Title
In vitro studies of single-stranded RNA virus self-assembly

Permalink
https://escholarship.org/uc/item/7t98443c

Author
Comas Garcia, Mauricio

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

In vitro studies of single-stranded RNA virus self-assembly

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Mauricio Comas Garcia

2013

ABSTRACT OF THE DISSERATION

In vitro studies of single-stranded RNA virus self-assembly

by

Mauricio Comas Garcia

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2013

Professor William M. Gelbart, Chair

Viruses are self-replicating nucleoprotein complexes that have the ability to cause diseases and are composed of, at least, a single copy of genetic material (either RNA or DNA) and a virally encoded proteinaceous (capsid protein [CP]) protective shell. In single-stranded RNA (ss-RNA) viruses genome packaging is unequivocally coupled to capsid assembly. This spontaneous process is driven by CP-CP and CP-RNA interactions. This thesis is concerned with the in vitro self-assembly of a ss-RNA virus, cowpea chlorotic mottle virus (CCMV), and the effects of the length of the nucleic acid on capsid assembly. We show that the size of a capsid depends not only on the preferred curvature of the CP, but also on the length and number of packaged RNA molecules. By carrying out a series of titration curves of CP with RNAs of varying lengths we find that such high encapsidation yield is possible if and only if the assembly reaction is carried out under very specific experimental conditions, namely at a CP:RNA mass-ratio of 6:1 ("Magic ratio"). By doing head-to-head competition experiments, between a reference and a competitor

RNA, under limiting assembly conditions we show that the relative packaging efficiency of ssRNA by CCMV CP is a non-monotonic function of RNA length. The relative packaging efficiency increases as the length of the competitor RNAs approaches that of the wild-type RNA (3.2 kilobases [kb]), while for equal-length RNAs it depends on the secondary and tertiary structure of the RNAs. We demonstrate that the length and number of packaged RNAs determines the assembly pathway; unlike the case of long RNAs (> 1.0 kb), when multiple short RNAs (0.5 kb) are packaged by CCMV CP, to virus assembly is a highly cooperative process. This is the first study in which it is shown that in contrast to the accepted dogma the degree of cooperativity of virus assembly depends not only on the relative strength of the CP-CP and CP-RNA complexes but also on the length and number of packaged RNA molecules.

The dissertation of Mauricio Comas Garcia is approved.

Miguel Garcia-Garibay

Thomas G. Mason

Ayala L. N. Rao

William M. Gelbart, Committee Chair

2013

To my beloved grandparents Engracia, Jose, Amparo y Jose Manuel

and to my granduncles Luis, Olga and Conchita

because they never gave up and were brave enough

to cross the Atlantic and start a new life

with nothing but their dreams.

To my parents and to my brother for being the light at the end of the tunnel.

TABLE OF CONTENTS

<div align="center">VITA/BIOGRAPHICAL SKETCH</div>

EDUCATION

- *Autonomous University of San Luis Potosi, Mexico (UASLP)*, Department of Chemistry; B.Sc. in Chemistry (2002-2007).

EMPLOYMENT HISTORY

- *University of California, Los Angeles (UCLA)*. Graduate Student Researcher, September 2007 – present.

- *UCLA*. Teaching Assistant and Graduate Student Researcher, September 2007 – June 2012.

- *UASLP*, Physics Institute. Physical Chemistry Lab. Assistant; January 2005 – June 2007 (Mexico).

- Producer and announcer of the popular science radio program *"Life is Chemistry"* at the *Autonomous University of San Luis Potosi Radio Station*; January 2004 – September 2006 (Mexico).

- Assistant of the (undergrad) Chemistry advisor at *UASLP*; January 2004 – January 2005 (Mexico).

AWARDS/DISTINCTIONS

- *UCLA* Department of Chemistry and Biochemistry, George Gregory Award for excellence in research in Physical Chemistry, November 19th 2012.

- *UCLA* Dissertation Year Fellowship 2012 – 2013.

