriophage fr. J. Mol. Biol. 43:191-200.
8. Pfeiffer, P., and L. Hirth. 1974. Aggregation states of brome mosaic virus protein. Virology 61:160-7.
9. van der Schoot, P., and R. Bruinsma. 2005. Electrostatics and the assembly of an RNA virus. Phys. Rev. E 71:061928.

10. Ting, C., J. Wu, and Z.-G. Wang. 2011. Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. Proc. Natl. Acad. Sci. USA 108:16986-16991.

11. Perlmutter, J. D., C. Qiao, and M. F. Hagan. 2013. Viral genome structures are optimal for capsid assembly. eLife:http://arxiv.org/abs/1305.3556.

12. Erdemci-Tandogan, G., J. Wagner, P. v. d. Schoot, R. Podgornik, and R. Zandi. 2014. RNA topology remolds electrostatic stabilization of viruses. Phys. Rev. E 89:032707.

13. Elrad, O. M., and M. F. Hagan. 2010. Encapsulation of a polymer by an icosahedral virus. Phys. Biol. 7:045003.

14. Zhang, R., E. Wernersson, and P. Linse. 2013. Icosahedral capsid formation by capsomer subunits and a semiflexible polyion. RSC Advances.

15. Zhang, R., and P. Linse. 2013. Icosahedral capsid formation by capsomers and short polyions. J. Chem. Phys. 138:154901/1-14.

16. Mahalik, J. P., and M. Muthukumar. 2012. Langevin dynamics simulation of polymer-assisted virus-like assembly. J. Chem. Phys. 136:135101.

17. Forrey, C., and M. Muthukumar. 2009. Electrostatics of capsid-induced viral RNA organization. J. Chem. Phys. 131:105101.

18. Cadena-Nava, R. D., M. Comas-Garcia, R. F. Garmann, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart. 2012. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J. Virol. 86:3318-3326.

19. Garmann, R. F., M. Comas-Garcia, A. Gopal, C. M. Knobler, and W. M. Gelbart. 2014. The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. J. Mol. Biol. 426 1050–1060.

20. Bancroft, J. B. 1970. The self-assembly of spherical plant viruses. Adv. Virus Res. 16:99-134.

21. Annamalai, P., and A. Rao. 2005. Dispensability of 3′ tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. Virology 332:650-658.

22. Dreher, T. W., A. L. Rao, and T. C. Hall. 1989. Replication in vivo of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. 206:425-438.

23. Adolph, K., and P. Butler. 1977. Studies on the assembly of a spherical plant virus III. Reassembly of infectious virus under mild conditions. J. Mol. Biol. 109:345-357.

24. Comas-Garcia, M., R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, and W. M. Gelbart. 2012. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. J. Virol. 86:12271-12282.

25. Johnson, J. M., D. A. Willits, M. J. Young, and A. Zlotnick. 2004. Interaction with capsid protein alters RNA structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus. J. Mol. Biol. 335:455-464.

26. Physical arguments based on the entropy gain associated with mobile counterion release upon polycation/nucleic acid binding estimate the free energy benefit to be roughly kT per charge. Since there are +20 charges on the CP dimer N-termini, we expect binding energies on the order of 20 kT per CP.

27. Choi, Y. G., and A. L. N. Rao. 2000. Molecular studies on bromovirus capsid protein. VII. Selective packaging on BMV RNA4 by specific N-terminal arginine residuals. Virology 275:207-217.

28. Johnson, J. M., J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick. 2005. Regulating self-assembly of spherical oligomers. Nano Lett. 5:765-770.

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

30. Hiebert, E., and J. B. Bancroft. 1969. Factors affecting the assembly of some spherical viruses. Virology 39:296-311.

31. Prinsen, P., P. van der Schoot, W. M. Gelbart, and C. M. Knobler. 2010. Multishell structures of virus coat proteins. J. Phys. Chem. B 114:5522-33.

32. Ni, P., Z. Wang, X. Ma, N. C. Das, P. Sokol, W. Chiu, B. Dragnea, M. Hagan, and C. C. Kao. 2012. An examination of the electrostatic interactions between the N-terminal tail of the Brome Mosaic Virus coat protein and encapsidated RNAs. J. Mol. Biol. 419:284-300.

33. McPherson, A. 2005. Micelle formation and crystallization as paradigms for virus assembly. Bioessays 27:447-458.

34. Devkota, B., A. S. Petrov, S. Lemieux, M. B. Boz, L. Tang, A. Schneemann, J. E. Johnson, and S. C. Harvey. 2009. Structural and Electrostatic Characterization of Pariacoto Virus: Implications for Viral Assembly. Biopolymers 91:530-538.

CHAPTER 4

Exploiting Fluorescent Polymers to Probe the Self-Assembly of Virus-Like Particles

ABSTRACT The inside surfaces of the protein shells of many viruses are positively charged, thereby enhancing the self-assembly of capsid proteins around their (oppositely-charged) RNA genome. These proteins have been shown to organize similarly around a variety of non-biological, negatively-charged, polymers, e.g., poly(styrene sulfonate) (PSS), forming virus-like particles (VLPs). We have demonstrated recently that the VLPs formed from cowpea chlorotic mottle virus (CCMV) capsid protein increase in size (from T=2 to T=3 structures) upon increase in PSS molecular weight (from 400 kDa to 3.4 MDa), and that the total charge on the PSS exceeds that of the capsid protein by as much as a factor of 3. Here we extend studies of this kind to PSS molecules that are sufficiently small that two or more can be packaged into VLPs. The use of 38 kDa PSS polymers that have been fluorescently labeled with Rhodamine B allows us to determine the number of PSS molecules per capsid. Electron micrographs of the VLPs show a bimodal distribution of particle diameters, with one peak centered around 19 nm, typical of a T=1 triangulation number, and the other around 21 nm, consistent with a pseudo T=2 structure; increasing the molar ratio of protein to PSS in the reaction mix shifts the VLP distribution from T=1 to T=2 structures. By combining fluorescence and gel electrophoresis measurements it is determined that, on average, there are two polymers in each T=1 capsid and three in each T=2, with the PSS charge less than that of the capsid protein by as much as a factor of 2. VLPs of this kind provide a versatile model system for determining the prinicples underlying self-assembly of controlled numbers of cargo molecules in nanocontainers of increasing size.

INTRODUCTION

Several strategies have been employed to generate fluorescently labeled and potentially addressable biomolecular nano-containers. For example, fluorescently labeled poly(styrene sulfonate) (PSS) has been incorporated into colloidal particles 8 -10 μm in diameter that have been prepared by layer-by-layer assembly onto latex particles that can then be decomposed to leave hollow shells[1]. More recently, fluorescent RNA has been synthesized from DNA templates within 4 μm poly(methacrylic acid) hydrogel capsules that have been assembled layer-by-layer.[2] Self-assembly of viral capsids has also been employed. Minten, et al.[3] packaged enhanced green fluorescent protein (EGFP) in Cowpea Chlorotic Mottle Virus (CCMV) capsids by the self-assembly of mixtures of EGFP-CCMV capsid protein complexes with pure capsid protein. The EGFPs are located inside the 28-nm diameter capsids that form, with the number incorporated depending on the ratio of EGFP-capsid protein complex to pure capsid protein in the reaction mix.

It has long been known that CCMV capsid protein can form shells of different sizes associated with different Caspar-Klug triangulation numbers[4] and that it can package a variety of anionic polymers[5]. In particular, in recent studies we have demonstrated that PSS molecules with molecular weights 400 kDa to 3.4 MDa can be incorporated into CCMV capsids[6]. We observed that the polymers with molecular weight of 1 MDa and lower were packaged in capsids 22-23 nm in diameter, consistent with a "T=2" triangulation number[7]. Such capsids contain 120 proteins organized into 12 pentamers of dimers. The capsids that self-assembled around 2.0- and 3.4-MDa polymer had diameters of 27-28 nm, consistent with T=3 structures that contain 180 proteins arranged in 12 pentamers and 20 hexamers. We argued that the variations in capsid size with the length of the polymer are consistent with the capsid size being determined by a

balance between the preferred radius of curvature of the protein shell and the polymer interactions with itself and with the charges on the capsid interior[6,8,9]. This suggested that with a sufficiently short polymer, capsids with a T=1 triangulation number composed of 60 proteins (12 pentamers) and having a diameter of about 18 nm could form.

Indeed, Sikkema, et al.[10] found that CCMV protein self-assembled around a 9.9 kDa PSS labeled with a dansyl chloride fluorophore. The capsid diameter, as determined by light scattering and electron microscopy, was about 16 nm, consistent with capsids having a T=1 triangulation number. But they attributed the formation of the small capsids to the positively-charged N-terminus of the capsid protein having been inadvertently cleaved, which had been shown earlier[7] to lead to T=1 structures for empty capsids. Minten, et al.[1] also reported the formation of T=1 capsids, for the packaging of poly(ferrocenylsilane) by CCMV protein. While CCMV proteins lacking the N-terminus are known to assemble into empty capsids, the packaging of an anionic species is thought to require its presence. Sikkema, et al. suggested that the packaging of the PSS in the absence of the positively charged protein tails might be due to the presence of a small amount of the full-length protein. (Minten, et al. do not report the purity of the protein they employed.) The relation between protein cleavage and the formation of T=1 capsids has been well established for a number of other viruses[11,12,13,14]. The formation of T=1 capsids from uncleaved CCMV has been demonstrated in experiments in which the self-assembly was carried out around double-stranded DNA oligomers[15]. In this case, however, the 17-nm diameter capsids budded off from the ends of 17-nm-diameter closed tubes. A related virus, brome mosaic virus, has been found to self assemble around 6-nm diameter gold particles to form VLPs corresponding to T=1 capsids[16].

In the present paper we demonstrate that T=1 VLPs can be formed by CCMV capsid proteins whose N-termini are intact, i.e., uncleaved. More importantly, we extend our earlier work on VLP formation from PSS and CCMV capsid protein by working with PSS of much lower molecular weight. We find bimodal distributions of VLP sizes that involve T=1 and T=2 particles. Their relative numbers are shown to depend on the molar ratio of capsid protein (CP) to PSS, ranging from predominantly (90%) T=1 at CP:PSS = 4.2 to around 70% for CP:PSS ratios of 34 and higher. In our previous studies of the packaging of PSS[6] we noted that we had begun to examine the packaging of PSS labeled with Rhodamine B (PSS-Rh) with the aim of demonstrating that the polymer was not adsorbed on the *outer* capsid surface, where its fluorescence could be diminished by the action of a quencher molecule. This work is reported here, but with the additional purpose of facilitating the determination of the number of PSS molecules packaged by the capsid protein. We also examine the dependence of capsid size on the CP:PSS molar ratio and report on the broad range of values found (upon change in polymer molecular weight) for the ratio of anionic-to-cationic charge brought to the VLPs by the polymer and capsid protein.

RESULTS AND DISCUSSION

In order to determine the optimal conditions for assembly, we carried out a series of studies in which the ratio of protein to PSS-Rh was varied. In these experiments, assemblies were performed with a fixed weight of CP (18 μg) and with increasing weights of PSS-Rh ranging from 0.2 to 8 μg, corresponding to CP:PSS molar ratios ranging from approximately 170 to 4.2. The assembly products were examined by gel electrophoresis and sedimentation in a sucrose gradient, from which it was determined that the optimal assembly occurs for a molar

ratio of approximately 34 CP per PSS-Rh. This is the minimum ratio for which a single band is observed when the assembly mixture is run on a sucrose gradient and at which only a strong single fluorescent band is visible in the gel.

A typical TEM image of the products of assembly at the optimal ratio is shown in Figure 4-1 and a histogram of the VLP sizes is shown in Figure 4-2a; for comparison a histogram of a micrograph for wild-type CCMV virions is also shown. It is evident from the histogram that there are two populations of VLP capsids, one centered around a diameter of 19 nm and the other around 21 nm, distinctly smaller in size than the (26 nm) T = 3 capsids for wild-type CCMV. It is reasonable to assume that the VLP capsids have triangulation numbers T = 1 and T = 2, although rigorous assignment of T numbers would require a structural determination. A least-squares fit of two Gaussian peaks to the histogram shows that the 72 % of the particles have T = 1 capsids and 28% T = 2 capsids. The fraction of T=2 capsids drops by a factor of 3 (to about 10%) when the CP:PSS molar ratio is decreased from 34 to 4.2, as shown by a similar fit to the lower histogram in Figure 4-2b.

TOF-MALDI mass spectrometer (Supporting Information Figure 4-S1) and SDS-PAGE studies of the capsid protein carried out just before assembly demonstrate (Figure 4-S2) that its N-terminus was fully intact, and SDS-PAGE analyses of disassembled VLPs establishes that no cleavage takes place during assembly, confirming that the formation of T = 1 VLPs can occur without protein cleavage.

In principle, the average number of PSS-Rh per capsid could be determined from measurements of UV-Vis spectra. But the weak characteristic absorptions for the PSS-Rh and its low concentration make this impossible. We can instead obtain an estimate of the number of

polymer molecules encapsidated by determining the PSS-Rh concentrations from fluorescence and the number of capsid proteins by gel electrophoresis.

Fluorescence spectra of free PSS-Rh at different concentrations in Virus Buffer (100 mM Sodium Acetate, 1 mM EDTA, pH 4.8) are shown in Figure 4-3a; excitation was at 525 nm. Also shown is the fluorescence spectrum of a solution of PSS-Rh VLPs, showing the same shape but with a maximum blue-shifted by 3 nm. Using the observed linear dependence of the fluorescence of PSS-Rh on concentration, we can determine the average concentration of the polymer in the solution of VLPs, under the assumption that there is no quenching (or enhancement) of the fluorescence by interactions with the assembled capsid protein. This assumption has been tested by following the variation of the fluorescence with time as the capsids are disassembled by the addition of 5 M urea[17]. As shown in Figure 4-4, the emission intensity of the enscapidated polymer remains essentially constant, increasing by only about 5% over the course (80 min) of capsid disassembly.

Given the effective concentration of the encapsidated polymer in solution as determined by the fluorescence intensity – see Figure 4-3a – it is possible to estimate the average number of polymers per capsid from a knowledge of the number of capsids in the solution and the distribution of capsid sizes. The number of capsids can be obtained by gel electrophoresis and the size distribution from TEM as described above. A measured amount of the VLP solution is run on an SDS-PAGE gel and the protein in the resulting band is stained with Coomassie Blue. The protein concentration is then determined from a calibration of densitometer tracings of bands obtained from capsid protein solutions of known concentration, as shown in Figure 4-3b. (Comparisons of gels of solutions of known concentrations of CCMV and those of capsid proteins confirm that the staining is identical.) By combining the results of these gel stain

measurements with the TEM and fluorescence data previously described, we find that that there are on average 2 PSS-Rh polymers in the T=1 capsids and 3 copies in the T=2 capsids.

The degree to which the PSS-Rh in the VLPs is protected by the protein capsid can be assessed by examining the change in fluorescence emission as a result of adding a quencher to the surrounding solution. We have carried out a series of experiments in which the VLPs and unencapsidated PSS-Rh, separately, were mixed with the fluorescence quencher methyl viologen (MV). Fluorescence spectra of solutions of free PSS-Rh in virus buffer at a concentration of 3.3 μg/mL in the presence and absence of quencher are shown in Figure 4-5. The addition of MV at a concentration of 1 mM reduced the fluorescence intensity by 75%. In contrast, while the emission from encapsidated PSS-Rh under the same conditions shows a decrease, it levels off at 5% after about 1200 s. No change in fluorescence was observed when free PSS-Rh was mixed with intact VLP capsids, nor is the quenching of free PSS-Rh by MV affected by the presence of intact VLPs. Hence, formation of VLPs protects the PSS-Rh against quenching by MV, with no effects due to direct interaction with capsid protein and only negligible polymer adsorbed on the capsid surface.

It is tempting to argue that the MV is ineffective in quenching polymer fluorescence in VLPs because the doubly-charged cation is repelled by the positive charge on the protein tails and/or is too big for the holes in the capsid and is therefore unable to enter the VLP. Other observations suggest, however, that this interpretation may not correct. We have regularly used ethidium bromide (EtBr) to stain RNA and short ds DNA oligos inside T = 3 and T =2 capsids, and Zlotnick has reported[18] the staining of DNA-oligo-containing T = 1 capsids. It follows, from the fact that the ethidium cation fluoresces only when it intercalates into double-stranded regions of nucleic acid, that it enters the capsid. From space-filling molecular models it is evident that

the ethidium cation is significantly larger than MV and it therefore is unlikely that the capsid is porous to EtBr and not to MV. It thus appears that the ineffectiveness of MV as a quencher of fluorescence in the VLPs is caused by the ionic environment of the polymer within the capsid rather than by the quencher not being able to permeate the capsid. Measurements of the efficacy of quenching of PSS-Rh by MV in virus buffer as a function of the concentration of Mg^{+2} (Figure 4-6) show that it decreases upon addition of Mg^{+2} and that at 500 mM the MV reduces the fluorescence by only 25% rather than by 75% as in the absence of Mg^{+2}. This suggests that locally-high ionic strength in the capsid, may account for the small degree of quenching observed for the VLPs. More explicitly, the N-terminus of each of the CCMV capsid proteins, which has 10 positively charged amino acid residues, protrudes into the VLP interior and interacts strongly with encapsidated polymers, which are concentrated in a region near the capsid interior surface. Thus, the PSS-Rh is in an environment with an effectively high ionic strength that is likely the cause of the ineffectiveness of the MV as a quencher., although we cannot demonstrate conclusively that other effects such as differences in polarity of the environment or conformation of the polymer play a role.

CONCLUSIONS

The 600 (1200) positive charges on the interior of T = 1 (T=2) capsids made up of CCMV protein, associated with the 60 (120) proteins each with 10 + charges per N-terminus, are relevant to other physical features of the self-assembled VLPs, in particular to the ratio of anionic-to-cationic charge. The charge on each encapsidated polymer is only -180, from which it follows that the ratio of negative to positive charge is 0.6 and 0.45 for the T=1 capsids (containing 2 polymers) and the T = 2 capsids (containing 3), respectively. These values are

much lower than the ratio of 1.6 that holds for a large variety of wild-type RNA viruses of different (genome and capsid) sizes[19]; in other terms, there is under-charging, i.e. under compensation of the protein positive charge.

Daniel, et al.[20], who examined the self-assembly of Brome Mosaic Virus capsid proteins around gold particles functionalized with thiolated tetraethylene glycol ligands, found that VLPs that contained 6-nm diameter particles were not stable at low pH and ionic strength and at least partially lose their protein coat. They attributed the lack of stability to the low surface charge on the gold particles, which is only 0.40 that of the positively charged CP tails in a T = 1 capsid. As noted above, in the case of the T = 1 and T = 2 VLPs containing PSS-Rh, the charges on the encapsidated polymer are similarly low compared to those of the CP. However, in contrast to Figure 5 in the Daniel, et al. paper, there is no evidence of particle dissociation in our Figure 4-1, which was made under similar conditions of pH and ionic strength. Moreover, PSS-Rh VLPs have been stored under these conditions for several weeks with no evidence of instability. The lower stability of the VLPs containing the gold nanoparticles may perhaps be attributable to the tethering of the polymers on the gold cores, which will restrict their ability to interact with the CP tails. It cannot be excluded that despite a high degree of homology, there are differences between BMV and CCMV CPs that affect the VLP stability.