- *UCLA* Department of Chemistry and Biochemistry, Teaching Assistant award 2009.

- Doctoral CONACyT Fellowship 2007-2012 (Mexican Government).

- *UCLA* Chemistry and Biochemistry Department, First Year Fellowship 2007 - 2008.

- 2nd Place in the "IV Contest of Project development of the Chemistry Major", Department of Chemistry, *UASLP*. With the title: "Air/water interfacial study of the Cowpea Chlorotic Mottle Virus: Phase diagram determination and observation of meso/micro viral structures". November 24th 2006.

- Best GPA of the 2002-2006 Class, Chemistry Major, *UASLP*.

- Mexican Institute of Chemical Engineering Award as the Best Chemistry Major Student, 2002 Generation. November 23th 2006.

- *UASLP* Chemistry Science Department Best Student Award; November 7th 2005.
- *UASLP* Chemistry Science Department Best Student Award; November 11th 2004.
- *UASLP* Chemistry Science Department Best Student Award; November 2003.

PUBLICATIONS

- 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 efficiency of a single-stranded RNA molecules by viral capsid protein. *Journal of Virology*. Volume 86, Number 22, 12271-12282 (2012)

- 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. *Journal of Virology*. Volume 86, Number 6, 3318-3326 (2012)

TEACHING ASSISTANT AT UCLA:

- Spring Quarter 2012: Chen 154 (Biochemistry Methods II) with Dr. S. Nakamoto.
- Fall Quarter 2011 and 2009: Chem 110A (Physical Chemistry; Chemical Thermodynamics) with Prof. W. M. Gelbart.
- Spring Quarter 2011 and 2010: Chem 156 (Physical Biochemistry) with Prof. M. Quinlan.
- Fall Quarter 2010: Chem 156 (Physical Biochemistry) with Prof. T. O. Yates.
- Summer Session C 2011: Chem 14D (Organic Chemistry and pharmaceuticals) with Dr. C. Chanmugathas.
- Summer Session A 2010: Chem 156 (Physical Biochemistry) with Prof. P. Zavodszky.
- Winter Quarter 2010, 2009 and 2008: Chem 14B (Thermodynamics, Kinetics, Electrochemistry and Organic Chemistry) with Dr. L. Lavelle.
- Fall Quarter 2008: Chem 14A (Atomic and Molecular Structure, Equilibria, Acids and Bases) with Dr. L. Lavelle.
- Spring Quarter 2008: Chem 110A (Physical Chemistry; Chemical Thermodynamics) with Prof. D. Neuhauser.
- Fall Quarter 2007: Chem 20A (Chemical Structure) with Dr. E. Scerri.

Chapter I

Introduction

Viruses are self-replicating nucleoprotein complexes that have the ability to cause diseases; with only a few exceptions they are composed of, at least, a single copy of genetic material (either RNA or DNA) and a virally encoded protective shell. H.R. Gelderblom defined them as mobile genetic elements, most probably of cellular origin and characterized by a long co-evolution of virus and host[1]. Viruses are infectious, obligate molecular parasites, do not produce their own energy (i.e., ATP), respire, grow or move[2]. Most viruses are ~ 100 times smaller than a bacteria $(0.5 - 5 \; \mu m)$ and their protective shell is called a capsid and when enclosing its genetic material, a nucleocapsid; some giant viruses can have several concentric shells. Capsids can be classified as symmetric (icosahedral or helical) and pleomorphic (not characterized by a single size or shape). Complex viruses also contain a host-derived lipid bilayer (envelope), spike proteins (glycoproteins), and host- and/or virally derived enzymes. Fig. 1 shows cryo-EM reconstructions of a double-stranded DNA (ds-DNA) bacterial virus (bacteriophage), a non-enveloped single-stranded RNA (ss-RNA) bacterial virus, an enveloped ss-RNA animal virus and a retrovirus. Viruses infect all of the life kingdoms and can retain their infectivity even outside their host and/or vector, even under very extreme conditions. All of these characteristics highlight the great diversity in the virus world, making generalizations almost impossible and opening the doors to novel biochemical processes and therefore creating new paradigms.