In our previous studies[6] of the packaging of polystyrene sulfonate polymers with molecular weights ranging from 400 kD to 3.4 MDa we found that stable VLPs can be assembled in which the charge ratios well *in excess* of 1.6. It is therefore evident for any given system (e.g., polymer type and capsid protein) a specific charge ratio is not the only criterion for VLP stability.

Finally, we expect that the polymorphism of VLPs, as a function of polymer length and molar ratio of capsid protein to polymer, is a general feature that follows from the competition between the preferred sizes of the polymer and capsid. In the present case, the co-existence of both T=1 and T=2 capsids reflects the closeness in their stability[6], with the law-of-mass-action determining the percentage of each form in response to changes in the relative amount of protein and polymer[21].

METHODS

CCMV CP Purification

CCMV is harvested and purified from infected California cowpea plant (*Vigna ungiculata*) following established protocols[6,22]. It is dialyzed against 500 mM Calcium Chloride, 50 mM Tris pH=7.5, 1 mM EDTA pH=8, 1 mM DTT, and 0.5 mM PMSF (Disassembly Buffer) to disrupt the nucleocapsids and precipitate the RNA. This disassembled mixture is then centrifuged at 100,000 RPM for 100 min at 4°C in a Beckman TLA110 rotor to pelletize the RNA and recover the purified capsid protein, which is immediately dialyzed against 1M Sodium Chloride, 20 mM Tris pH=7.2, 1mM EDTA, 1mM DTT and 1mM PMSF (Protein Buffer). The protein concentration and its purity against RNA contamination are checked by UV-Vis spectrophotometry – only protein solutions with 280/260 ratios greater than 1.5 are used for assembly. SDS-PAGE gel electrophoresis and mass spectrometry are used to ascertain that the purified protein is not cleaved.

Synthesis and characterization of PSS-Rh

The PSS-Rh is a homopolymer containing 5% Rh. The detailed synthesis protocol for its synthesis has been given elsewhere[23]. Briefly, sodium salt of styrene sulfonic acid (Sigma Aldrich, Australia) is copolymerized with methacryloxyethyl thiocarbamoyl rhodamine B (Polysciences Inc., Warrington, PA) via the reversible addition fragmentation polymerization technique using 4-cyanopentanoic acid dithiobenzoate as a chain-transfer reagent and 4,4. azobis-4-cyanopentanoic acid as an initiator. After extensive dialysis against bi-distilled water, the polymer is recovered by freeze-drying. Its molecular weight (38 kDa) and polydispersity index (1.17) were determined using aqueous gel permeation chromatography with PSS standards.

VLP Assembly

In the assembly studies a fixed amount of purified capsid protein was combined with different amounts of PSS-RH in a total volume of 100μL Protein buffer and dialyzed against 50 mM Tris pH=7.2, 50 mM Sodium Chloride, 10 mM Potassium Chloride, 5 mM Magnesium Chloride, and 1mM DTT (virus assembly buffer) in the dark overnight. The products of an assembly are separated in a sucrose gradient in virus assembly buffer. The VLPs form a sharp fluorescent band and are easily extracted. The extracted band is dialyzed against 100 mM Sodium Acetate, 1 mM EDTA, pH 4.8 (Virus Buffer) to remove the sucrose and the VLPs are then concentrated using Amicon Ultra 500 μL filters of 100 kDa pore size.

TEM imaging of VLPs

Commercial 400-mesh copper grids (Ted Pella, Redding, CA) were modified by first coating the grids with Parlodion, followed by vacuum deposition of a thin layer of carbon. They were glow-discharged in vacuum to ensure good sample spreading. Aliquots of the final assembly products are added to the grids, stained with 1% uranyl acetate, and imaged using a JEM 1200-EX transmission electron microscope equipped with a wide angle (top mount) BioScan 600W 1x1K pixel digital camera. The microscope is operated at 80keV. Images are analyzed by recording two orthogonal diameter measurements of each VLP – analysis is carried out with the help of ImageJ (U. S. National Institutes of Health) software. The geometric mean of these orthogonal measurements is displayed in histograms with 0.4 nm diameter bins.

ACKNOWLEDGEMENTS

We thank Professor Frank Caruso for his assistance in obtaining the labeled polymer, Adriana Anivitarte for carrying out some of the measurements of size distributions from electron micrographs, Professor Joseph Loo for allowing us access to the TOF-MALDI instrument and Dr. Sergey Ryazantsev for his assistance with the TEM imaging. This work was supported by the U.S. National Science Foundation (Grant CHE 0714411, to WMG and CMK). RCN acknowledges partial support from CONACyT and UCMexus.

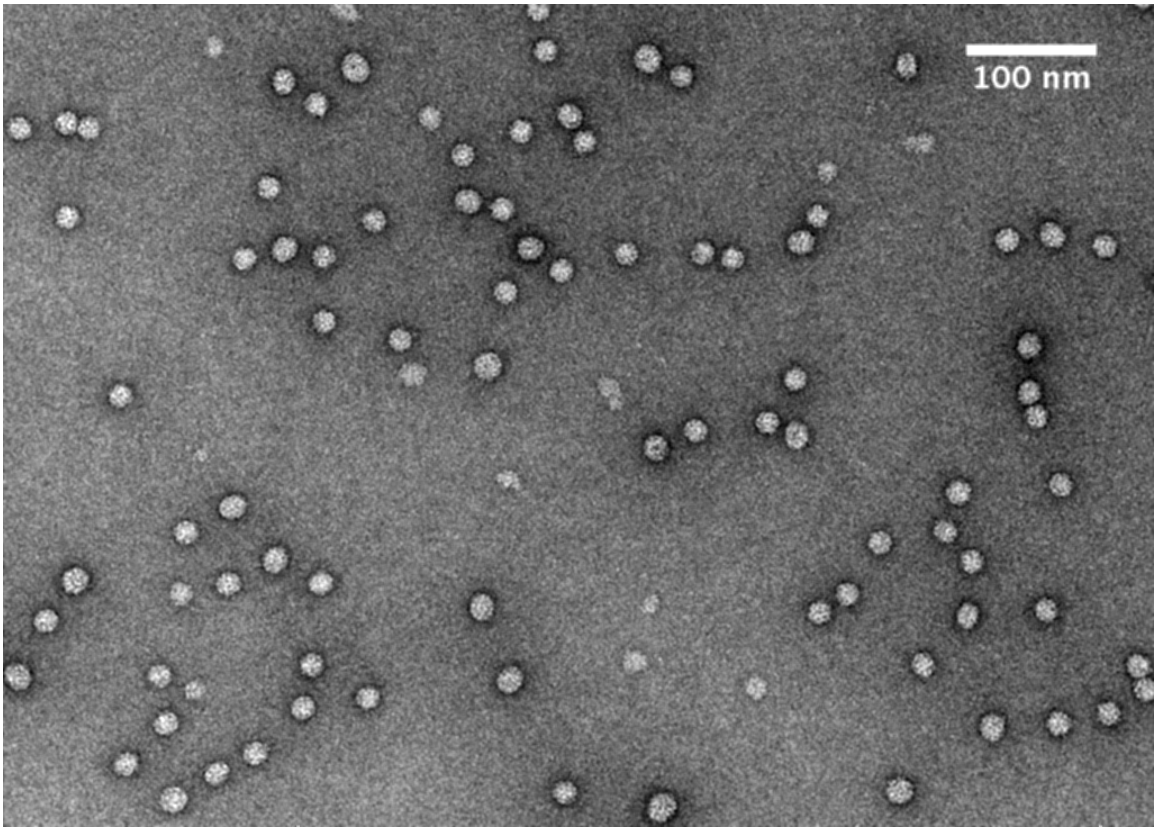

Figure 4-1. PSS-Rh VLPs, formed in a reaction mix with a CP:PSS ratio of 34:1, visualized by negative-stain TEM. Two populations of particles are present, those with a diameter of about 19nm (T=1) and those with diameters of about 21 nm ("T=2").

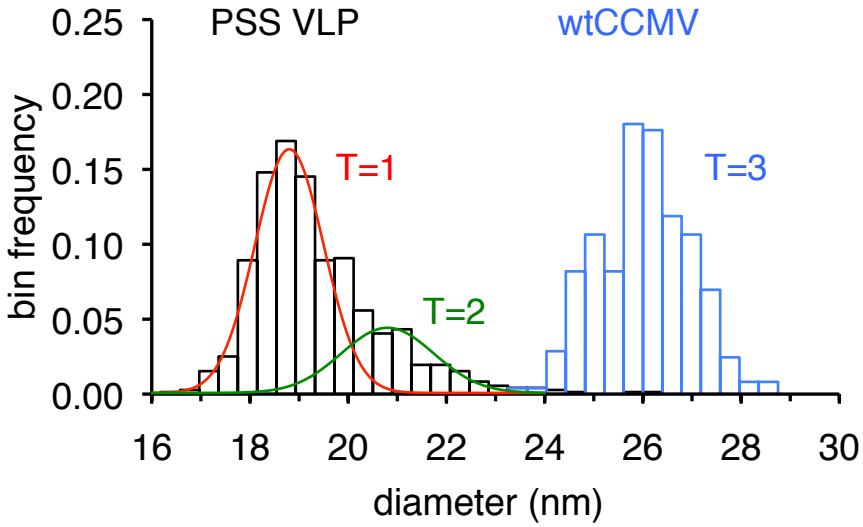

Figure 4-2a. Diameter histogram of PSS-Rh VLPs (in black) assembled using a molar ratio of 34 CP to 1PSS-Rh. The distribution has been determined from images of 900 capsids. A pair of best-fit Gaussian curves, centered at 18.8 nm and 20.8 nm, have been plotted to quantify the relative numbers of T=1 (72%) and "T=2" (28%) particles. Measurements of wt CCMV are plotted in blue as a reference.

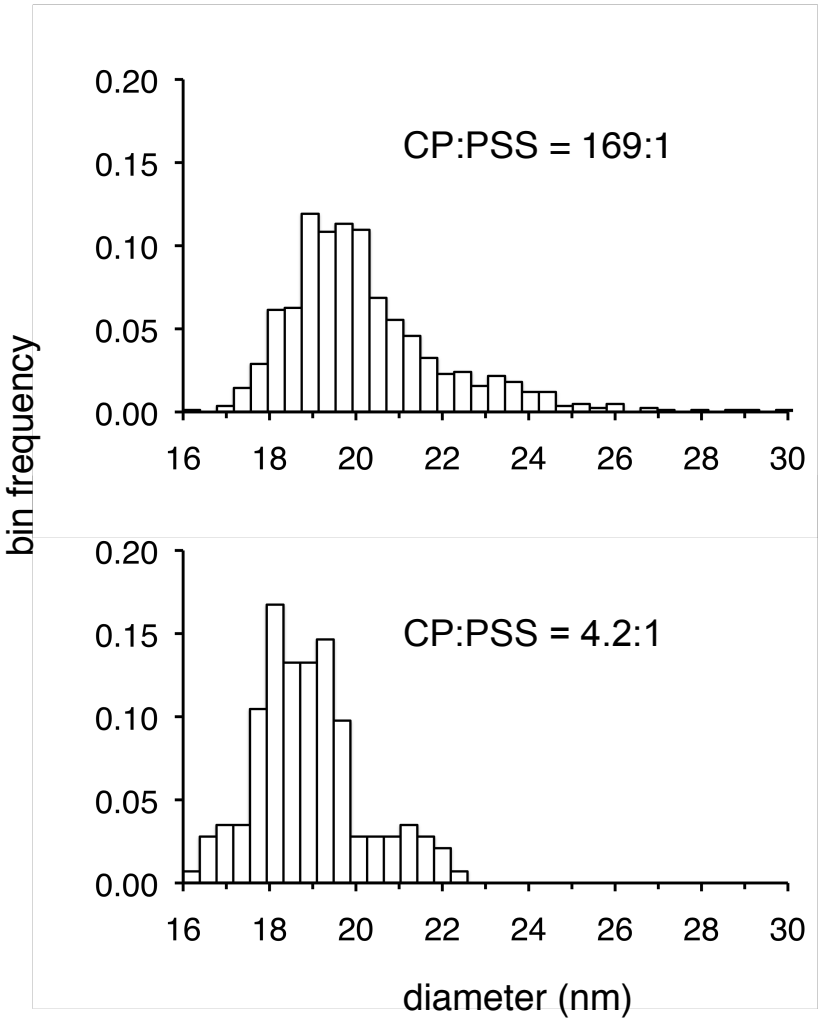

Figure 4-2b. Size distribution histograms for PSS-Rh VLPs assembled in reaction mixes with CP:PSS-Rh mole ratios of 169 (top) and 4.2 (bottom). Note the increase in the number of particles with diameters larger than 20 nm as the relative protein concentration increases.

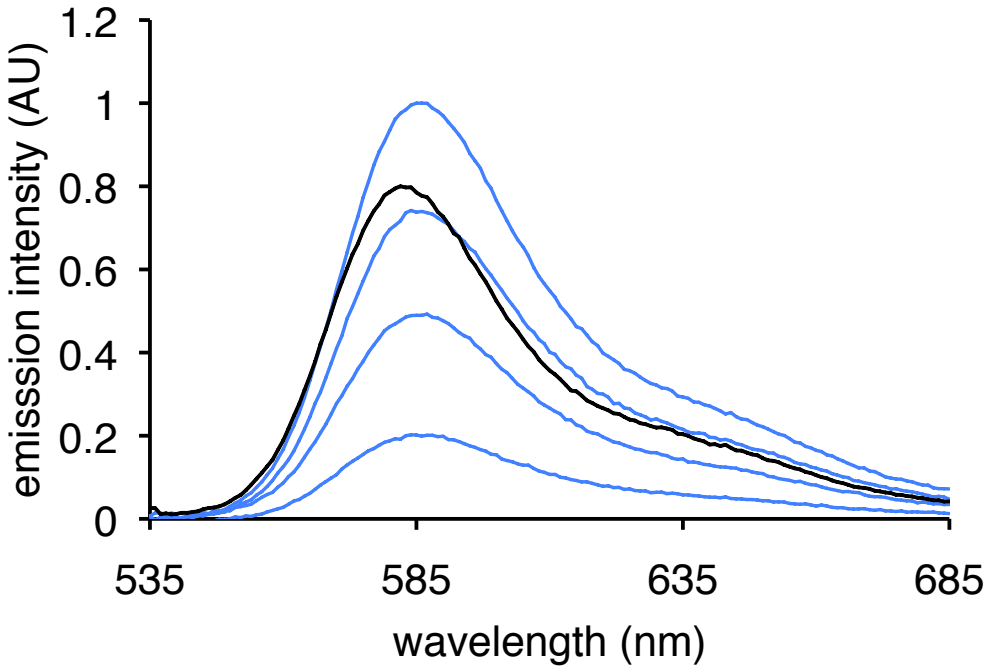

Figure 4-3a. PSS-Rh fluorescence emission spectra of four solutions of free PSS-Rh polymer in Virus Buffer are shown in grey (concentrations of 0.27 pmol/μL [top curve], 0.21 pmol/μL, 0.14 pmol/μL, and 0.07 pmol/μL [bottom curve]). The spectrum of a sample of PSS-Rh VLPs is shown in black. There is a 3-nm blue shift in the peak position for the packaged PSS-Rh.

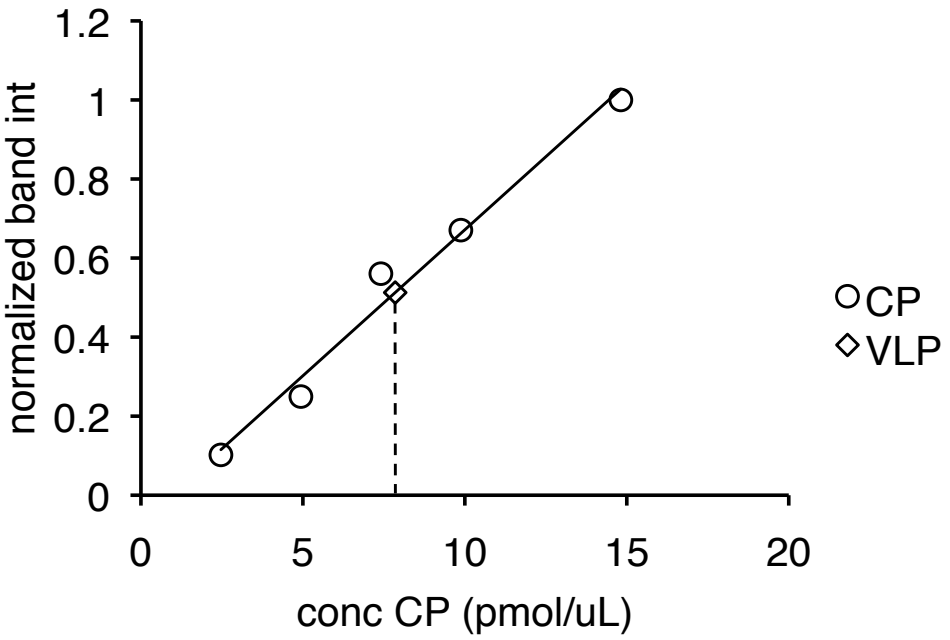

Figure 4-3b. SDS-PAGE stain intensity measurements taken from a set of five concentrations of purified CCMV CP (shown as circles) as well as from a sample of PSS-Rh VLPs (shown as a diamond). The protein was stained with Coomassie blue.

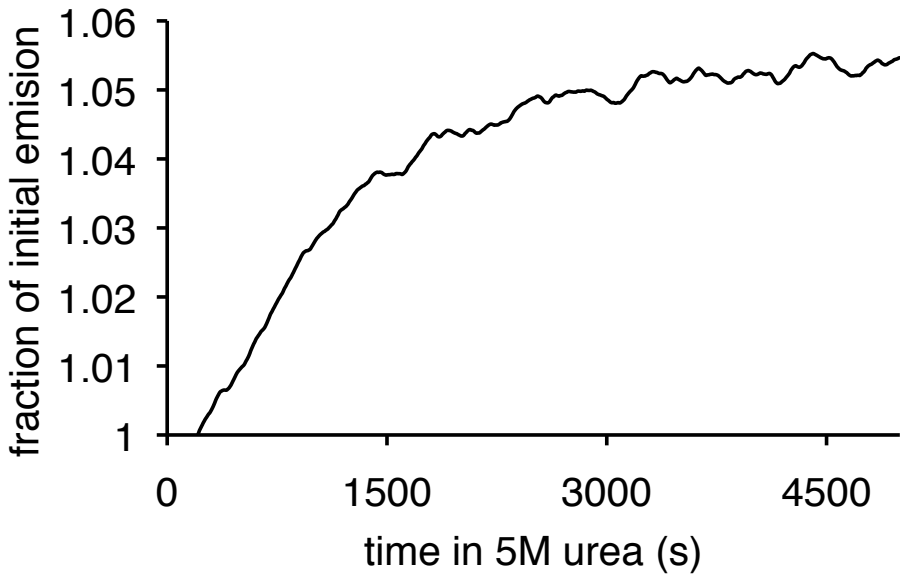

Figure 4-4. Fluorescence emission intensity at 580 nm of a sample of PSS-Rh VLPs as a function of the time after addition of 5M urea. No further increase was observed after 4500 sec. The intensity values are averages of 20 points taken at 4 sec intervals.

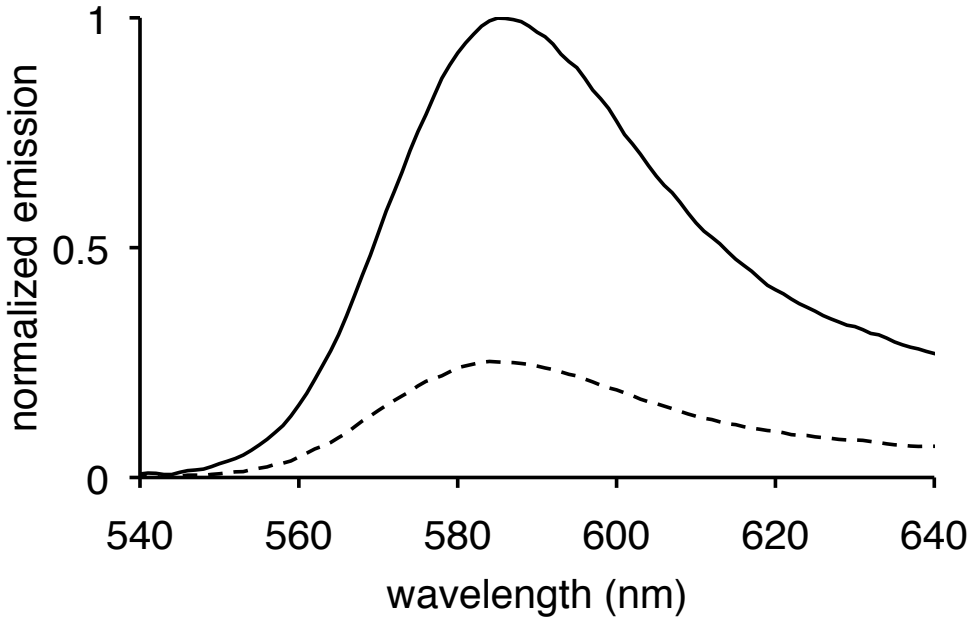

Figure 4-5. The emission spectrum of free PSS-Rh after the addition of 1mM MV (dashed line) shows a 75% reduction in peak intensity from its original value (solid line).

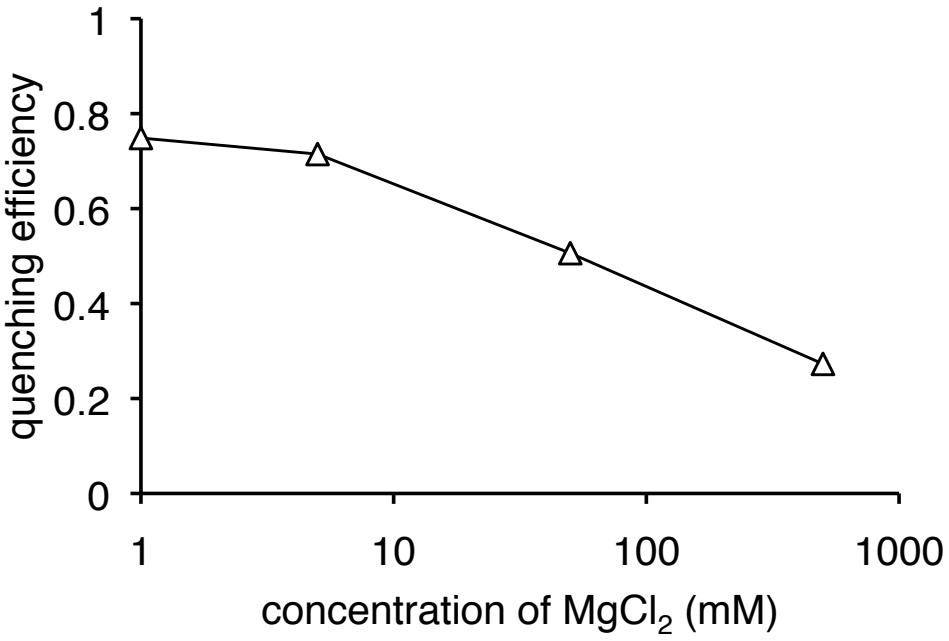

Figure 4-6. Quenching efficiency of free PSS-Rh by MV, as defined by the percentage of emission intensity reduction accompanied by the addition of 1mM MV, varies with the ionic strength of the PSS-Rh solution.

SUPPLEMENTARY FIGURES

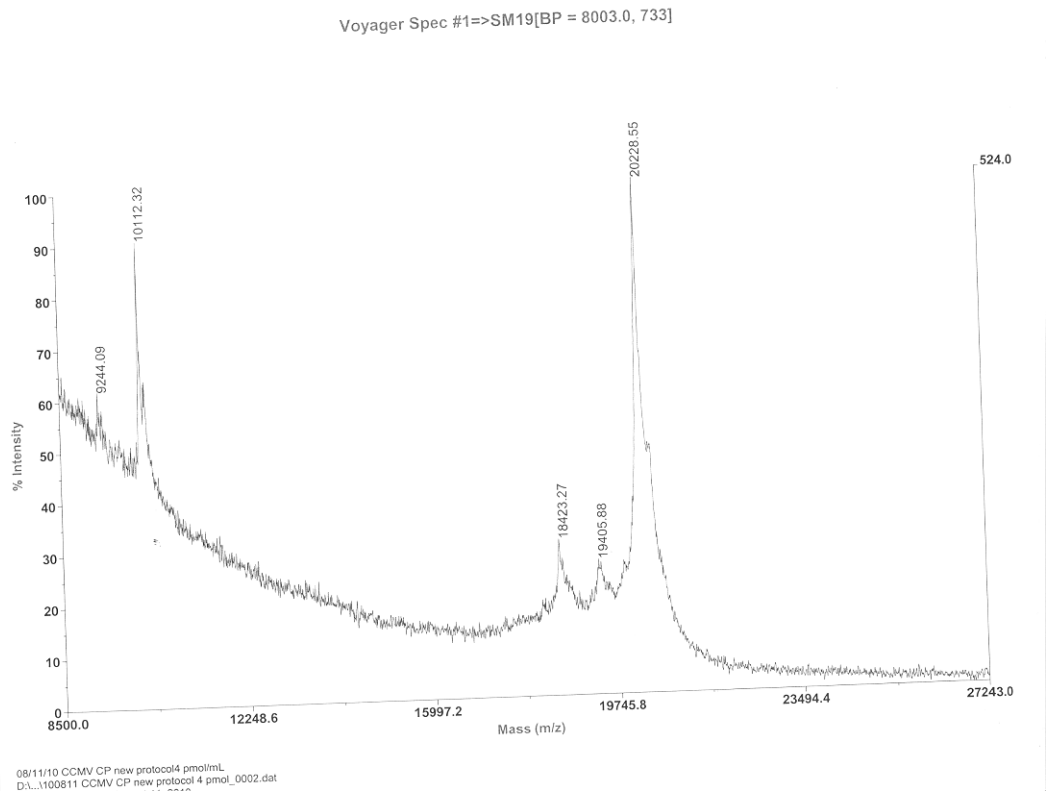

Figure 4-S1A. Demonstration of the intactness of capsid protein used in the assemblies. TOF MALDI MS of the freshly-purified CCMV capsid protein. The peak at m/z= 20228 is that of the uncleaved protein and that at 10112 is attributable to the doubly-ionized protein. Very small amounts of impurities are indicated by the peaks at lower m/z values (e.g., at 18423 and 19405).

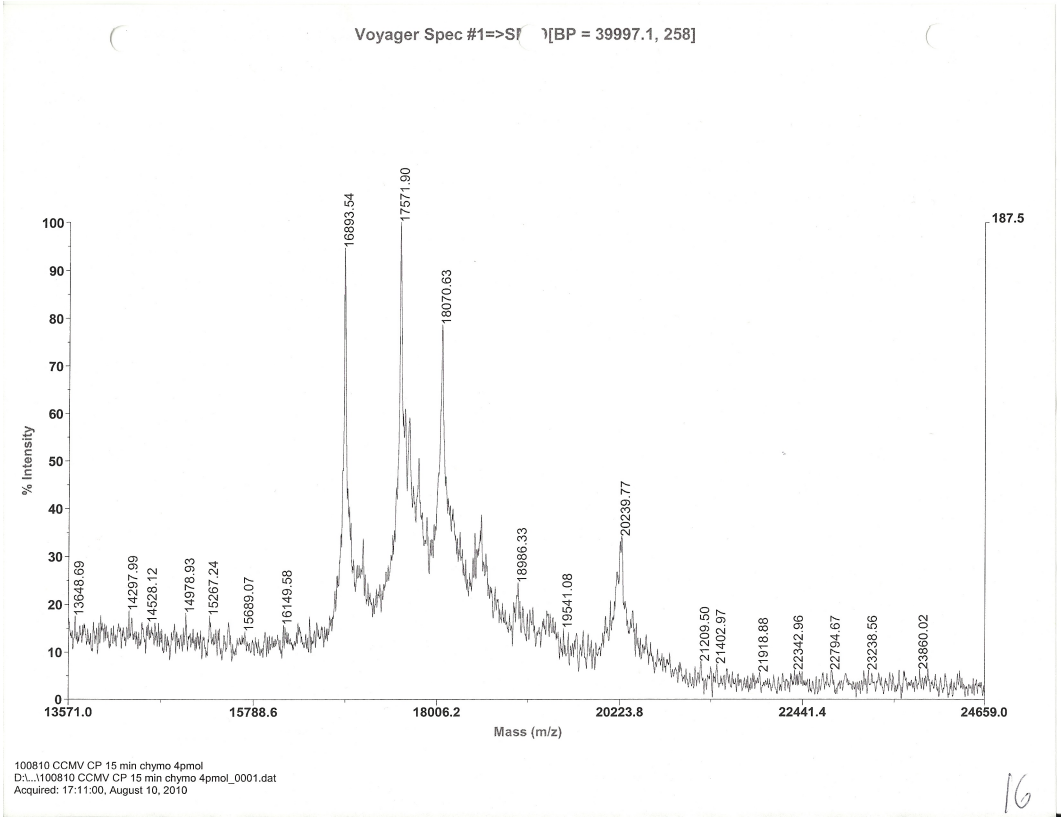

Figure 4-S1B. TOF MALDI MS of purified CCMV capsid protein that has been digested for 15 min with chymotrypsin, an enzyme that cleaves the N terminus of the protein[1]. The major peaks are at m/z values of 18070, 17571 and 16893, and correspond respectively to the loss of 21, 25 and 30 residues from the N-terminus. Note that the peak at 20239.77 corresponds to the small amount of intact protein that survives the enzyme digestion.

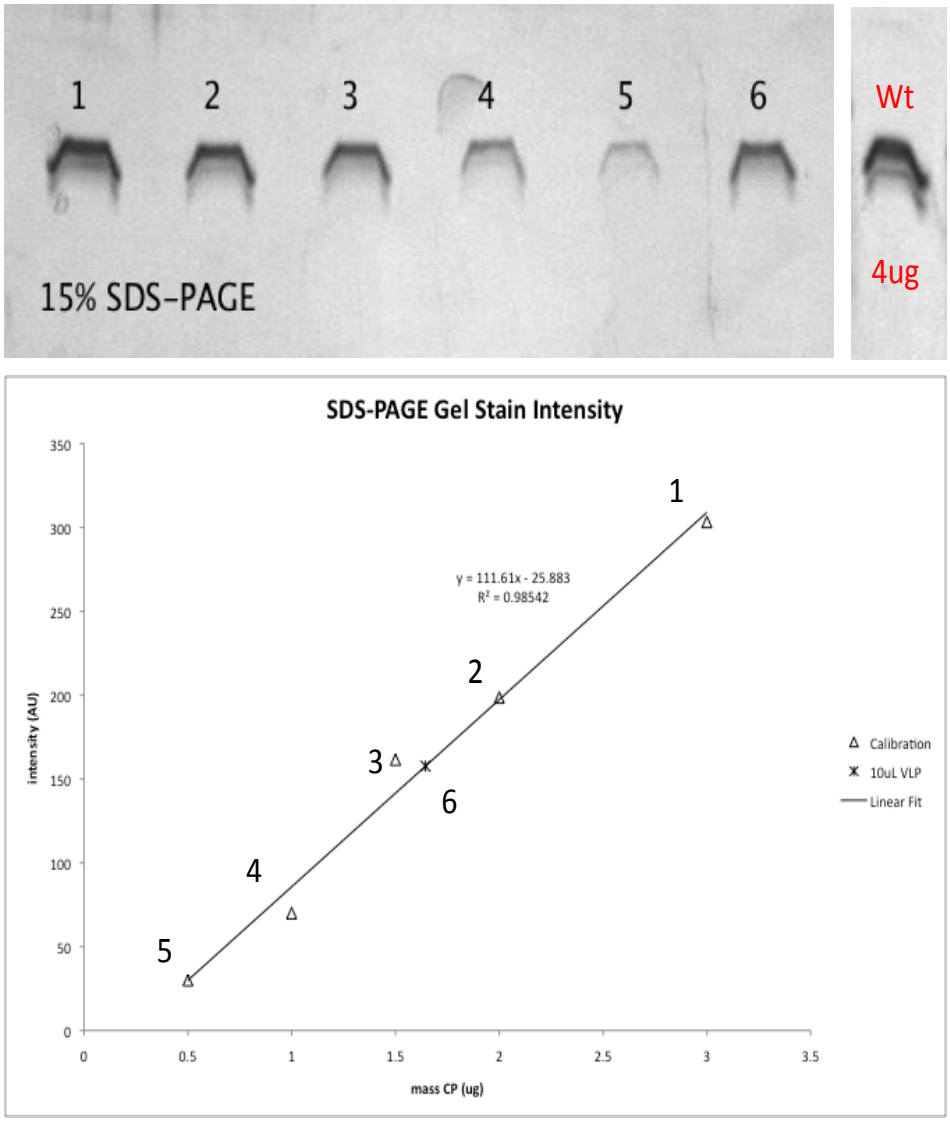

Figure 4-S2A. Demonstration of intactness of capsid protein in VLPs. 15%-SDS-PAGE results for capsid protein. Lane 6 shows the capsid protein obtained from the disassembly of PSS-Rh-containing VLPs by incubating the VLPs for 3 min at 100 °C in a buffer consisting of 50mM Tris (pH 6.8), 100 mM dithiothreitol, 2% SDS. The staining is with Coomassie Blue. Lanes 1-5 show the gel bands for the freshly-purified capsid protein used in the assembly reactions, subjected to the same 3 min/100 ˚C incubation in Tris/DT/SDS solution. The amounts of protein in lanes 1-5 range from 0.5 to 3 µg, determined by UV-Vis absorption. (The plot of gel-stain

intensity vs CP mass included below shows the determination of VLP concentration from these purified protein samples.) For comparison, the right-most lane shows 4 μg capsid protein obtained from the same treatment of wt CCMV. A small amount of a faster-running second component is always found in the disassembly of wt capsids, likely due to the disassembly treatment; a smaller amount is evident as well (see lane 6) in the protein from the VLPs.

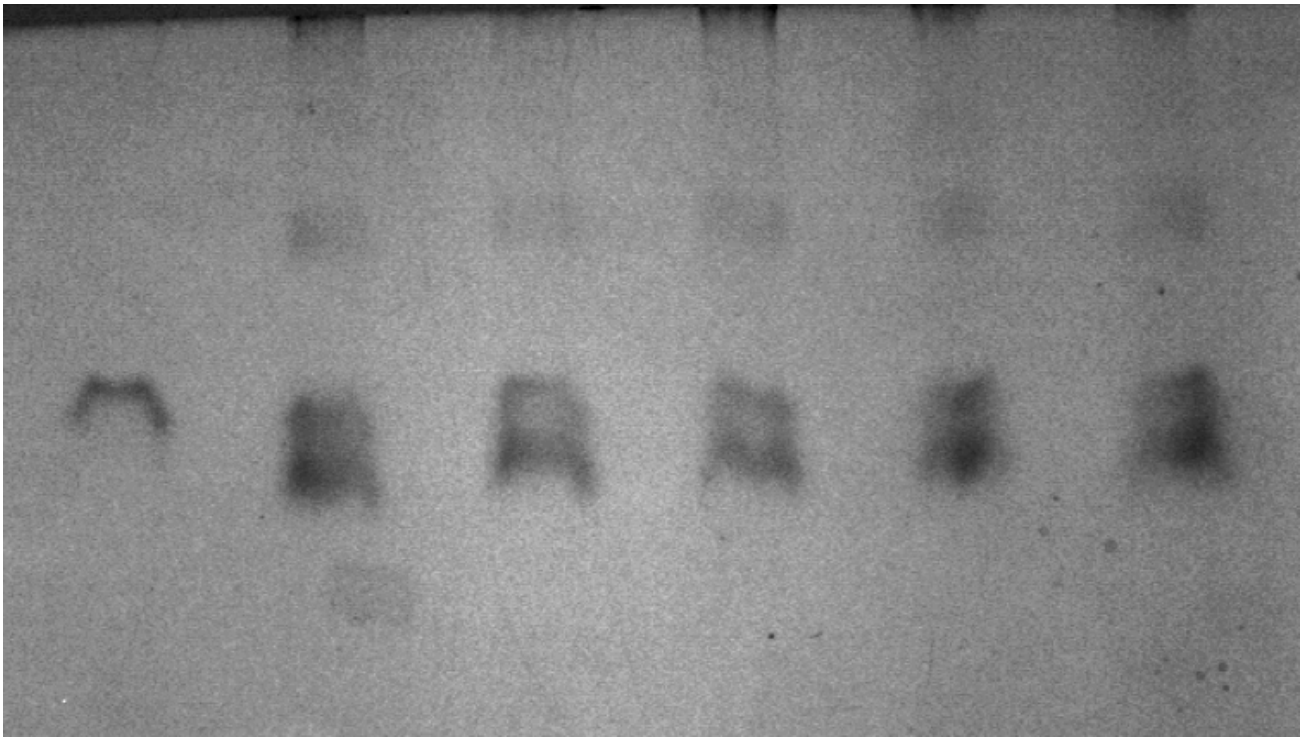

Figure 4-S2B. SDS-PAGE gel of purified capsid protein, both intact and cleaved. The left-most lane is the purified protein used in the assembly, whose TOF-MALDI spectrum is shown in Fig. 4-S1A. The second lane is protein obtained with a commonly-used procedure[2] that involves treatment with PEG 8000, which we have found leads to protein cleavage; this is demonstrated by the other lanes, which show the intact protein (obtained from the purification procedure that does *not* involve PEG) after digestion with chymotrypsin for periods ranging from 15 min (3rd

lane) to 24 hr (6th lane), all of which have similar TOF-MALDI spectra (like that shown for the 15-min digestion in Fig. 4-S1B).