A particular paradigm that has been challenged is the definition of life and inanimate objects. A. Rein recently proposed that a virus can be viewed as a rather regular, relatively simple physical object. Alternatively, it can be seen as a living organism, evolving in response to

selective pressure[3]. This duality concept is equivalent to that of light; depending on the measured property it behaves like a wave or like a particle. Viruses present such duality, like all "living organisms" they react to ecological and evolutionary pressures to ensure successful replication of their genes. On one hand, they have evolve to have a fast mutation rate to avoid been detected by the host immune response system (i.e., influenza)[4] and/or incorporate genes from other viruses and/or hosts (horizontal gene transfer)[5]. On the other hand, they can *"survive"* outside of their host, they can be disassembled and further purified into their elemental constituents and then reassembled into infectious particles[6]. They can be chemically modified and used for macromolecular assemblies in the absence of their genetic material, making them valuable tools for biotechnology[7-10].
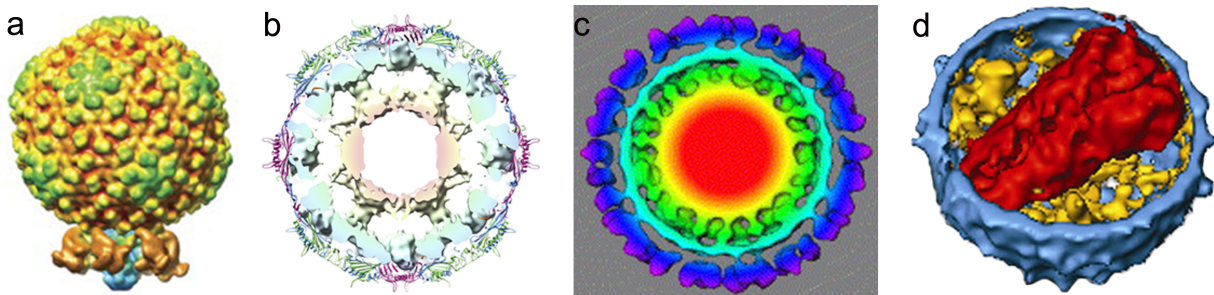

FIG 1. Cryo-EM reconstruction of four viruses a) ds-DNA icosahedral bacteriophage epsilon15; the packaging motor (brown and blue) is bound to the capsid at a pentamer[11]. b) The cross-section of the reconstruction of the ss-RNA bacteriophage MS2 was superimposed on the crystal structure of the capsid. The solid turquoise color represents the electron density of the disordered part of the capsid protein bound to the RNA and the yellow-red part is the RNA that is not in contact with the capsid protein[12]. c) Cross-section of a low-resolution reconstruction of Sindbis virus which is a ssRNA enveloped virus. This reconstruction shows the glycoproteins (dark blue), the host-derived envelope (turquoise) and the capsid protein (green)[13]. d) Cross section of a cryo-EM tomograph of HIV. As for all retroviruses it is an enveloped virus, the red structure represents the cone-like capsid, which contains the genomic ssRNA[14].

Virus classification

Given the incredible diversity of virus composition, structure and biochemical properties, it is not possible to describe them by a taxonomic system equivalent to the one used for traditional life kingdoms, therefore a great deal of work has been done on trying to find a systematic way

that better describes them. As a result several criteria have been proposed for the classification of viruses:

a) On the basis of their evolutionary relationships with other viruses: *ICTV classification*[15].

b) On the folding motifs of their main capsid protein[16].

c) By a combination of their nucleic acid, strandedness (double or single), sense and method replication: *Baltimore classification*[17].

d) Based on chemical and physical characters like nucleic acid, symmetry, presence of envelope, diameter of the capsid and number of capsomers: *LHT system*[18].

The ICTV classification (International Committee of Taxonomy of Viruses) shares many features with the classification system of cellular organisms and is based on the assumption that virus families have evolved from a common ancestor. David Baltimore's classification system divides viruses into seven groups based on the way they go from their original genomic material to their messenger RNAs (see Fig. 2). The LHT system proposed in 1962 is based on a set of structural characteristics, which might be useful if one aims to understand the physical aspects of viruses. However, from the point of view of physical virology all of these classifications are far too complicated, making it almost impossible to find general rules that explain the assembly of viruses. Abrescia and co-workers[16] proposed a simpler and elegant scheme by realizing that they could build a structured-based phylogenetic tree showing the four different lineages based on the folding motif of the viral capsid proteins (PRD1-, Picorna-, HK967- and BTV-like foldings). Unfortunately, this classification includes within the same viral lineage double- and single-stranded viruses. While all these classifications are extremely useful when looking at the

structure of the capsids, evolution or biochemical properties they are either far too complicated or not appropriate when trying to understand virus assembly mechanisms.
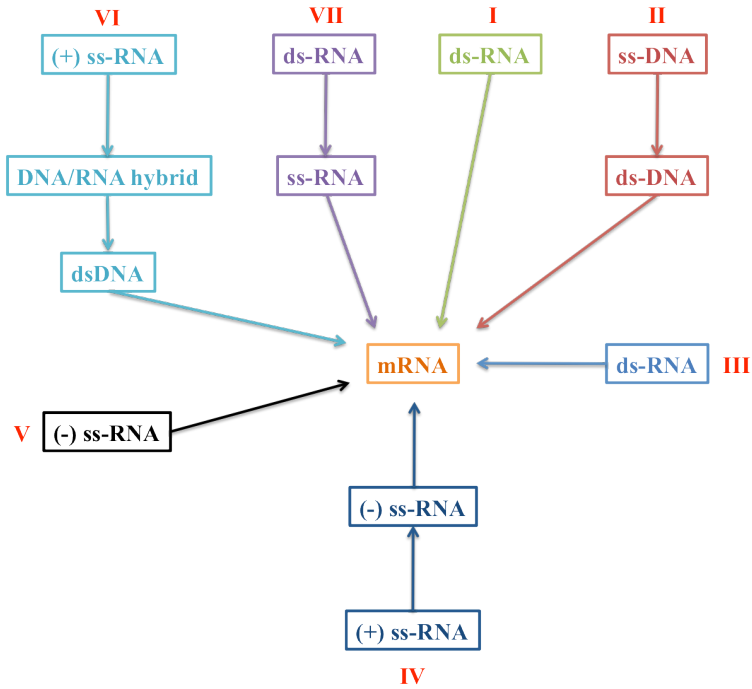

FIG 2. The Baltimore classification is a system that places viruses into seven groups depending on how their replication method and the strandedness of the encapsidated viral genome in the mature infectious particle.

A very simple (and drastic) way to classify viruses from an assembly point of view is by grouping them whether on their genome is either single- or double-stranded *during encapsidation*. This extremely simple classification is independent of post-encapsidation processes in which the genome might go from single- to double-stranded encapsidation, i.e. Hepatitis B and HIV. Independent of the chemical nature of their nucleic acid (DNA or RNA) we can merge groups II, IV, V and VI from Baltimore's classification into a single-stranded virus category and groups I, II and VII into a double-stranded one:

    a) ***Double-stranded*** genome packaging involves the formation of empty capsids and

        subsequent translocation of the DNA or RNA into the capsid by a virally-encoded

enzyme. The latter process requires chemical energy, i.e. ATP, and the performance of work.

b) ***Single-stranded*** genome capsid assembly and genome packaging are coupled to each other, and are spontaneous, not involving ATP.