REFERENCES

1. Radtchenko, I. L.; Sukhorukov, G. B.; Möhwald, H. *Colloids Surf. A* 2002, *202,* 127-33.
2. Price, A. D.; Zelikin, A. N.; Wark, K. L.; Caruso, F. *Adv. Mater.* 2010, *22*, 720-3.
3. Minten, I. J.; Hendriks, L. J. A.; Nolte, R. J. M.; Cornelissen, J. J. M. *J. Am. Chem. Soc.* 2009, *131*, 17771-3.
4. Caspar, D. L.; Klug, A. *Cold Spring Harb. Symp. Quant. Biol.* 1962, *27*, 1-24.
5. Bancroft, J. B. *Adv. Virus Res.* 1970, *16*, 99-134.
6. Hu, Y.; Zandi, R.; Anavitarte, A; Knobler, C. M.; Gelbart, W. M. *Biophys. J.* 2008, *94* , 1428-36.
7. Tang, J.; Johnson, J. M.; Dryden, K. A.; Young,M. J.; Zlotnick, A.; Johnson, J. E. *J. Struct. Biol.* 2006, *154*, 50-67.
8. Zandi, R.; Reguera, D.; Bruinsma, R. F.; Gelbart, W. M.; Rudnick, J. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 1555-15560.
9. van der Schoot, P.; Bruinsma, R. *Phys. Rev. E.* 2005, *71*, 061928.
10. Sikkema, F. D.; Comellas-Aragonès, M.; Fokkink, R. G.; Verduin, B. J. M.; Cornelissen, J. J. L.M.; Nolte, R. J. M. *Org. Biomol. Chem*. 2007, *5*, 54-7.
11. Minten, I. J.; Ma, Y.; Hempenius, M. A.; Vancso, G. J.; Nolte, R. J. M.; Cornelissen, J. J. L. M. *Org. Biomol. Chem.* 2009, *7*, 4685-88.
12. Erickson, J. W.; Rossman, M. G. *Virol.* 1982, *116*, 128-36.
13. Sorger, P. K.; Stockley,P. G.; Harrison, S. C. *J. Mol. Biol.* 1986,*191*, 639-58.

14. Chidlow, J.; Tremaine, J. H. *Virol.* 1971, *43*, 267-78.

15. Agrawal, H. O.; Tremaine, J. H. *Virol.* 1972, *47*, 8-20.

16. Burns, K.; Mukherjee, S.; Keef, T.; Johnson, J. M.; Zlotnick, A. *Biomacromol.* 201, *11*, 439-42.

17. Sun, J.; DuFirt, C.; Daniel, M.-C.; Murali, A.; Chen, C.; Gopinath, K.; Stein, B.; De, M.; Rotello, V. M.; Holzenberg, A.; Kao, C. C.; Dragnea, B. *Proc. Natl. Acad. Sci. USA* 2007*, 104,* 1354-9.

18. Adolph, K. W.; Butler, P. J. G. *Phil. Trans. R. Soc. Lond.* 1976, *276*, 113-22.

19. Zlotnick, A., private communication.

20. Belyi, V. A.; Muthukumar, M. *Proc. Natl. Acad. Sci. USA* 2007, *103*, 17174-8.

21. Daniel, M.-C.; Tsvetkova, I. B.;Quinkert, Z. T.; Murali, A.; De, M.; Rotello, V. M.; Kao, C. C.; Dragnea, B., *ACS Nano* 2010, *4*, 3853-60.

22. Zandi, R.; van der Schoot, P. A. M. *Biophys. J.* 2009, *96*, 9-20.

23. Choi, Y. G.; Rao, A. L. N. *Virology* 2002, *275*, 207-17.

24. Johnston, A. P. R.; Zelikin, A. N.; Lee, L.; Caruso, F. *Anal. Chem.* 2006, *78*, 5913-9.

CHAPTER 5

Reconstituted Plant Viral Capsids Can Release Genes to Mammalian Cells

ABSTRACT The nucleocapsids of many plant viruses are significantly more robust and protective of their RNA contents than those of enveloped animal viruses, making them an attractive option for virus-based gene delivery research. In particular, the capsid protein (CP) of the plant virus Cowpea Chlorotic Mottle Virus (CCMV) is of special interest because it has been shown to spontaneously package, with high efficiency, a large range of lengths and sequences of single-stranded RNA molecules. In this work we demonstrate that hybrid virus-like particles, assembled *in vitro* from CCMV CP and a heterologous RNA derived from a mammalian virus (Sindbis), are capable of disassembly in the cytoplasm of mammalian cells. Following this disassembly step and release of the RNA cargo, the RNA undergoes replication and translation similar to that of its naked RNA counterpart transfected under the same conditions. This result establishes the first step in the use of plant viral capsids as vectors for gene delivery and expression in mammalian cells. Furthermore, the CCMV capsid protects the packaged RNA against nuclease degradation and serves as a robust external scaffold with many possibilities for further functionalization and cell targeting.

INTRODUCTION

The use of animal virus-like particles (VLPs) as vectors for the delivery of genes to mammalian cells has been explored for many decades and by a large variety of very different approaches.[1,2,3] The major driving force motivating these studies is the fact that the raison d'être

for viruses – virtually their sole evolutionary pressure – is to protect, and deliver to host cells, the genes coding for their replication.

In this paper, we consider the use of a well-characterized plant RNA virus for expression of heterologous genes in mammalian cells. Unlike mammalian viruses – the majority of whose nucleocapsids are enveloped by an extra layer of protection in the form of a viral envelope – plant viruses are almost without exception "just" genetic material (DNA or RNA) surrounded by a shell composed of the capsid protein (CP). Because the viral genome is protected *only* by this protein shell, the resulting nucleocapsid (nucleic acid packaged inside the capsid) is significantly more robust than its counterpart in enveloped animal viruses.

Consider, for example, the "superfamily" of single-stranded RNA viruses comprised of bromoviruses (infecting plants) and alphaviruses (infecting animals), and whose members include Cowpea Chlorotic Mottle Virus (CCMV) and Sindbis Virus (SINV), respectively.[4] The nucleocapsids of both CCMV[5, 6, 7] and SINV [8, 9, 10, 11, 12] have been reconstituted *in vitro* from purified components. However, the nucleocapsids of CCMV are much more stable in solution than those of SINV. Furthermore, in contrast to SINV nucleocapsids that have evolved to depend on their viral envelope for protection against nucleases, CCMV capsids protect their RNA content against digestion.

Equally important, CCMV capsid protein has been shown to be capable of packaging, via spontaneous *in vitro* self-assembly, a wide range of *non*-viral cargo, including: heterologous RNAs,[13, 14, 15] synthetic anionic polymers,[16, 17, 18, 19] mineralized salts,[16] negatively charged colloidal particles such as gold nanocrystals[20] and oil-in-water nanoemulsion droplets,[21] fluorescent proteins,[22] and pH-controlled chromophores.[23] In addition, by functionalizing and chemically modifying their capsids, closely-related plant viruses like Cowpea Mosaic Virus

(CPMV) have been used for targeting of and uptake by specific mammalian cells for imaging and drug delivery.[24, 25, 26, 27, 28]

Plant-derived VLPs, however, have not been used for direct gene delivery. Interestingly, high-level expression of mammalian genes in plant hosts, mediated by Agrobacterium plasmids containing the translation-enhancement elements from messenger RNAs of the plant virus CPMV, has been demonstrated for various target proteins and vaccines[29]. Also, the insect virus baculovector system has been effectively used for expression of mammalian genes in a wide variety of insect and mammalian hosts[30], but unlike the system we describe here cannot be reconstituted *in vitro* and must be prepared by recombinant plasmid engineering in cell culture.

There have been no attempts to use an *intact* plant viral capsid to deliver genes for expression in *mammalian* cells, even though there are several independent demonstrations of the internalization of plant virus by such cells. Examples include the membrane-protein-mediated uptake of CPMV by mouse vascular endothelial cells[31] and the uptake of brome mosaic virus (BMV) by human bronchial epithelial cells.[32]

The possibility of uptake and release of the contents of a plant capsid in a mammalian cell has been suggested in the case of red clover necrotic mosaic virus (RCNMV).[33] In particular, capsids of purified virus loaded with doxorubicin were functionalized with HeLa cell-targeting peptides, and the survivability of HeLa cells overlaid by them shown to decrease with doxorubicin concentration. While disassembly of the capsids in the HeLa cells was not directly established, doxorubicin was argued to be released by the same divalent cation-controlled mechanism as is known to occur with RCNMV RNA in *plant* cell hosts.

In light of there being no direct demonstration of plant viruses disassembling and thereby releasing their contents in animal cells, we consider here the question: Can the genetic material

of intact plant VLPs be made available to a mammalian cell and their protein products synthesized?

In answering this question, and thereby providing a new platform for gene delivery to mammalian cells, we take advantage of many unique *in vitro* reconstitution properties of the spherical plant virus, CCMV. In particular – in contradistinction to the CPMV and RCNMV examples mentioned above – the purified capsid protein of CCMV is capable of efficiently packaging a large range of RNA lengths and sequences, including arbitrary transgenes of interest. If these *in vitro* synthesized VLPs can be shown to disassemble and release their RNA contents in mammalian cells, then the stage is set for functionalizing the capsids to target those cells *in vivo*.

Our strategy is the following. We reconstitute *in vitro* a well-characterized VLP consisting of a reporter ssRNA molecule packaged inside a CCMV capsid. To this end, we designed a SINV-derived defective interfering RNA (DI[EYFP], 1800 nt), which upon replication produces the mRNA for expression of high levels of enhanced yellow fluorescent protein (EYFP). The VLPs containing this RNA are transfected into a monolayer of baby hamster kidney (BHK) cells. The machinery for the replication of the DI[EYFP] RNA is supplied by a SINV-like particle (GP-Rep vector) providing the replicase proteins that will efficiently replicate the DI[EYFP] RNA molecule (into thousands of mRNA copies for translation) *if and only if the encapsidated DI[EYFP] RNA is released from the VLPs upon internalization*. Therefore, the expression of EYFP in vector-infected, VLP-transfected, cells can report on the successful disassembly of the VLPs and the release of the encapsidated RNA for replication and translation.

Figure 5-1 shows schematically the coding and regulatory sequences of the full SINV RNA genome (i), highlighting the two open reading frames (ORFs) coding for the nonstructural (NS) genes (RNA-dependent RNApolymerases, the replicases) and structural (S) genes ([CP] and membrane glycoproteins [GP]). Also shown is the SINV-derived glycoprotein "replicon" (GP-Rep) RNA (ii), obtained by deleting the CP gene in the full SINV genome (See Methods). A defective-interfering (DI) RNA molecule (iii) is derived from it by deleting most of the nonstructural ORF and replacing the structural genes by the EYFP gene (see METHODS).

RESULTS AND DISCUSSION

VLP synthesis and characterization

As described in the Methods section, DI[EYFP] RNA and an excess of CCMV CP are subjected to the standard protocol for *in vitro* reconstitutions of RNA and CCMV CP. Basically, they are incubated together at neutral pH and then dialyzed against pH 4.8 buffer solution.

The resulting mixture was run in a 1% agarose electrophoresis gel and showed no band associated with free (unpackaged) RNA. Rather, the only band observed was one corresponding to VLPs (see gel on right side of Figure 5-2, leftmost panel [-]) which are resistant to RNase digestion ([+]). Note (gel, middle panel) that this RNase treatment was sufficient for complete digestion of naked RNA. Comparing the band intensity with lanes containing a set of CCMV standards of known concentration (Figure 5-2 left, densitometry plot) showed that the assembled VLPs contained a concentration of RNA equal to 18 ng/mL. Furthermore, the VLPs were homogeneous (see, however, discussion below TEM data) and had electrophoretic mobilities corresponding to well-formed VLPs.

The assembly mix was imaged by negative-stain transmission electron microscopy (TEM). The images showed spherical VLPs (Figure 5-3, micrograph) with a distribution of diameters corresponding to roughly equal numbers of T=2 and T=3 capsids,[34, 35] overlapping the range of wt CCMV virions (Figure 5-3, histogram) – in agreement with what was previously found for CCMV CP VLPs packaging 2,000-nt long RNA.[36]

CCMV VLPs can release their RNA content in the cytoplasm of mammalian cells

Having characterized the plant virus VLPs containing DI[EYFP] RNA, we then transfected mammalian cells with them, using Lipofectamine-2000 (see METHODS). The transfected cells were incubated for 30 min to allow for internalization of the VLPs and for their disassembly in the cytoplasm. At this point, GP-Rep vector was added at a multiplicity of infection (MOI) of 100 (i.e., 100 vector particles per cell), in order to produce a high copy number of EYFP mRNA from DI[EYFP] RNAs that had been released from their CCMV capsids and successfully delivered to the cytoplasm. The intracellular fluorescence signal generated by the subsequent translation of this amplified mRNA into EYFP could then be used to report on the level of RNA delivery.

In parallel, equal amounts of the naked DI[EYFP] RNA were transfected by Lipofectamine-2000 into BHK cells under similar conditions. In addition, a sample consisting of "naked" VLPs (i.e., VLPs in final assembly buffer without Lipofectamine) were incubated with BHK cells.

The transfected cells were imaged by fluorescence microscopy to estimate the number of cells positive for EYFP. The number of transfected cells (positive for EYFP expression) was

also quantified by flow cytometry; the transfection efficiencies are shown as a function of RNase treatment in the top plot of Figure 5-4.

As the bottom of Figure 5-4 shows (lower, leftmost, panel), the transfected VLPs can deliver their RNA content into the cytoplasm of mammalian cells, where the RNA is involved in downstream processes (replication and translation of EYFP reporter) leading to gene expression, as is the case for the naked RNA control (see upper, leftmost, panel). As expected, the naked VLPs overlaid on cells without any Lipofectamine could not transduce cells with EYFP (data not shown).

The number of fluorescent cells is lower (by a factor of 5) in the case of VLP transfection than for RNA, associated with the fact that not all of the VLPs disassemble to make their RNA content available for replication and translation. The conclusion that only about a fifth of the VLPs are releasing their gene cargo – as opposed to 5 times fewer VLPs being transfected than RNAs – is consistent with experiments in which equal numbers of the same VLPs and RNA molecules are manually *microinjected* – rather than *transfected* – into the same BHK cells, and the VLPs found to give rise to a factor of 5 fewer fluorescent cells (data not shown).

CCMV VLPs remain intact prior to cell entry and protect their RNA content against RNase digestion

Two variations on the transfection procedure involving RNase A were carried out in parallel to gain further information about the physical state of the VLPs during transfection. They are referred to as "RN1" and "RN2" and are described below:

RN1: Before the initial equilibration with Lipofectamine-2000, RNase A was added to both the packaged and naked RNA samples. This digestion step was designed to remove all naked

full-length RNAs from solution and ensure that any transfection signal (reported by EYFP production) generated by the VLPs did not come from trace amounts of unpackaged RNA or from RNA that escaped from the VLPs during transfection. We observed (Figure 5-4, comparing leftmost and rightmost lower panels) essentially identical transfection efficiencies for VLPs with and without RNase digestion, indicating that unpackaged RNA was not transfected. We also observed the complete absence of transfection and gene expression in the digested naked RNA sample (see upper, rightmost, panel), indicating that the concentration of RNase A used was adequate for complete removal of all naked full-length RNA in solution. These results are also reflected quantitatively in the leftmost and rightmost pairs of points in the plot at the top of Figure 5-4.

RN2: RNase A was added immediately *after* the addition of Lipofectamine. This digestion step was designed to test whether Lipofectamine-2000 is capable of protecting unpackaged RNA against RNase A digestion. If this were the case, the RNase A digestion assay described above (i.e., [RN1]) would not distinguish between RNA that was transfected within intact VLPs and RNA that had escaped from VLPs that prematurely disassembled and entered the cell as an RNase-protected lipid-RNA complex. Our results show (see Figure 5-4, leftmost and middle panels – and also the corresponding pairs of points in the plot at the top) that adding RNase A at this step of the transfection procedure results in a large attenuation of the transfection efficiency of the naked RNA while having no effect on the transfection efficiency of the VLPs, indicating again that VLPs are being transfected as intact capsids.

The intact state of the plant viral capsids, even after RNase treatment in the absence and presence of Lipofectamine, is only one measure of their robustness. Another is their "shelf-life"

between synthesis and transfection: in the experiments described above, for example, the VLPs were left for 1 week at 4°C before being subjected to RNase treatment and transfection.

CONCLUSIONS

We have shown that plant VLPs formed *in vitro* from CCMV capsid protein and several-thousand-nt-long RNA: (1) are resistant to RNase; (2) can be transfected into mammalian cells, and (3) disassemble in these cells, resulting in the expression of the target transgene.

This demonstration establishes that plant virus capsids – here CCMV – are capable of protecting their RNA contents outside the cell, and yet making them available in the intracellular environment of mammalian cells. We have chosen the CP as well as the RNA of a superfamily of *plus*-strand RNA viruses like CCMV and SINV – as opposed to minus-strand or retro RNA viruses – in order to avoid the complications of needing to package directly the delicate RNA-dependent RNA polymerase or reverse transcriptase proteins themselves, Further, with plus-strand viruses, all of the nucleic acid replication takes place in the cytoplasm rather than requiring trafficking of RNA into and out of the host cell nucleus. In addition, utilizing the DI[EYFP] RNA in conjunction with GP-Rep vector infection, we are able to exploit the powerful messenger RNA amplification scheme unique to plus-strand RNA viruses, i.e., translation of messenger RNA and ensuing protein syntheses that take place only after the original delivered RNA is amplified.[4, 37] It is this enhancement that significantly increases the sensitivity for detecting disassembly of the VLPs in the cytoplasm.

These studies were undertaken to demonstrate that nucleocapsids of plant viral capsid protein are capable of making available their RNA content to mammalian cells. The results of this work can now be directly combined with functionalization of the capsid protein for its

targeting and uptake by mammalian cells,[24-28, 31-33] thereby facilitating the use of reconstituted plant viral capsids for gene delivery.

MATERIALS AND METHODS

DI[EYFP]-RNA

The DI[EYFP] construct was the result of (1) deletion of the region spanning BamHI and BspEI restriction sites on the NS ORF of SINV cDNA, and (2) replacement of the S ORF with nuclear EYFP ORF, using standard molecular biology protocols. The resulting plasmid – after linearization – was used as the template for *in vitro* transcription using Ambion Sp6 mMessage mMachine in vitro transcription kit. The RNA was purified by Qiagen mini RNeasy kit, and quantified by UV absorption at 260 nm.

GP-Rep Vector

First, two constructs, GP-Rep and DHBB plasmids, were prepared: GP-Rep was made by deleting the CP gene from the structural ORF in the SINV cDNA while DHBB was made by deleting the region between BamHI and BspEI sites on the NS ORF of SINV cDNA. These constructs were made into corresponding RNA transcripts, using *in vitro* transcription (see above). Next, the two RNAs were cotransfected into BHK cells using electroporation protocol. The vector particles were harvested from the medium on top of transfected cells after 24 hr. The titer of this GP-Rep vector was determined by the following procedure. (1) BHK cell monolayers were infected with serial dilutions of the prepared vector and an agar/medium layer was overlaid on top of infected cells. (2) After two days, the cells were stained with a 1xPBS solution of phenol red. At an appropriate dilution, small white puncta in the background of red stained cells

locate the sites of infection by GP-Rep vector particles. They were counted and their number was used to determine the titer of prepared GP-Rep vector. As the procedure outlined above is a standard plaque assay used for SINV, this also enables one to ensure that no wild-type virus contamination is present, which could easily be identified by the large characteristic plaques generated by SINV infection.

CCMV CP purification

CCMV was purified from infected California cowpea plant (Vigna ungiculata cv Black Eye)[38] and CP was isolated as described previously[36, 39]. SDS-PAGE and MALDI-TOF were employed to ascertain that the purified protein was intact.

VLP Assembly reactions

DI[EYFP] RNA was packaged by CCMV CP at a CP:RNA mass ratio of 6.5:1 to ensure complete RNA packaging.[36] Assembly was carried out by mixing CP and RNA in buffer B (1M NaCl, 20 mM Tris-HCl pH 7.2, 1 mM EDTA, 1 mM DTT and 1 mM PMSF), followed by dialysis against RNA assembly buffer (RAB: 50 mM NaCl, 10 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 50 mM Tris-HCl pH 7.2) for 4 hr, followed by dialysis against virus buffer (VB: 0.1 M sodium acetate, 1 mM EDTA, pH 4.8) for 16 hr, all at 4°C. Free (excess) CP was removed by washing the sample with VB in a 100kDa Amicon centrifugal filtration unit at 3,000 G at 4°C.

Electron microscopy characterization of VLPs

A 6 µl-aliquot of the final assembly mixture was applied to a glow-discharged copper grid (400-mesh) that previously had been coated with Parlodion and carbon. The mix was spread

onto the grid for 1 min, blotted with Whatman filter paper and then stained with 6 μL of 1% uranyl acetate for 1 min. Excess stain was removed by blotting with filter paper. The samples were stored overnight in a desiccator and analyzed with a JEM 1200-EX transmission electron microscope operated at 80 keV and equipped with a wide-angle (top mount) BioScan 600 W 1 × 1K pixel digital camera. The reported average diameter of VLPs is that of the geometric mean of two orthogonal measurements of the capsids obtained with ImageJ (U.S. National Institutes of Health) software from recorded images.

Preparation of BHK cell culture

Prior to transfection, BHK-21 cells were grown as monolayers on 6-well plates at 37°C in a CO_2 incubator. EMEM (ATCC) supplied with 10% (v/v) FCS and Penicillin-streptomycin was used to grow and maintain cells.

Lipofectamine transfections

Monolayers of BHK cells were grown on six-well plates to 90% confluence. Transfection with Lipofectamine-2000 was carried out following a standard protocol normally used for the transfection of nucleic acids. Briefly, the protocol consisted of the following (5) steps:

(1) An aliquot of VLPs (or naked RNA) containing 1ug of DI[EYFP] RNA was diluted into Opti-MEM to 300 μl for each transfection. Similarly, 3 μl of Lipofectamine-2000 was diluted into 300 μl Opti-MEM and the dilutions were incubated at room temperature for 5 min.

(2) The two diluted reagents were mixed and (RNA/VLP + Lipofectamine) complexes were allowed to form by incubating at room temperature for 30 min.

(3) The cells were washed with 1x nuclease-free PBS before overlaying the transfectant onto the cells, followed by incubation at 37°C in a CO_2 incubator with occasional shaking to distribute the transfectant evenly on the cells.

(4) After 30 min of incubation, a 150-mL aliquot of GP-Rep vector at an MOI of >100 was added to the cells in order to supply the replication machinery necessary for EYFP production.

(5) After 4 hr the transfectant was removed and the cells were gently washed twice with pre-warmed 1x PBS and 1 ml of 2% FCS medium was overlaid on the cells. The cells were left overnight at 37°C in a CO_2 incubator.

RNase A assays

All RNase A digestion steps were carried out using 0.2 ng/µl RNase A for 30 min at room temperature. For digestions that were performed during the transfection protocol (i.e. scenarios RN1 and RN2), this concentration of RNase A was maintained throughout each subsequent step of the protocol. These conditions were found to be sufficient to remove all full-length RNA band intensity from the agarose gel assay (Figure 5-2, gel, middle panel) as well as completely destroy the transfection efficiency of naked RNA by Lipofectamine-2000 (Figure 5-4, rightmost panel).

Flow cytometry

24 hr post-transfection, the cells were washed with 1x PBS, trypsinized, and resuspended in 1x PBS buffer with 2% FCS at a density of 3 x 10^6 cells/ml. Analysis was performed using a Becton Dickinson SORP BD LSRII Analytic Flow Cytometer equipped with a 488-nm blue laser

for excitation of EYFP. Yellow-green YFP fluorescence was collected after a 530/30 bandpass filter. FACSDIVa software was used to control the parameters during the run and to analyze collected data.

ACKNOWLEDGEMENTS

O.A. thanks A. Berk for hosting and mentoring her in his laboratory. We also thank M. Jiang and Y. Wang for performing the microinjection experiments, and A. L. N. Rao for helpful discussions and comments on the manuscript. The counting of fluorescent cells was carried out in the UCLA Flow Cytometry Core Facility, and the TEM images were obtained in the California NanoSystems Institute (CNSI) Electron Imaging Center for Nano-Machines supported by NIH (1S10RR23057). This work was supported by the U. S. National Science Foundation, in the form of grant CHE 1051507 to W.M.G. and C.M.K.

FIGURES

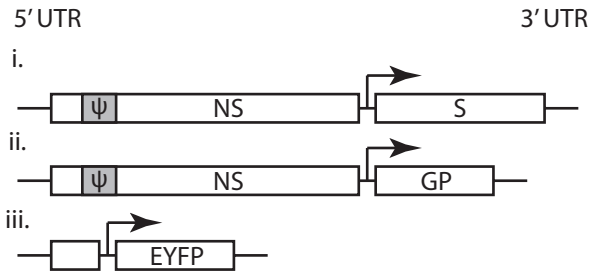

Figure 5-1. Schematics of the relevant RNA reagents. i: Wt SINV genome. The blocks labeled NS and S are the open reading frames coding for the replicase proteins and the structural proteins ([CP] and [GPs]). Ψ denotes the packaging signal, a sub-sequence in NS responsible for the

preferential packaging of the RNA by its capsid protein. The hook-arrow denotes the promoter sequence controlling replication and translation of the downstream (structural, S) genes. ii: GP-replicon RNA. Same as i, except for deletion of the CP gene. iii: DI[EYFP] RNA. Derived from i upon deletion of most of the NS ORF and replacement of the structural genes (S ORF) with EYFP ORF. The Ψ signal is not present, but the cis-acting elements – the 5' and 3' untranslated regions (UTRs) needed for replication by the NS complex – are retained.

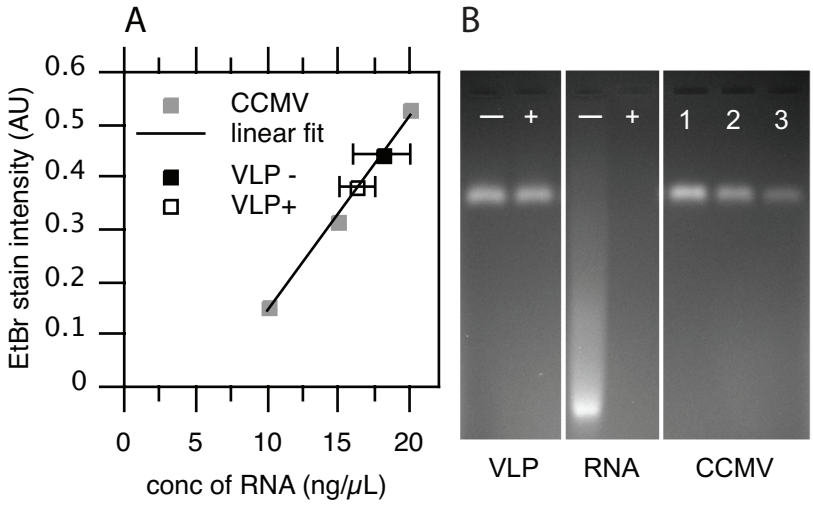

Figure 5-2A: Densitometry plot used to quantify the amount of RNA packaged in VLPs by measuring the EtBr fluorescence intensity from the bands. 2B: Agarose gel electrophoresis of synthesized VLPs, naked RNA, and CCMV particle standards with known concentrations. Note that no calibration ladder is included because only one length (1800nt) of (DI[EYFP]) RNA is involved, and wildtype CCMV virions serve as the size markers for the VLPs. Electrophoresis was carried out in a 1% agarose gel and in virus buffer (VB: see METHODS); the gel was stained with EtBr. (+) and (-) designations refer to the presence or absence of RNase A. Rightmost panel: CCMV standards prepared by adding 10 mL of CCMV at concentrations of 93 (1), 69.8 (2), and 46.5 (3) ng/mL, respectively, corresponding in turn to concentrations of RNA (within the virions) of 20, 15, and 10 ng/mL.

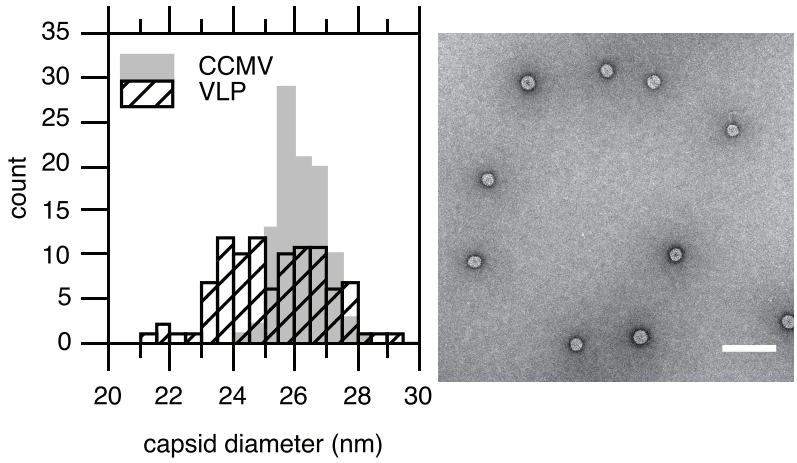

Figure 5-3. Left: Distributions of diameters of CP + DI[EYFP] VLPs (cross-hatched) and of wt CCMV virions (grey). Right: A negative-stain TEM micrograph of the VLPs assembled from CCMV CP + DI[EYFP] RNA. Scale bar = 100 nm.

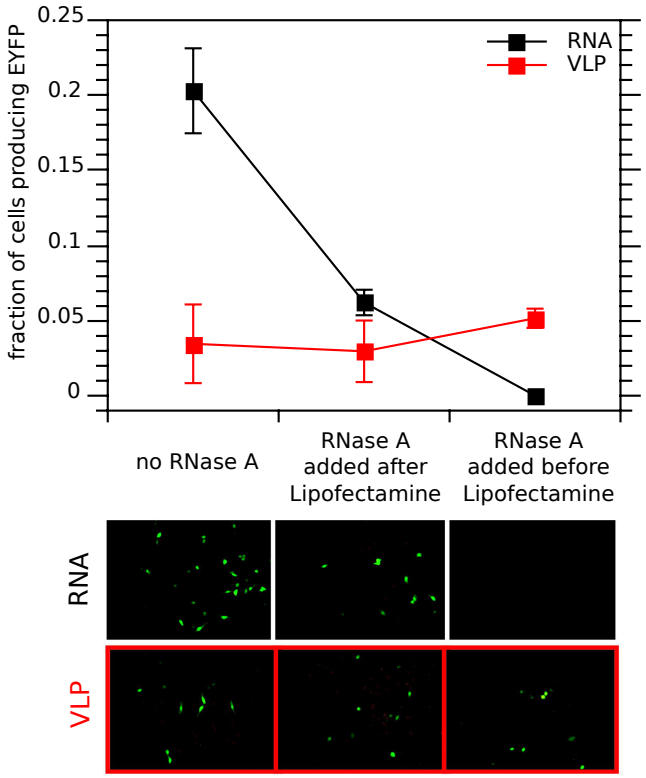

Figure 5-4. Top: Plot of fraction of fluorescent cells as quantified by flow cytometry. The leftmost points show the average fraction of positive cells for a duplicate set of transfections of VLPs (red) and naked RNA (black) that were not treated with RNase A. The middle set of points shows the effect of adding RNase A after mixing with Lipofectamine-2000 (RN2, see text). The leftmost points show the result of incubation with RNase A before addition of Lipofectamine (RN1). The line connecting the separate trials is meant to aid the eye. Error bars show the differences between measured values in pairs of the duplicate trials. Bottom: Fluorescence micrographs of representative fluorescent cell densities for each scenario.

REFERENCES

1. Schaffer, D. V.; Koerber, J. T.; Lim, K. Molecular Engineering of Viral Gene Delivery Vehicles. *Annu. Rev. Biomed. Eng.* 2008, *10*, 169-94.

2. Lavillette, D.; Russell, S. J.; Cosset, F. L. Retargeting Gene Delivery Using Surface-engineered Retroviral Vector Particles. *Curr. Opin. Biotech*. 2001, *12*, 461-66.

3. Soong, N. W.; Nomura, L.; Pekrun, K.; Reed, M.; Sheppard, L. et al. Molecular Breeding of Viruses. *Nat. Genet.* 2000, *25*, 436-39.

4. Strauss, J. H.; Strauss, E. G. The Alphaviruses: Gene Expression, Replication, and Evolution. *Microbiol. Rev.* 1994, *58*, 491-562.

5. Bancroft, J. B.; Hiebert, E. Formation of an Infectious Nucleoprotein from Protein and Nucleic Acid Isolated from a Small Spherical Virus. *Virology* 1967, *32*, 354 –356.

6. Zhao, X.; Fox, J. M.; Olson, N. H.; Baker, T. S. ;Young, M. J. In vitro assembly of cowpea chlorotic mottle virus from coat protein expressed in Escherichia coli and in vitro-transcribed viral cDNA. *Virology* 1995, *207*, 486-494.

7. Fox, J. M.; Wang, G.; Speir, J. A.; Olson, N. H.; Johnson, J. E.; Baker, T. S.; Young, M. J. Comparison of the native CCMV virion with in vitro assembled CCMV virions by cryoelectron microscopy and image reconstruction. *Virology* 1998 *244*, 212-8.

8. Wengler, G.; Boege, U.; Wengler, G.; Bischoff, H.; Wahn, K. The Core Protein of the Alphavirus Sindbis Virus Assembles into Core-like Nucleoproteins with the Viral Genome RNA and with other Single-stranded Nucleic Acids in Vitro. *Virology* 1982, *118*, 401-410.

9. Tellinghuisen, T. L.; Hamburger, A. E.; Fisher, B. R.; Ostendorp, R.; Kuhn, R. J. In vitro Assembly of Alphavirus Cores by Using Nucleocapsid Protein Expressed in E. coli. *J. Virology* 1999, *73*, 5309-19.

10. Mukhopadhyay, S.; Chipman, P. R.; Hong, E. M.; Kuhn, R. J.; Rossmann, M. G. In vitro-assembled Alphavirus Core-like Particles Maintain a Structure Similar to that of Nucleocapsid Cores in Mature Virus. *J. Virol.* 2002, *76*, 11128-11132.

11. Goicochea, N. L.; De, M.; Rotello, V. M.; Mukhopadhyay, S.; Dragnea, B. Core-like Particles of an Enveloped Animal Virus Can Self-Assemble Efficiently on Artificial Templates. *Nano Lett.* 2007, *7*, 2281-90.

12. Cheng, F.; Mukhopadhyay, S. Generating virus-like particles with in vitro assembled cores. *Virology* 2011, *413*, 153-60.

13. Hiebert, E.; Bancroft, J. B.; Bracker, C. E. The Assembly in vitro of Some Small Spherical Viruses, Hybrid viruses, and Other Nucleoproteins. *Virology* 1968, *34*, 492–508.

14. Bancroft, J.B.; Hiebert, E.; Bracker, C. E. The Effects of Various Polyanions on Shell Formation of Some Spherical Viruses. *Virology* 1969, *39*, 924–930.

15. Rao, A. L. N. Genome Packaging by Spherical Plant RNA Viruses. *Annu. Rev. Phytopathol*. 2006, *44*, 61– 87.

16. Douglas, T.; Young, M. Host–guest Encapsulation of Materials by Assembled Virus Protein Cages. *Nature* 1998, *393*, 152-5.

17. Hu, Y.; Zandi, R.; Anavitarte, A.; Knobler, C. M.; Gelbart, W. M. Packaging of a Polymer by a Viral Capsid: The Interplay between Polymer Length and Capsid Size. *Biophys. J.* 2008, *94*, 1428-36.

18. Brasch, B.; Cornelissen, J. J. L. M. Relative size selection of a conjugated polyelectrolyte in virus-like protein structures. *Chem. Commun.* 2012,*48*, 1446-1448.

19. Cadena-Nava, R. D.; Hu, Y.; Garmann, R. F.; Ng, B.; Zelikin, A. N.; Knobler, C. M.; Gelbart, W. M. Exploiting Fluorescent Polymers to Probe the Self-Assembly of Virus-Like Particles. *J. Phys. Chem. B* 2011, *115*, 2386-93.

20. Chen, C.; Daniel, M. C.; Quinkert, Z. T.; De, M.; Stein, B.; Bowman, V. D.; Chipman, P. R.; Rotello, V. M.; Kao, C. C.; Dragnea, B. Nanoparticle-Templated Assembly of Viral Protein Cages. *Nano Lett.* 2006, *6*, 611–615.

21. Chang, C. B.; Knobler, C. M.; Gelbart, W. M.; Mason, T. G. Curvature Dependence of Viral Protein Structures on Encapsidated Nanoemulsion Droplets. *ACS Nano* 2008, *2*, 281-6.

22. Minten, I. J.; Hendriks, L. J. A.; Nolte, R. J. M.; Cornelissen, J. J. L. M. Controlled Encapsulation of Multiple Proteins in Virus Capsids. *J. Am. Chem. Soc.* 2009, *131*, 17771–17773.

23. Brasch, M.; de La Escosura, A.; Ma, Y.; Uetrecht, C.; Heck, A. J. R.; Torres, T.; Cornelissen, J. J. L. M. Encapsulation of Phthalocyanine Supramolecular Stacks into Virus-like Particles. *J. Am. Chem. Soc.* 2011, *133*, 6878-6881.

24. Destito, G.; Yeh, R.; Rae, C. S.; Finn, M. G.; Manchester, M. Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells. *Chem. Biol.* 2007, *14*, 1152-62.

25. Gonzales, M. J.; Plummer, E. M.; Rae, C. S.; Manchester, M. J. Interaction of Cowpea Mosaic Virus (CPMV) Nanoparticles with Antigen Presenting Cells In Vitro and In Vivo. *PLoS One* 2009, *4*, e7981.

26. Yildiz, I.; Shukla, S.; Steinmetz, N. F. Applications of Viral Nanoparticles in Medicine. *Curr. Opin. In Biotechnol.* 2011, *22*, 901-908.

27. Steinmetz, N. F. Viral Nanoparticles as Platforms for Next-Generation Therapeutics and Imaging Devices. *Nanomed.* 2010, *6*, 634-641.

28. Wu, Z.; Chen, K.; Yildiz, I.; Dirksen, A.; Fischer, R.; Dawson, P. E.; Steinmetz, N. F. Development of Viral Nanoparticles for Efficient Intracellular Delivery. *Nanoscale* 2012, *4*, 3567-76.

29. Sainsbury, F.; Lomonossoff, G. P. Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. *Plant Physiol.* 2008, *148*, 1212-1218.

30. Chen, C.-Y.; Lin, C.-Y.; Chen, G.-Y.; Hu, Y.-C. Baculovirus as a gene delivery vector: Recent understandings of molecular alterations in transduced cells and latest applications. Biotechnology Advances 2011, 29, 618-631.

31. Koudelka, K. J.; Destito, G.; Plummer, E. M.; Trauger, S. A.; Siuzdak, G.; Manchester, M. Endothelial Targeting of Cowpea Mosaic Virus (CPMV) via Surface Vimentin. *PLoS Pathogens* 2009, *5*, e1000417/1-10.

32. Jung, B.; Rao, A. L. N.; Anvari, B. Optical Nano-Constructs Composed of Genome-Depleted Brome Mosaic Virus Doped with a Near Infrared Chomophore for Potential Biomedical Applications. *ACS Nano* 2011, *5*, 1243-1252.

33. Lockney, D. M.; Guenther, R. N.; Loo, L.; Overton, W.; Antonelli, R.; Clark, J; Hu, M.; Luft, C.; Lommel, S. A.; Franzen, S. The RCNMV Capsid as a Multifunctional Cell Targeting Plant Viral Nanoparticle. *Bioconjugate Chem*. 2011, *22*, 67-73.