Single-stranded viruses

The first ever *in vitro* reconstituted virus was a cylindrical ss-RNA virus: tobacco mosaic virus (TMV)[19]. Not long after, the first icosahedral *in vitro* reconstituted virus was also an ss-RNA one: cowpea chlorotic mottle virus (CCMV)[20]. It was no coincidence that both are ssRNA viruses, in fact this was possible for two main reasons; ss-RNA viruses are extremely simple viruses (their capsids are made up multiple copies of a single gene product – the capsid protein [CP]) and they spontaneously self-assemble into RNA-containing capsids under specific solution conditions without the need of chemical energy, i.e. ATP.

In contrast to ds-DNA viruses, capsid assembly and genome packaging in ss-RNA viruses are unequivocally coupled processes. With a few exceptions (i.e. cowpea mosaic virus [CPMV]) no empty capsids are formed during the viral cycle, hence it is difficult to decouple assembly and packaging from each other. The main force driving virion assembly is the electrostatic interaction between the highly-negatively charged RNA phosphate backbone and the highly-positively charged interior of the CP. In general there are two main ways the CP interacts with the RNA, either by a structural disordered arginine-rich motif (ARM) that protrudes into the interior of the capsid or by a positively-charged interior surface[21] (see Figure 3). The latter mechanism is less common and it is used mostly by ssRNA bacteriophages like

MS2. In the case of ARM-bearing CPs, the basic residues are essential for nucleic acid encapsidation, as shown by the inability of mutant CPs with deletions in the ARM region to encapsidate RNA; moreover such deletions promote T = 1 empty capsid assembly[22,23].

The role of the CPs in ss-RNA viral life cycles goes beyond just forming a protective shell; they have been associated with the switching from minus- to plus-strand synthesis[24], suppression of RNA silencing[25] and in plant viruses they are involved in efficient cell-to-cell movement[26]. As mentioned above, although it is hard to uncouple packaging from assembly some viruses like CCMV and brome mosaic virus (BMV) can form empty shells in the absence of RNA under non-physiological conditions. This ability to form empty capsids has been exploited to understand capsid assembly[27], though in some cases it was not explicitly acknowledged that this process might be different in the presence of the RNA. While this approach was extremely helpful for understanding the physics of assembly, few viruses form empty capsid during their infectious cycle and it therefore has no biological significance for ss-RNA virus assembly.

Recent theoretical studies and extensive computational simulations have shown that the assembly pathway for ss-RNA viruses greatly depends on the relative strength of the CP-CP and CP-RNA interactions[28,29]. In a particularly appealing work Elrad and Hagan[30] explored two limiting scenarios depending on the strength of these interactions: (i) nucleation of a capsid "embryo" involving a small number of proteins, followed by growth; and (ii) an "*en masse*" adsorption of protein subunits onto the nucleic acid that approaches or exceeds the stoichiometric number required for assembly, followed by a cooperative rearrangement to form a capsid. The

nucleation pathway is favored for strong CP-CP and weak CP-RNA interactions, while the "*en masse*" scenario is expected for the opposite case. In this work no specific interactions between the RNA and the CP were assumed, which is consistent with a large set of experimental work in which it was shown that a viral CP can encapsidate non-viral RNAs *in vitro* and *in vivo*. The fact that a CP can package a non-viral RNA *in vitro* is a reflection that under specific CP and RNA concentrations and buffer conditions the most energetically favorable structure is a capsid. However this is normally not the case during an infection in which only the viral RNA is packaged. It has been argued that this difference arises because *in vivo* assembly occurs in specialized compartments where there is only viral RNA[31,32] or because the CPs associate specifically with the RNA-dependent RNA polymerase and only package nascent copies of the viral genome[33]. A rather different view has been taken by the group of Peter Stockely in which they have studied viruses with a very high packaging selectivity even in *in vitro* conditions. Their work has shown that faithful and efficient assembly depends on the presence of specific RNA sequences ("packaging signals"), which in some cases are not necessary if the CP concentration is sufficiently high[34,35].