34. T=2 and T=3 refer to Caspar-Klug structures[35] consisting of 120 and 180 copies of the capsid protein, respectively. While we do not establish directly their structures, we deduce them[36] from their relative sizes and the fact that the T=3 VLPs are identical in size to those of the wildtype, T=3, CCMV capsids.

35. Caspar, D. L. D.; Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symp*. 1962, *27*, 1-24.

36. Cadena, R. D.; Comas-Garcia, M.; Garmann, R. F.; Rao, A. L. N.; Knobler, C. M.; Gelbart, W. M. Self-assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *J. Virology* 2012, *86*, 3318-26.

37. Frolov, I.; Hoffman, T. A.; Pragai, B. M.; Dryga, S. A.; Huang, H. V.; Schlesinger, S.; Rice, C. M. Alphavirus-based Expression Vectors: Strategies and Applications. *Proc. Natl. Acad. Sci. (USA)* 1996, *93*, 11371-7.

38. Bancroft, J. B. The Self-Assembly of Spherical Plant Viruses. *Adv. Virus Res.* 1970, *16*, 99-134.

39. Annamalai, P.; Rao, A. L. N. Dispensability of 3'tRNA-Like Sequence for Packaging Cowpea Chlorotic Mottle Virus Genomic RNAs. *Virol.* 2005, *332*, 650-658.

CHAPTER 6

Visualizing the Global Secondary Structure of a Viral Genome with Cryo-Electron Microscopy

ABSTRACT The lifecycle, and therefore the virulence, of single-stranded (ss)-RNA viruses is largely regulated by the secondary and tertiary structure of their genomes. The secondary structure of the entire genomic RNA of satellite tobacco mosaic virus (STMV) was recently determined by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). The SHAPE analysis suggested a single highly extended secondary structure with much less branching than occurs in the ensemble of structures predicted by purely thermodynamic algorithms. Here we examine the solution-equilibrated STMV genome by direct visualization with cryo-electron microscopy (cryo-EM), using an RNA of similar length transcribed from the yeast genome as a control. The cryo-EM data reveal an ensemble of branching patterns that are collectively consistent with the SHAPE-derived secondary structure model. Thus, our results both elucidate the statistical nature of the secondary structure of large ss-RNAs and give visual support for modern RNA structure determination methods. Additionally, this work introduces cryo-EM as a means to distinguish between competing secondary structure models if the models differ significantly in terms of the number and/or length of branches. Furthermore, we suggest the possibility of developing methods that incorporate constraints from cryo-EM into the next generation of algorithms for RNA structure determination.

SIGNIFICANCE Single-stranded (ss)-RNAs encode cellular function in various ways, involving both their primary sequence and their secondary and tertiary structures. Consistent with this fact, modern RNA structure determination methods (e.g. SHAPE) suggest that the genomes of some

ss-RNA viruses adopt specific secondary structures that direct different aspects of the viral lifecycle. This result appears at odds with thermodynamic arguments describing a large number of thermally accessible secondary structures at equilibrium. Our results demonstrate that the two scenarios are, in fact, compatible. Specifically, a viral genome in solution can adopt multiple conformations whose ensemble properties are consistent with the SHAPE-determined structure. More generally, we introduce cryo-EM as a tool for studying the large-scale features of viral RNA secondary structures.

INTRODUCTION

The viral genome encodes the entire parasitic lifecycle of the virus at multiple levels. In addition to the primary sequence that directly codes for the viral proteins, single-stranded (ss)-RNA viruses exploit their secondary and tertiary structures to direct various events during infection. For example, specific secondary structure motifs within the viral genome have been implicated in packaging (1-4), genome replication (5-7), the regulation of protein expression (8), protein folding (9) and conformational switching (10), and evading degradation by the host cell (11).

However, relatively little is known about the large-scale – global – secondary-tertiary structures of these ss-RNA genomes. In part, this is because RNA is a flexible polymer that is unlikely to adopt a single unique 3D-structure at equilibrium. In addition, large RNAs (of order 1000 nt) may assume an ensemble of secondary structures, each containing a different combination of intra-molecular base pairings between nucleotides in both local and distant regions of the primary sequence, resulting very different branching patterns. Moreover, the secondary structures may well contain kinetically trapped portions that are formed during

replication. Furthermore, it is not known whether, or to what extent, the secondary structure is refolded upon packaging into the virus.

All of these factors preclude determination of the detailed global 3D-structure of the genomic RNA by the techniques traditionally used for high-resolution molecular structure studies that require a large number of identical structures – e.g., x-ray crystallography or cryo-electron microscopy (cryo-EM) single-particle reconstruction. At best, when these techniques are applied to the whole virions of small RNA viruses, they can reveal double-helical regions if those are organized with the same icosahedral symmetry as the capsid proteins. Such double helices have been seen in pariacoto virus (12), bacteriophage MS2 (13), and satellite tobacco mosaic virus (STMV) (14, 15).

The lack of direct structural data can be supplemented by RNA secondary structure predictions. Thermodynamic prediction algorithms (16-22) are based on estimates of the free energies associated with the known structural motifs – helices, loops, bulges, and junctions – that are derived from melting studies on small model RNAs (23). When applied to the genomes of ss-RNA viruses (24), these algorithms predict a multiplicity (of order 1000) of energetically allowed structures due to the large number of degenerate (within thermal energy, kT) ways of forming distinct base-pairings between both local and distant regions of the primary sequence.

A more recent refinement to RNA secondary structure prediction algorithms involves the incorporation of chemical probing data, which provide information on the extent to which individual nucleotides are constrained by base pairing or other interactions (25). When such data are used to supplement thermodynamic predictions on relatively short RNAs, significantly more reliable secondary structure maps can be produced (26). Furthermore, when applied to the genomes of ss-RNA viruses, these techniques suggest a much smaller (of order unity) ensemble

of structures (9, 11, 27, 28). We previously incorporated data from one such technique, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (29), to suggest the most likely secondary structure of the STMV genomic RNA generated by *in vitro* transcription (11). Figure 6-1A shows our SHAPE-directed secondary structure model. It contains three main branches and is significantly more extended than the ensemble of structures predicted by purely thermodynamic algorithms, a representative member of which, the minimum free energy (MFE) structure, is shown in Figure 6-1B.

In a subsequent study, Archer *et al.* (27) used SHAPE-directed analysis to determine the secondary structure of STMV RNA that had been extracted from the wild-type virion, finding a structure very similar to that of Figure 6-1A: a compact, highly branched domain containing the 5' and 3' ends of the RNA, and a long central arm with minimal branching. They also probed the RNA inside the virus, finding only modest differences between the structure of the RNA *in virio* and the two *ex virio* structures. In addition, those same authors used atomic force microscopy (AFM) to visualize the *ex virio* RNA and found that the approximate lengths and number of branches were consistent with their SHAPE-directed secondary structure map. While AFM has been successful in elucidating various structural properties of viral RNA genomes (28, 30, 31), inferring the equilibrium secondary structure of solvated RNA from AFM measurements is made difficult by poor lateral resolution and the fact that the sample RNA must be adsorbed onto the imaging surface before visualization. If the energy of interaction between the RNA and the imaging surface is comparable to the energy stabilizing the secondary structure motifs of the RNA, the observed conformations of the adsorbed RNA may not represent those realized in solution (32). Therefore, the observations made by AFM need to be corroborated by complementary measurements on RNA molecules free in solution.

Cryo-EM has emerged as a powerful technique for the direct imaging of solvated macromolecules in their native conformations (33). Recent experiments by Gopal *et al.* (34) demonstrated that cryo-EM is capable of resolving the large-scale features of long RNA molecules (thousands of nucleotides) and elucidating coarse-grained features of the ensemble of conformations they assume in solution. Specifically, they were able to determine the average 3D size and shape of long RNAs and distinguish between different lengths and sequences.

In the work presented here, we extend cryo-EM to the study of the secondary structure of a viral RNA. To do this we record the 2D projections of freely-solvated RNAs in their native conformations and quantify the compactness of the observed branching patterns. We compare *in vitro* transcribed molecules of STMV RNA with a control RNA of comparable length, taken from a non-coding region of the yeast genome. We find that while neither sequence adopts a single unique structure in solution, the ensemble of STMV RNA structures is significantly more extended than that of the yeast genome-derived RNA. We are able to correlate the observed extendedness of STMV RNA with a lack of high-order junctions and an unusually long central arm motif within the predicted secondary structure map (Figure 6-1A). We conclude that cryo-EM can be used to study the large-scale secondary structure feature of large RNA molecules, and we suggest that data from cryo-EM could be used to inform structure predictions on long RNA molecules.

RESULTS AND DISCUSSION

The secondary structure of the genomic RNA of STMV, both *in virio* and *ex virio,* has been previously determined by two independent SHAPE analyses (11, 27), both of which find a structure very similar to that shown in Figure 6-1A. In order to test the validity of the SHAPE-

derived structure we have directly visualized an *in vitro* transcript of the STMV genome by cryo-EM. Unfortunately, reconstructing a precise secondary structure map from cryo-EM micrographs is made difficult by fundamental limits in spatial resolution as well as the inability to resolve structural features that lie perpendicular to the imaging plane (an inherent limitation of projection imaging; see Supplemental Information). While single-particle cryo-EM reconstruction and cryo-electron tomography are capable of resolving the high-resolution 3D-structure of certain biomolecules, at present these techniques cannot be used to study large RNAs due to their structural heterogeneity and sensitivity to radiation damage, respectively. Rather, we are constrained to study the features of the secondary structures that reveal themselves within the observed ensemble of 2D-projected images.

The compactness of individual secondary structures can be described by a metric, the maximum ladder distance (MLD), that was defined by Yoffe, *et al.* (24, 35). The ladder distance associated with an arbitrary pair of bases (nucleotides) in a secondary structure had been introduced earlier by Bundschuh and Hwa (35); it is defined as the *minimum* number of base pair "rungs" that must be crossed in reaching one base from the other. The MLD is the *maximum* value of all ladder distances defined in this way. Accordingly, it is the longest path across the secondary structure (see Figure 6-1). For random-sequence RNAs, it was found that the relationship between the ensemble-averaged MLD (<MLD>) and the length of the RNA (the number of nucleotides, *N*) follows a simple power law:

$$\langle MLD \rangle = 1.37 N^{0.67} \qquad (1)$$

The SHAPE-determined secondary structure model of STMV RNA (Figure 6-1A) has an MLD of 205, which is 40% larger than that expected from Eq. (1).

To verify the extendedness of the STMV genome, we compare the cryo-EM images of

STMV RNA to those obtained from a 975-nt sequence chosen randomly from a non-coding region of chromosome XII of the yeast genome (*Saccharomyces cerevisiae*); we refer to this control molecule as "yeast RNA". We choose yeast RNA as a reference structure with which to compare STMV RNA because it is of roughly equal length and has a predicted secondary structure that closely resembles those of random sequence RNAs. More specifically, yeast RNA has an MLD of about 100-150, depending on which prediction algorithm one uses (Supplementary Information, Figure 6-S1). These values are close to the predicted <MLD> of 138 for a randomly generated sequence of equal length (Eq. (1)). Thus, by comparing the cryo-EM-determined structure of STMV RNA to that of yeast RNA, we are effectively comparing STMV RNA to an equal-length random sequence.

121 images of STMV RNA were obtained from 11 micrographs, along with 122 images of yeast RNA from 6 micrographs. Figure 6-2A shows a representative micrograph of STMV RNA. The observed branched and wispy images reflect the complicated networks of double-helical segments connected by various kinds of single-stranded junctions (i.e., the secondary structure). The projected image of each molecule was converted to a one-pixel-wide skeleton (Figure 6-2B) and the total length (TL_{2D}) of each skeleton was measured. Only skeletons with TL_{2D} values between 25nm and 85nm were kept (100 of each sequence); those with lower or higher TL_{2D} values were discarded due to concerns regarding degradation and aggregation, respectively. The upper bound was chosen by noting that about 50-65% of the nucleotides are paired in RNAs with compositions typical of viral genomes. Taking 65% as the upper limit, an RNA with 1000 nt will have a maximum of ~325 base pairs. Using the rise of the A-form RNA double helix, 0.26 nm/bp, we obtain a maximum base-paired length of 325 bp * 0.26 nm/bp ≈

85 nm. In any case, molecules with TL_{2D} values above the upper bound or below the lower bound are rare.

For each sequence we find a diverse ensemble of skeletons (Figure 6-3), which we attribute to three main factors. First, RNA is flexible, so even molecules with identical branching will have different three-dimensional conformations at equilibrium. Second, 3D rotations of the imaged RNA act to shuffle the relative orientation of the branches as well as hide portions of the structure (see Supplementary Information, Figure 6-S2). Lastly, it is likely that the secondary structure varies at equilibrium. While flexibility and rotation alone may account for the structural differences between many pairs of skeletons (e.g. compare the yeast RNA skeleton found in row 1/column 6 of Figure 6-3 to that found in row 1/column 7), some pairs differ so dramatically in their branching that they clearly must represent different secondary structures (e.g. compare the yeast RNA skeleton found in row 1/column 6 or 7 of Figure 6-3 to that found in row 1/column 8). Examples of such strongly dissimilar pairs are present in both the viral and yeast RNA sequences and support the statistical – ensemble – interpretation of long ss-RNA secondary structures at equilibrium. Throughout the remainder of this study we focus on identifying the average properties of the skeletons of Figure 6-3 that allow us to characterize the underlying ensembles of secondary structures.

The TL_{2D} distributions for STMV RNA and yeast RNA are seen to overlap (Figure 6-4A), with mean values of 59 ± 16nm and 56 ± 15nm (SD), respectively. Because portions of the RNA structure are obscured by projection, the TL_{2D} cannot be interpreted as the collective length of every branch within the secondary structure (the TL_{3D}). However, because the TL_{3D} of long RNAs depends primarily on the fraction of base pairing and less on the specific connectivity, the TL_{2D} distributions from distinct populations of equal length RNAs should be roughly equivalent

– irrespective of sequence or secondary structure – as long as the populations sample all possible spatial orientations with equal weight. The distributions of Figure 6-4A are consistent with this scenario.

Analysis of the connectivity between segments of the skeletons of the two RNAs reveals average structural differences between them; we define a segment as the path connecting two neighboring branch points or a branch point and its neighboring endpoint. On average, STMV contains only two 3-fold branch points (V_3 vertices) per skeleton, while yeast RNA has three (Figure 6-4B). Additionally, 4-fold branch points (V_4 vertices) are more common in yeast RNA than in STMV RNA (Figure 6-4B, inset). While limited spatial resolution and projection artifacts preclude the accurate determination of the true 3D-branching of each molecule, the decreased number of vertices in the skeletons of STMV RNA relative to yeast RNA suggests a smaller number of junctions within the secondary structure map (i.e., less branching). While the most common segment length for both RNAs is ≈6nm, STMV RNA has a relative abundance of longer (>20nm) segments (Figure 6-4C).

Collectively, these observations are consistent with STMV being more extended than yeast RNA, and thus support the SHAPE-constrained model (Figure 6-1A) over the MFE structure derived from purely thermodynamic algorithms (Figure 6-1B). Furthermore, plotting the distribution of the longest-segment length of each skeleton (Figure 6-4D) shows that STMV RNA contains a broad range of such segments, with a significant excess of longest-segment lengths over 20nm when compared to yeast RNA. This observation supports the existence of the long central arm present in the SHAPE-derived model of the STMV genome.

As an alternate metric of extendedness, akin to the MLD, we define the maximum length (ML_{2D}) of an RNA as the length of the longest connected path between any two points within its

projected image skeleton. We see in Figure 6-4E that plotting the distribution of ML_{2D} values also distinguishes the relative extendedness of STMV RNA compared to yeast RNA – their mean ML_{2D} values are 46 ± 13nm and 37 ± 9nm (SD), respectively.

Lastly, we investigate the two subsets of 10 skeletons with the highest accepted TL_{2D} values (i.e. the highest TL_{2D} values less than 85 nm) for each sequence. We claim that these skeletons (Figure 6-3, bottom-most row) most accurately represent the true secondary structures that are adopted in solution because each corresponding RNA was imaged in an orientation that has maximized the amount of resolvable structure (i.e., minimized the amount of hidden structure). Within this subset, we find good agreement with the previously reported vertex distributions of Figure 6-4B. Specifically, 20 three-fold vertices and a single four-fold vertex were found for STMV RNA, and 30 three-folds and 6 four-folds were found for yeast RNA.

The average TL_{2D} values for these subsets of STMV RNA and yeast RNA are 82 ± 3nm and 79 ± 3nm, respectively. These values – which should be the same for equal-length molecules – reflect the slightly greater length of the STMV RNA (1058 nt) *vs.* the yeast RNA control (975 nt). The average ML_{2D} values, on the other hand, are 63 ± 7nm and 45 ± 7nm, respectively. These values are within the ranges of the physical length corresponding to the MLD of the SHAPE-derived STMV RNA secondary structure (51-71nm, calculated from Figure 6-1A, depending on whether or not one includes the approximate lengths of the unpaired regions), and that of the yeast RNA structure shown in Figure 6-1C (37-51nm). In contrast, the physical length of the MLD of the non-SHAPE-constrained STMV RNA structure shown in Figure 6-1D is 25-35nm. Thus, after accounting for the inherent underestimation of 3D distances associated with projection imaging (see Supplementary Information, Figure 6-S5), we find rough quantitative

agreement between the cryo-EM resolved structures and the SHAPE-derived secondary structure of STMV RNA and the structure of yeast RNA generated by purely thermodynamic analysis.

CONCLUSION

This work extends an earlier cryo-EM study on large ss-RNA molecules in solution (34), demonstrating that the large-scale features of RNA secondary structure can be reliably determined by cryo-EM. The direct visualization of an *in vitro* transcript of genomic STMV RNA reveals an ensemble of highly-extended secondary structures. Comparing the cryo-EM data to the STMV RNA models generated with and without SHAPE constraints (Figure 6-1A and 6-1B, respectively) demonstrates that the SHAPE-derived model better represents the average features (i.e. the overall extendedness) of the conformations realized in solution. In particular, the relatively low degree of branching, the large MLD, and the viroid-like central arm were directly observed. Thus cryo-EM supports the strength of chemical probing techniques, at least those using the SHAPE technology (9, 11, 27, 28), as a method for determining the global secondary structure of viral RNAs.

At the same time, the variety of structures we observe with cryo-EM cautions against interpreting the SHAPE-derived secondary structure map as the only structure found in solution. Instead, it appears that the global architecture of the STMV genome is better described as a family of structures that, as a whole, closely resemble the SHAPE-derived model. In a recent study by Wu et al. (28), the entire 4,800-nt genome of the tomato bushy stunt virus was solved by SHAPE. However, these authors found that only two of the six previously confirmed long-range secondary structure motifs required for biological function were present in their model. From this it was proposed that the global secondary structure acted as a basic scaffold for the

dynamic development of distinct substructures. Our results are consistent with this scenario and may reflect a general strategy exploited by ss-RNA viruses.

Continual improvements in the detection hardware of electron microscopes (36) may soon allow for the 3D imaging of large RNAs by cryo-EM tomography. More sensitive detectors allow images to be acquired with less beam intensity and thus permit multiple exposures (i.e., a tomographic tilt-series) of the same RNA molecule. These advancements will only improve our ability to characterize the structure of viral RNAs by direct visualization. Looking forward, it may be possible to incorporate cryo-EM-based measurements on branching as an additional experimental constraint within secondary structure prediction algorithms. Much as SHAPE has aided secondary structure prediction at the scale of single nucleotide base pairing, cryo-EM can produce useful information about larger-scale secondary structure features such as the degree and nature of branching.

MATERIALS AND METHODS

RNA synthesis and purification

Both RNAs were synthesized by *in vitro* transcription. After transcription and purification, RNA was equilibrated in low-ionic-strength buffer in order to promote electrostatic repulsions that facilitate the visualization of secondary structure by preventing inter-helix clumping. Yeast RNA was imaged in the absence of Mg^{2+} and STMV RNA was imaged in 5 mM Mg^{2+}. We included Mg^{2+} in the STMV buffer, because Mg^{2+} normally stabilizes any tertiary interactions, so that it might be expected to make the molecule more compact. The demonstration that the molecule is extended in the presence of magnesium confirms our previous observation that Mg^{2+} does not affect the SHAPE-restrained secondary structure of STMV RNA (11).

STMV RNA was prepared by *in vitro* transcription from an STMV DNA sequence within a pCR 2.1-TOPO plasmid, as previously described (11), then lyophilized for storage. To rehydrate and refold the STMV RNA, the lyophilized sample was added to 50 mM Hepes, 50 mM sodium acetate (pH 8), heated to 90 °C for 1 minute, cooled on ice, then equilibrated at 37 °C in the presence of 10 mM MgCl for 20 minutes. STMV RNA was then equilibrated at 4 °C for 24 h. Directly before imaging, the sample was diluted 2-fold with dd-H_2O to reduce the background noise caused by the buffer.

The 5' end of the yeast RNA studied here corresponds to the 874,269-th base of chromosome XII of *S. cerevisiae* (37). Yeast RNA was generated by *in vitro* transcription and purified as previously described (34), followed by equilibration in TE buffer at 4 °C for 24 h before imaging.

Each RNA sample was assayed for degradation after transcription as well as directly before imaging by gel electrophoresis in a 1% native agarose gel run in TAE buffer.

Cryo-EM

Both RNA molecules were imaged by conventional cryo-EM (33). Briefly, cryo-EM is performed by preparing thin films of RNA solution and rapidly cooling them to liquid nitrogen temperature. Rapid cooling vitrifies the solution and acts to lock in the structure of each tumbling, fluctuating RNA at the time of the quench. The process is so rapid that the water freezes to vitreous ice, which has the same density as the liquid; this prevents the formation of ice crystals that can disrupt the native structures of large RNAs and hinder visualization in the electron microscope.

Cryo-EM experiments were performed as described by Gopal *et al.* (34). Briefly, 3 µL of RNA solution was deposited on a Quantifoil holey carbon grid (200 mesh; R2/1) that had been previously glow-discharged. The grids were then blotted and flash-frozen by rapid plunging into liquid ethane cooled to liquid nitrogen temperature. Images of RNA molecules were recorded in transmission mode as 2D-projections. Micrographs were acquired using an FEI Tecnai G2 TF20 microscope operated at an accelerating voltage of 200 kV. Images were recorded at 3-4 microns underfocus with a TIETZ F415MP 4k × 4k pixel CCD camera. Total beam exposure was maintained between 20-40 e/Å^2.

Image analysis

RNA images were semi-manually converted to 1-pixel-wide skeletons using the NeuronJ plugin (38) within ImageJ (http://imagej.nih.gov/ij/) as described in (34). Determination of the total projected length, maximum projected length, length of each individual segment, as well as the connectivity of the skeletons was performed using the AnalyzeSkeleton plugin within ImageJ (39).

RNA secondary structure predictions

Secondary structures were predicted using three different algorithms, Mfold (16, 17), UNAfold (16, 17), and RNAstructure (21, 22), to get a sense of uncertainties in the data (Supplementary Information, Figure 6-S1). MLDs for each structure were generated using a program developed by Aron Yoffe, described in (24).

ACKNOWLEDGMENTS

We are indebted to Loren Williams for critical discussions, and to Emily Norris for the calculations of MLDs. Supported by grant R01 GM70785 to SCH from the National Institutes of Health. RFG was supported by the U. S. National Science Foundation, in the form of grant CHE 1051507 to William Gelbart and Charles Knobler. Additional funding for RFG was provided by a University of California, Los Angeles Dissertation-Year Fellowship. TEM images were obtained in the California NanoSystems Institute (CNSI) Electron Imaging Center for Nano-Machines supported by NIH (1S10RR23057).

SUPPLEMENTARY INFORMATION

Studying 3D branched objects by 2D projection imaging

Consider a simple branched object. We define its total length (TL_{3D}) as the sum contour length of all branches extended in 3D. Additionally, we define its maximum length (ML_{3D}) as the length of the longest path between any two points within its 3D structure. We use the subscript "3D" to designate that these lengths correspond to distances measured in 3D space. To illustrate these quantities we introduce Mel, the 3D stickman shown on the left side of Figure 6-S2. Mel is not truly symmetric; his left leg is slightly longer than the right, while his right arm is a little longer than the left. Ignoring his neck and head, Mel's TL_{3D} is roughly 5nm; his ML_{3D} spans his left foot to his right hand (shown in black in Figure 6-2) and is roughly 3m.

Now consider the 2D projections of the same 3D branched object. We analogously define the total length (TL_{2D}) and maximum length (ML_{2D}) of a projection as the relevant distances measured in 2D space. In Figure 6-S2 a headless Mel tumbles in 3D space, and we have plotted 4 randomly selected orientations. We immediately see that the TL_{2D} of each projection depends on Mel's orientation in 3D. Furthermore, we observe that the TL_{2D} is necessarily shorter than the

TL_{3D}, due to regions of the 3D structure that extend perpendicular to the plane of projection and are thus hidden from view. Next, we notice that the path associated with the ML_{2D} need not coincide with the ML_{3D} path (shown in black), though the ML_{2D} can be used as a crude, and usually underestimated, approximation of the ML_{3D}.

To examine how the 2D projections of 3D branched objects depend on the extent of branching, we compare the projected images of Mel to those of a second toy model with more branches. The second model, Bug, has an equal total length in 3D (TL_{3D} = 4.9nm), but a significantly shorter maximum length (ML_{3D} = 2.0nm). It has six 3-way junctions and two 4-way junctions, and the arms off the longest path have lengths between 0.3nm and 0.6nm. In Figure 6-S3A we show 64 randomly generated orientations of Mel with the corresponding 2D projections shown in Figure 6-S3B. Similarly, Figure 6-S4A shows 64 randomly generated orientations of Bug with the corresponding 2D projections shown in Figure 6-S4B. We observe how increased 3D branching affects the 2D projections by measuring the distribution of TL_{2D}, the number of third- and fourth-order branch points (V_3 and V_4, respectively), segment length, longest segment length, and ML_{2D} (Figure 6-S5).

Fractal Dimension of RNA

In an effort to quantify the absolute extendedness of each RNA sequence, the fractal dimension (D_F) (40) of every skeleton of Figure 6-3 was calculated by the traditional box counting method. Neither sequence showed D_F values significantly larger than unity, most likely because of limited resolution of the fine-branching that provides self-similarity at short length-scales. STMV RNA skeletons have an average D_F of 1.01±0.03 (SD) and yeast RNAs have an average D_F of 1.05±0.05.

FIGURES

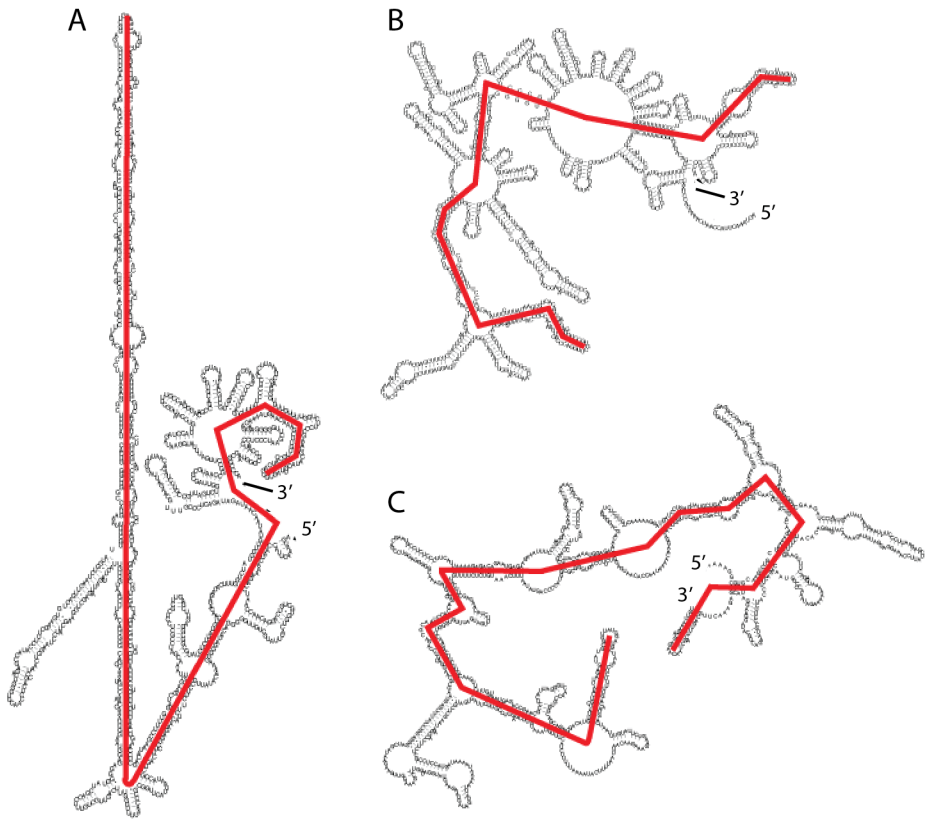

Figure 6-1. RNA secondary structure maps are shown with red lines that illustrate the paths associated with the Maximum Ladder Distance (MLD) (24) of each structure. (A) SHAPE-directed secondary structure model for free STMV RNA; MLD = 205. (B) The MFE secondary structure of free STMV RNA generated by purely thermodynamic analysis; MLD = 100. (C) The MFE secondary structure of a Yeast RNA of comparable length (see Methods) generated by purely thermodynamic analysis; MLD = 147.

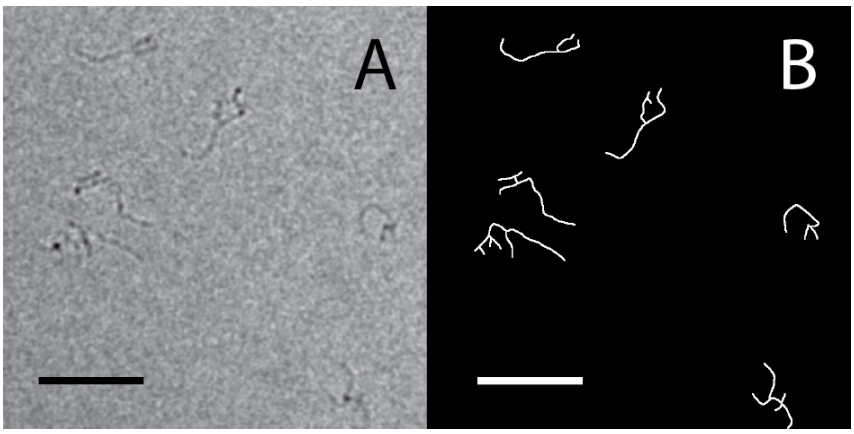

Figure 6-2. (A) Cryo-EM micrograph of STMV RNA. (B) The projected images were semi-manually converted to skeletons (see Image analysis, in MATERIALS AND METHODS) in order to analyze their secondary structure. Scale bar represents 50nm.

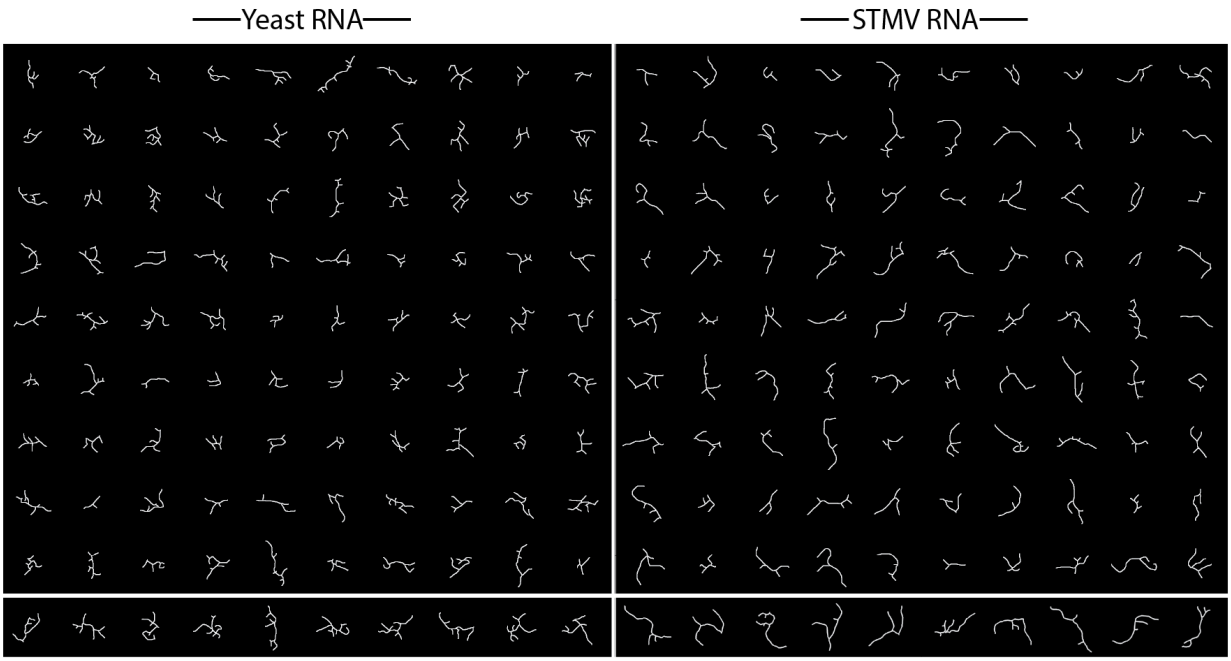

Figure 6-3. 200 skeletons of RNAs with TL_{2D} values between 25nm and 85nm. The 100 skeletons on the left were obtained from micrographs of yeast RNA, and the 100 on the right from STMV RNA. The bottom-most row shows the 10 skeletons from each sequence with the highest accepted TL_{2D} values, which are the most extended molecules in the plane of projection. To define a length scale, the skeletons have been positioned so that their centers form a simple square lattice, with 60nm taken as the distance between the centers of each nearest-neighbor pair of skeletons.

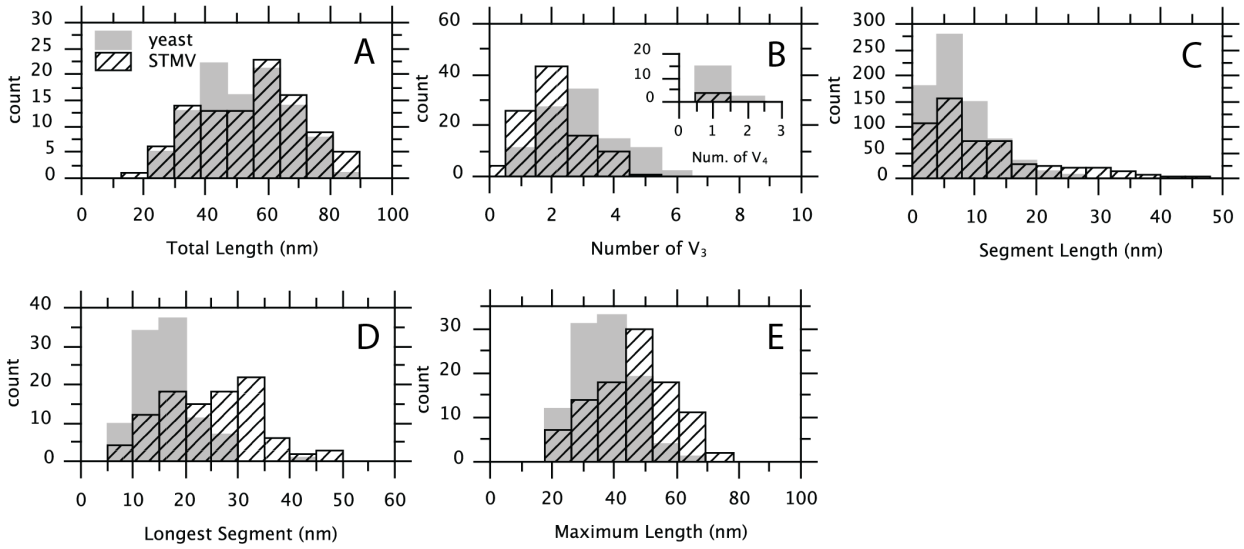

Figure 6-4. Analysis of the structures shown in Figure 6-3. Values corresponding to yeast RNA are plotted in gray, those for STMV are plotted in hatched black. (A) The distributions of total length (TL_{2D}). (B) The distributions of the number of third-order branch points (V_3) per skeleton; the inset shows the same for the number of fourth-order branch points (V_4). (C) The distributions of segment lengths. Segments are defined as the paths connecting two neighboring branch points or a branch point and its neighboring endpoint. (D) The distributions of the length of the longest segment of each skeleton. (E) Maximum length (ML_{2D}) distributions.

SUPPLEMENTARY FIGURES

Figure 6-S1. Predicted secondary structure of the control yeast RNA. The left panel shows the minimum free energy (MFE) structure as predicted by Mfold (16, 17) (MLD = 153); 29 suboptimal structures from Mfold gave MLDs ranging from 98-155. The MFE structure from UNAfold (16, 17) has an MLD of 147. The MFE structure from RNAstructure (21, 22) (MLD = 127) is shown at the right; 20 different predictions generated by RNAstructure using Boltzmann sampling gave MLDs ranging from 120-140. The expected MLD for an ensemble of random sequences of the same length is 138, by Eq. (1).

Figure 6-S2. A 3D stickman, Mel, is boxed o the left. Each of Mel's limbs, and his torso, measures approximately 1nm in length (TL_{3D} = 4.9nm). Because Mel's right arm is slightly longer than his left, and his left leg is a little longer than his right, he has a unique maximum path length (ML_{3D} = 2.8nm), shown in black. Four representative projections of Mel (excluding his head and neck) in random orientations are shown.