Even though ss-RNA viruses were the first to be reassembled *in vitro*, it has not been until recently that significant efforts have been addressed to understand their assembly pathways. This thesis focuses on understanding the general principles for RNA encapsidation and virus assembly by systematically studying the effects on virus assembly of the nucleotide length and sequence of RNA. Here we will not address the effects of packaging signals that are known to play a crucial role during *in vivo* assembly[34,35]; nonetheless the results exposed here highlight the importance of the RNA as a catalyst and scaffold, inducing capsid sizes different from the one

preferred by the CP and by inducing new assembly pathways. We also show how the packaging efficiency depends on the length of the RNA when sequence-specific interactions are not dominant. The main goal of this thesis is to understand the dominant interactions that give arise to virus efficient and faithful assembly in the presence of ss-RNA best in the absence of any specific interactions.

Chapters two and three were published in 2012 in the *Journal of Virology* and chapter four is ready to be submitted to the same journal. Chapter two deals with the self-assembly of CCMV CP with RNA molecules of different sizes[36]. In this study we examined the self-assembly of CCMV CP with RNA molecules ranging in length from 140 to 12,000 nucleotides, showing that (i) even though the CP has a preferred curvature associated with a particular size and triangulation number (for review on the concept of triangulation number read *Morgan, J.G. (2003) Historical review: Viruses, crystals and geodesic domes. Trends in biochemical Sciences, 28(2):86-90*), the length of RNA and the number of packaged molecules affect the size of a capsid; (ii) each of these RNAs is completely packaged if equal RNA and N-terminal-protein charges are provided in the assembly reaction ("*Magic ratio*").

Chapter three focuses on the *in vitro* quantification of the relative packaging efficiency of ssRNA molecules of varying lengths by CCMV CP[37]. It is the first systematic study of the *in vitro* packaging efficiency of ssRNA by a viral CP as a function of RNA length; we showed that the relative packaging efficiency is a non-monotonic function with its maximum at the length of the viral RNA. We found also, unexpectedly, that BMV RNA 1 (B1) had a higher efficiency of being packaged by CCMV CP than does CCMV RNA 1 (C1), suggesting that either B1 might contain a series of cryptic signals with high-affinity for CCMV CP or that B1 contains more of

these small sequences than C1 (or that B1 was a better size and shape for acting as a template for capsid nucleation and growth). This surprising outcome resulted in a collaboration with Prof. Peter G. Stockley to perform a series of SELEX experiments with both, CCMV and BMV CP's that we might determine if there are high-affinity binding sequences for these CP's.

Chapter 3 is aimed at understanding the assembly pathway of CCMV CP around short RNAs. This project was inspired by some of the results in chapter 2, as well as the *in vivo* scenario in which viruses from the *Bromoviridae* family copackage their subgenomic RNA and the shortest of the viral RNAs. This study showed that unlike long RNAs, when multiple copies of a short RNA are packaged the assembly pathway changes from a non-cooperative to a cooperative assembly pathway. By using a single-molecule technique we confirmed that four 500 nucleotides (nt). RNAs were packaged per capsid.

The last part of this thesis includes a summary of the findings and two appendixes. Appendix I describes a labeling protocol for CCMV CP by covalently linking solvent-exposed lysines with a fluorophore; Appendix II focuses on the technical aspects of single-molecule fluorescence correlation spectroscopy.

REFERENCES

1. **Gelderblom, H.R**. 1996. Medical microbiology; Chapter 24th Structure and Classification of viruses. Baron S. Editor

2. **Mishra, R**. 2004. Virus and Plant diseases, pp. 12. Discovery publishing House page 12)

3. **Rein, A**. 2011. Murine Leukemia Viruses: Objects and Organisms. Adv. Virol. **2011**:1-14

4. **Parvin, J.D., Moscona, A., Pan, W.T., Leider, J.M. and Palese, P**. 1986. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. J. Virol. **59**(2):377-383

5. **Benarroch, D., Claverie, J-M., Raoult, D. and Shuman, S.** 2006. Characterization of Mimivirus DNA Topoisomerase IB Suggest Horizontal Gene Transfer between Eukaryal Viruses and Bacteria. J. Virol. **80**(1):314-321

6. **Bancroft, J.B.** 1972. A virus made from parts of genomes of brome mosaic and cowpea chlorotic mottle viruses. J. gen. Virol. **14**:223-228

7. **Douglas, T. and Young, M.** 1998. Host-guest encapsulation of materials by assembled virus protein cages. Nature. **393**(6681):152-155

8. **Shenton, W., Douglas, T., Young, M., Stubbs, G and Mann, S**. 1999. Inorganic-organic nanotube composites from template mineralization of tobacco mosaic virus. Adv. Mater. **11**(3):253-256

9. **Steinmetz, N.F., Calder, G., Lomonossoff, G.P. and Evans, D.J.** 2006. Plant viral capsids as nanobuilding blocks: construction of arrays on solid supports. Langmuir. **22**(24):10032-10037

10. **Steinmetz, N.F., Lin, T., Lomonossoff, G.P. and Johnson, J.E.** 2009. Structure-based engineering of an icosahedral virus for nanomedicine and nanotechnology. Curr. Topo. Microb. Immun. **327**:23-58

11. **Johnson, J.E. and Chiu, W.** 2007. DNA packaging and delivery machines in tailed bacteriophages. Curr. Op. Struct. Biol. **17**:237-243

12. **Toropova, K., Basnak, G., Twarock, R., Stockely, P.G. and Ranson, N.A.** 2008. The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly. J. Mol. Biol. **375**:824-836

13. **Paredes, A.M., Ferreira, D., Horton, M., Saad, A., Tsuruta, H., Johnston, R., Klimstra, W., Ryman, K., Hernandez, R., Chiu, W. and Brown, D.T**. 2004. Conformational changes in Sindbis virions resulting from exposure to low pH and interactions with cells suggest that cell penetration may occur at the cell surface in the absence of membrane fusion. Virol. **324**(2):373-386

14. **Briggs, J.A.G., Grünewald, K., Glass, B., Förster, F., Kräusslich, H-G. And Fuller, S.** 2006. The mechanism of HIV-1 core assembly: insights from thee-dimensional reconstructions of authentic virions. **14**(1):15-20

15. http://ictvonline.org/index.asp?bhcp=1

16. **Abrescia, N.G.A., Bamford, D.H., Grimes, J.M. and Stuart, D.I.** 2012 Structure unifies the viral universe. Annu. Rev. Biochem. **81**:795-822

17. **Baltimore, D. 1971**. Expression of Animal Virus Genomes. Bacteriol. Rev. **35**(3):235-241

18. **Lwoff, A., Horne, R., Tournier, P.** 1962. A system of viruses. Cold Spring Harbor Symposia on Quantitative Biology. **27**(0)51-55

19. **Fraenkel-Conrat, H. and Williams, R.C.** 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. P.N.A.S. **41**(10):690-698

20. **Bancroft, J.B., Hills, G.J. and Markham, R.** 1967. A study of the self-assembly process in a small spherical virus. Virol. **31**:354-379

21. **Fox, J.M., Johnson, J.E. and Young, M.J.** 1994. RNA/protein interactions in icosahedral virus assembly. Virol. **5**:51-60

22. **Rao, A.L.N. and Grantham, G.L.** 1996. Molecular studies on bromovirus capsid protein II. Functional analysis of the amino-terminal arginine-rich motif and its role in encapsidation, movement, and pathology. Virol. **226**:294-305

23. **Annamalai, P., Apte, S., Wilkens, S. and Rao, A.L.N.** 2005. Deletion of highly conserved arginine-rich RNA binding motif in Cowpea Chlorotic Mottle Virus capsid protein results in virion structural alterations and RNA packaging constraints. J. Virol. **79**(6):3277-3288