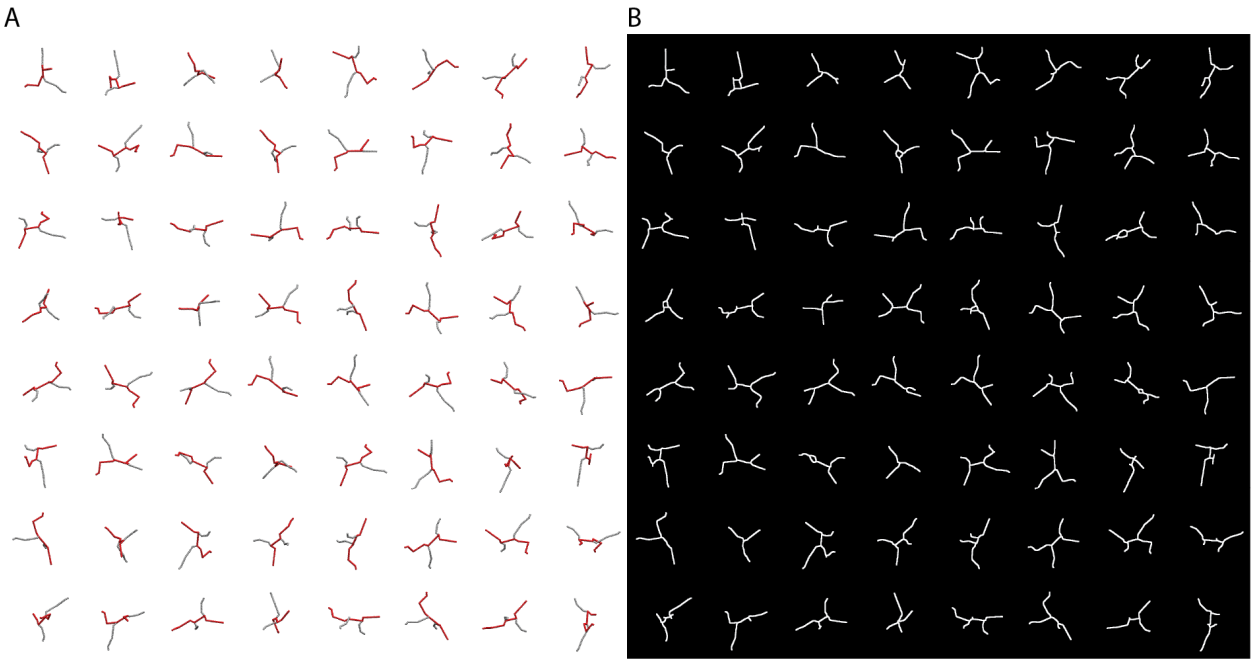

Figure 6-S3. (A) 64 randomly generated orientations of the 3D stickman, Mel (excluding his head and neck). Here, his ML_{3D} is shown in red. (B) Each of the 64 orientations has been converted to a 2D-projected path skeleton.

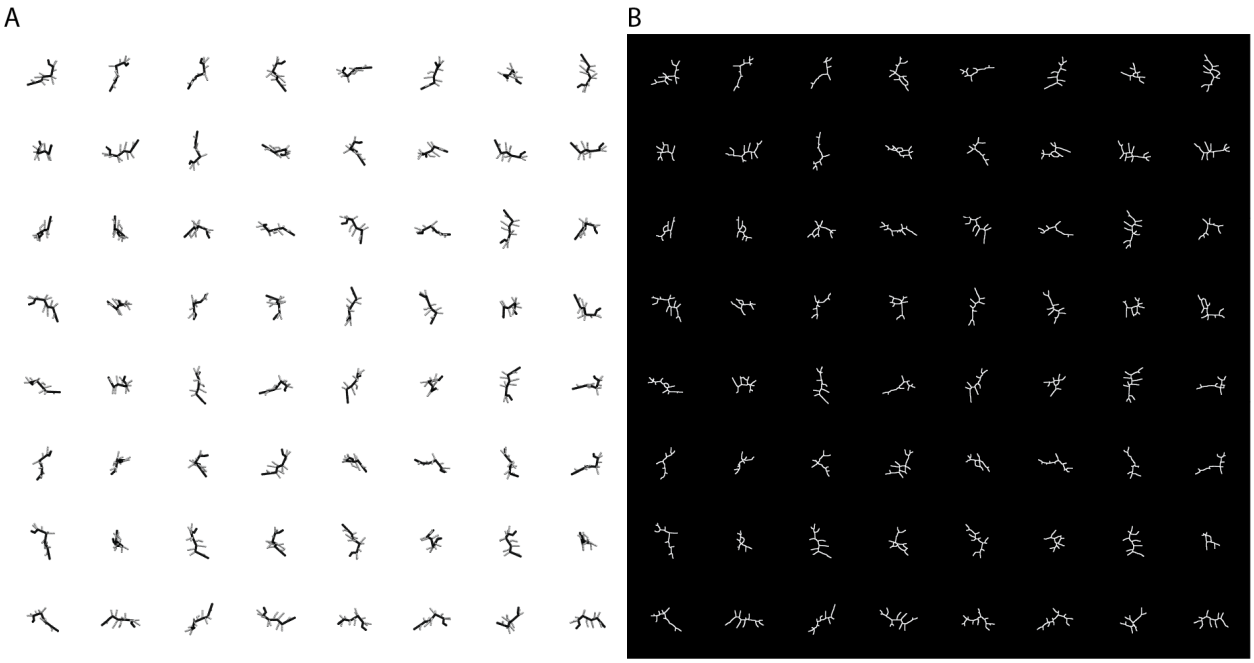

Figure 6-S4. (A) 64 randomly generated orientations of Bug, a branched object of equal total length as Mel (TL_{3D} = 4.9nm), but with significantly more branches. Bug's ML_{3D} is shown in black, and measures 2.0nm. (B) Each of the 64 orientations has been converted to a 2D-projected path skeleton.

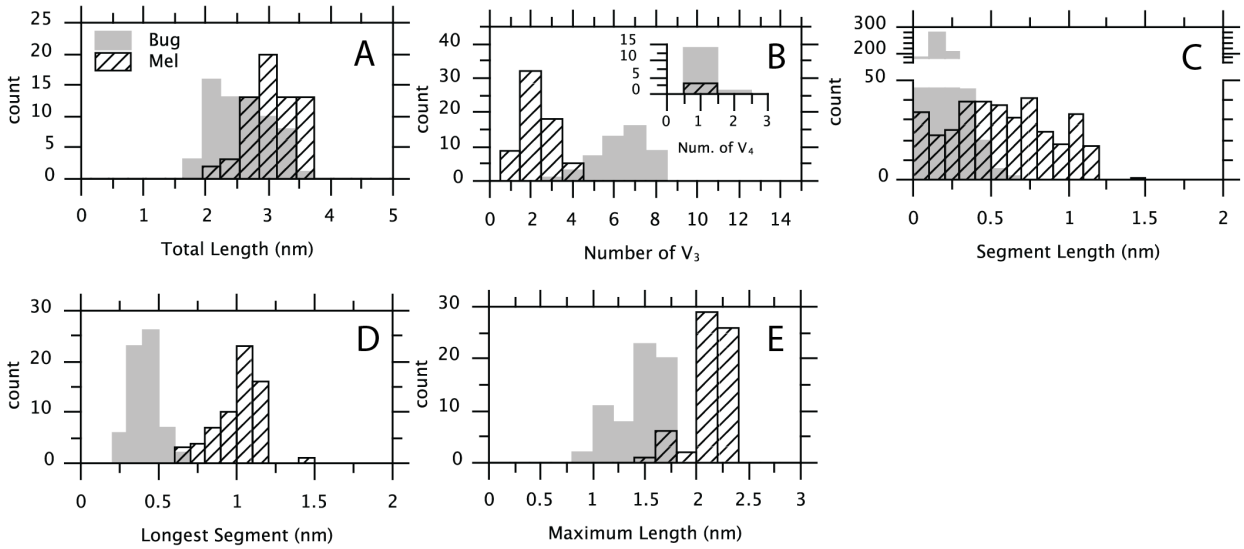

Figure 6-S5. Analysis of the structures shown in Figure 6-S3 and Figure 6-S4. Values corresponding to Bug are plotted in gray, those for Mel are plotted in hatched black. (A) The distributions of total length (TL_{2D}). (B) The distributions of the number of third-order branch points (V_3) per skeleton; the inset shows the number of fourth-order branch points (V_4). (C) The distributions of segment lengths. Segments are defined as the paths spanning two neighboring branch points or a branch point and its neighboring endpoint. (D) The distributions of the length of each skeleton's longest segment. (E) Maximum length (ML_{2D}) distributions.

REFERENCES

1. Choi Y, Dreher T, & Rao A (2002) tRNA elements mediate the assembly of an icosahedral RNA virus. *Proc Natl Acad Sci U S A* 99(2):655-660.

2. Valegård K, Murray JB, Stockley PG, Stonehouse NJ, & Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371(6498):623-626.

3. Bunka DHJ, *et al.* (2011) Degenerate RNA Packaging Signals in the Genome of Satellite Tobacco Necrosis Virus: Implications for the Assembly of a T=1 Capsid. *J. Mol. Biol.* 413(1):51-65.

4. Qu F & Morris T (1997) Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J. Virol.* 71(2):1428-1435.

5. Damgaard C, Andersen E, Knudsen B, Gorodkin J, & Kjems J (2004) RNA interactions in the 5′ region of the HIV-1 genome. *J. Mol. Biol.* 336(2):369-379.

6. Frolov I, Hardy R, & Rice C (2001) Cis-acting RNA elements at the 5'end of Sindbis virus genome RNA regulate minus-and plus-strand RNA synthesis. *RNA* 7(11):1638-1651.

7. Lindenbach BD, Sgro J-Y, & Ahlquist P (2002) Long-Distance Base Pairing in Flock House Virus RNA1 Regulates Subgenomic RNA3 Synthesis and RNA2 Replication. *J. Virol.* 76(8):3905-3919.

8. Barends S, Bink H, van den Worm S, Pleij C, & Kraal B (2003) Entrapping ribosomes for viral translation: tRNA mimicry as a molecular Trojan horse. *Cell* 112(1):123-129.

9. Watts JM, *et al.* (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460(7256):711-716.

10. Stockley P, *et al.* (2007) A Simple, RNA-Mediated Allosteric Switch Controls the Pathway to Formation of a T= 3 Viral Capsid. *J. Mol. Biol.* 369(2):541-552.

11. Athavale SS, *et al.* (2013) In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One* 8(1):e54384.

12. Tang L, *et al.* (2001) The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nature Struct Biol* 8:77-83.

13. Toropova K, Basnak G, Twarock R, Stockley P, & Ranson N (2008) The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* 375(3):824-836.

14. Larson SB*, et al.* (1993) Double-helical RNA in satellite tobacco mosaic virus. *Nature* 361(6408):179-182.

15. Larson SB, Day J, Greenwood A, & McPherson A (1998) Refined structure of satellite tobacco mosaic virus at 1.8Å resolution. *J Mol Biol* 277(1):37-59.

16. Zuker M (1989) On finding all suboptimal foldings of an RNA molecule. *Science* 244(4900):48-52.

17. Markham NR, Zuker M, Mathews DH, & Turner DH (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 453(11):3-31.

18. Hofacker IL*, et al.* (1994) Fast folding and comparison of RNA secondary structures. *Monatsch Chemie* 125:167-188.

19. Wuchty S, Fontana W, Hofacker IL, & Schuser P (1999) Complete Suboptimal Folding of RNA and the Stability of Secondary Structures. *Biopolymers* 49:145-165.

20. Swenson MS*, et al.* (2012) GTfold: Enabling parallel RNA secondary structure prediction on multi-core desktops. *BMC Res Notes* 5(1):341.

21. Bellaousov S, Reuter JS, Seetin MG, & Mathews DH (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* 41(Web Server issue):W471-474.

22. Reuter JS & Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129.

23. Turner DH & Mathews DH (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res* 38(Database issue):D280-282.

24. Yoffe AM*, et al.* (2008) Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci U S A* 105:16153-16158.

25. Low JT & Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods* 52(2):150-158.

26. Deigan K, Li T, Mathews D, & Weeks K (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* 106(1):97-102.

27. Archer EJ*, et al.* (2013) Long-range architecture in a viral RNA genome. *Biochemistry* 52(18):3182-3190.

28. Wu B, Grigull J, Ore MO, Morin S, & White KA (2013) Global Organization of a Positive-strand RNA Virus Genome. *PLoS Pathog* 9(5):e1003363.

29. Wilkinson KA, Merino EJ, & Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 1(3):1610-1616.

30. Day J, Kuznetsov Y, Larson S, Greenwood A, & McPherson A (2001) Biophysical studies on the RNA cores of satellite tobacco mosaic virus. *Biophys. J.* 80(5):2364-2371.

31. Kuznetsov Y, Daijogo S, Zhou J, Semler B, & McPherson A (2005) Atomic force microscopy analysis of icosahedral virus RNA. *J. Mol. Biol.* 347(1):41-52.

32. Giro A*, et al.* (2005) Single molecule studies of RNA secondary structure: AFM of TYMV viral RNA. *Microsc. Res. Tech.* 65(4-5):235-245.

33. Frank J (2009) Single-particle reconstruction of biological macromolecules in electron microscopy--30 years. *Q Rev Biophys* 42(3):139-158.

34. Gopal A, Zhou ZH, Knobler CM, & Gelbart WM (2012) Visualizing large RNA molecules in solution. *RNA* 18(2):284-299.

35. Bundschuh R & Hwa T (2002) Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E* 65(3):031903.

36. Faruqi A & Henderson R (2007) Electronic detectors for electron microscopy. *Curr. Opin. Struct. Biol.* 17(5):549-555.

37. Cherry JM*, et al.* (1997) Genetic and physical maps of Saccharomyces cerevisiae. *Nature* 387(6632):67-73.

38. Meijering E*, et al.* (2004) Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A* 58(2):167-176.

39. Arganda‐Carreras I, Fernández‐González R, Muñoz‐Barrutia A, & Ortiz‐De‐Solorzano C (2010) 3D reconstruction of histological sections: Application to mammary gland tissue. *Microsc. Res. Tech.* 73(11):1019-1029.

40. Mandelbrot BB (1982) *The Fractal Geometry of Nature* (Freeman, San Francisco).

Summary and additional preliminary results

In summary of Chapters 1-3, we have shown that any length or sequence of ss-RNA can be efficiently packaged by CCMV CP so long as: (i) the concentration of basic residues brought by the N-termini of the CP is sufficient to neutralize the negative charge brought by the phosphate backbone of the RNA, and (ii) the assembly is carried out in two steps, first at neutral pH and low ionic-strength (to drive CP-RNA interaction), then at low pH and low ionic-strength (to drive capsid formation). If condition (i) is not met, a fraction of the RNA will remain unpackaged. If condition (ii) is not met by neglecting the acidification step, the strength of CP-CP interactions will be insufficient to nucleate capsid growth. Alternatively, if condition (ii) is not met by neglecting the neutral-pH step, kinetically trapped aggregate structures are formed consisting of many nucleated capsids strung together by numerous RNA strands. We speculate that aggregation is caused by an effective RNA-RNA attraction between RNAs that are partially decorated by CP. Because CP-RNA and CP-CP interactions are simultaneously turned-on, partially decorated RNAs will effectively attract each other through the direct interaction of their bound CPs – *e.g.,* a CP bound to RNA molecule *i* may attract a CP bound to RNA molecule *j*. Such aggregation should be suppressed if the absolute concentration of RNA was low enough to prevent RNA-RNA collisions. Consistent with this scenario, recent preliminary data suggest that well-formed VLPs may be generated in a single step at mild acidity (pH 6) and an absolute concentration of 3 nM RNA (compared to 30 nM used throughout Chapters 1-3).

The discovery of multiplet capsids (Chapter 1), defined as clusters of essentially complete capsids with a single RNA molecule shared between them, inspired us to consider the possibility of generating a new construct, the “cherry-bomb”, consisting of a single VLP with a

portion of the RNA poking out through a small defect (or a natural pore) in the capsid. Such a construct would be interesting for a number of reasons. Most straightforwardly, this construct would make possible, for the first time, a single-molecule force pulling experiment to measure the force required for extracting the ss-RNA genome from a spherical virus. Such a measurement would test the proposed mechanism of co-translational disassembly (1), in which a host ribosome actively pulls the RNA of certain positive-sense viruses from their capsid during the initial translation event.

To generate the cherry bomb, a 200-nt DNA oligo was designed to be complimentary to the terminal 190 nucleotides (nt) of the 5'-end of BMV RNA1. The remaining 10 nts, located at the 3'-end of the DNA oligo, were designed as a 10-mer of poly-A. This DNA oligo was hybridized to BMV RNA1 in TE buffer with 150 mM NaCl by thermal annealing. The predicted structure of this RNA-DNA hybrid molecule (Figure 7-1-2) is roughly 3,000 nt of BMV RNA1 with a 190-bp DNA-RNA hybrid arm protruding from its 5'-end. Because 190 base pairs of double-stranded nucleic acid is relatively stiff (persistence length of the order of 50 nm), and roughly 46-65 nm in length (depending on whether one assumes the rise of double-stranded A-form RNA (0.24 nm/bp) or double-stranded B-form DNA (0.34 nm/bp)), we expect that this DNA-RNA hybrid portion will be too stiff and too long to be packaged into the 28 nm CCMV capsid (Figure 7-1-3).

We carried out the normal two-step assembly protocol using this DNA-RNA hybrid, and imaged the reaction products by negative-stained-EM. Some preliminary images are shown in Figure 3. Additionally, a fluorescent poly-T oligo was used to ascertain that a significant fraction of the assembled VLPs had a portion of DNA-RNA hybrid poking out. We are currently working to perform the first force pulling experiments on this new construct.

FIGURES

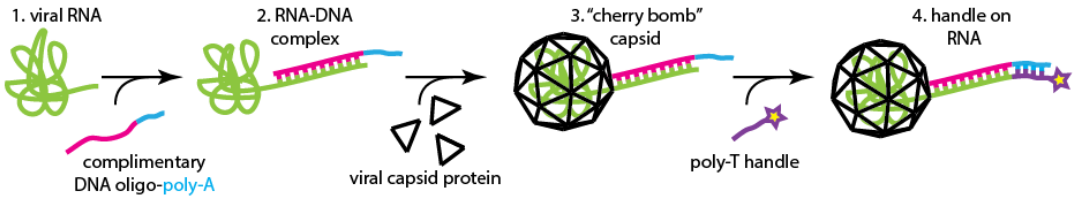

Figure 7-1. Assembly strategy for generating the cherry-bomb construct. (1) BMV RNA1 is shown in green. The DNA oligo is shown with red representing the 190-nt portion of the 3'-end that is complementary to BMV RNA1, and blue representing the 3' 10-mer of poly-A. (2) BMV RNA1 and the DNA oligo are hybridized. (3) The DNA-RNA hybrid molecule is subjected to the normal packaging protocol. We expect that the DNA-RNA hybrid portion is too stiff and too long to fit within the capsid of CCMV. (4) The terminal 10-mer of poly-A may serve as a handle for the DNA-RNA hybrid portion protruding from the capsid.

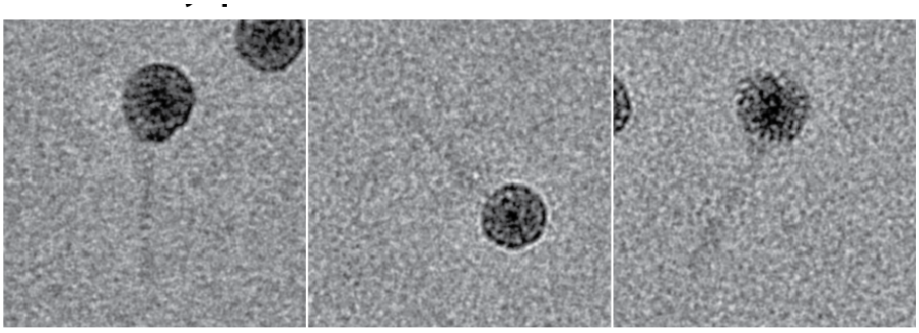

Figure 7-2. Negative-stain electron micrographs showing the cherry bomb constructs. To judge the scale, each capsid is 28 nm in diameter.

REFERENCES

1. Wilson, T.M. 1984. Cotranslational disassembly of tobacco mosaic virus in vitro. Virology. 137(2):255-65.