24. **Ahlquist, P.** 2002. RNA-Dependent RNA polymerases, viruses, and RNA silencing. Science. **296**(5571):1270-1273

25. **Qu, F. and Morris, T.J.** 1997. Encapsidation of Turnip Crinkle Virus is defined by a specific packaging signal and RNA size. J. Virol. **71**(2):1428-1435

26. **Schmitz, I. and Rao, A.L.N.** 1996. Molecular Studies on Bromovirus Capsid Protein I. Characterization of cell-to-cell movement-defective RNA3 variants Brome Mosaic Virus. Virol. **226**:281-293

27. **Kate, S. and Zlotnick, A.** 2009. The thermodynamics of virus capsid assembly. Methods in Enzymology. **455**:395-417

28. **Hagan, M.F**. 2008. Controlling viral capsid assembly with templating. Phys. Rev. E. **77**:051904

29. **Hagan, M.F.** 2009. A theory for viral capsid assembly around electrostatic cores. J. Chem. Phys. 130:114902

30. **Elrad, O.M. and Hagan, M.F.** 2010. Encapsulation of a polymer by an icosahedral virus. Phys. Biol. **7**(4):045003

31. **Restrepo-Hartwig, M. and Ahlquist, P.** 1999. Brome Mosaic Virus RNA replication proteins 1a and 2a colocalize and 1a independently localizes on the yeast endoplasmic reticulum. J. Virol. **73**(12)10303-10309

32. **Schwartz, M., Chen, J., Janda, M., Sullivan, M., den Boon, J. and Ahlquist, P.** 2002. A positive-strand RNA virus replication complex parallels form and function of retrovirus capsids. Mol. Cell. **9**(3):505-514

33. **Seo, J-K., Kwon, S-J. and Rao, A.L.N.** 2012. Molecular dissection of Flock House virus protein B2 reveals that electrostatic interactions between N-terminal domains of B2 monomers are critical for dimerization. Virol. **432**(2):296-305

34. **Ford, R.J., Barker, A.M., Bakker, S.E., Coutts, R.H., Ranson, N.A., Phillips, S.E.V., Person, A.R. and Stockley, P.G.** 2013. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. J. Mol. Biol. *In press*

35. **Borodavka, A., Tuma, R. and Stockely, P.G.** 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. P.N.A.S. **109**(39):15769-15774

36. **Cadena-Nava, R.D., Comas-Garcia, M., Garmann, R.F., Rao, A.L.N., Knobler, C.M. and 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**(6):3318-3326

37. **Comas-Garcia, M., Cadena-Nava, R.D., Rao, A.L.N., Knobler, C.M. and Gelbart, W.M.** 2012. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. J. Virol. **86**(22):12271-12282

Chapter II

Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio[*]

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 single-stranded RNA (ssRNA) molecules: RNA1, 3,171 nucleotides (nt); RNA2, 2,774 nt; and RNA3, 2,173 nt[2,18]. The first two genomic RNAs are packaged alone, and the third genomic RNA is copackaged with a subgenomic RNA4 of 824 nt, so that each CCMV virion contains about 3,000 nt. Just 2 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 4,000 to 6,400 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 masses 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, it has solution properties (e.g., radius of gyration) that are well understood, and it is readily available in a large range of molecular masses. Carrying out assemblies of virus-like particles (VLPs) for mixes of CCMV CP with PSS of various molecular masses, 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 findings 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 masses 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 mass 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 such as PSS.

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 Feng Guo (Department of Biological Chemistry, University of California, Los Angeles). 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) (3,234 nt) and BMV RNA 3 (B3) (2,117 nt), respectively[17], and plasmids p30B, pEYFP-Rep, and pTE12 encode the full length of TMV RNA (6,395 nt)[15], a Sindbis-derived replicon RNA containing enhanced yellow fluorescent protein (8,600 nt) and full-length Sindbis virus RNA (11,700 nt), respectively[30]. The plasmids were grown, purified, and analyzed by standard methods. All of the other chemicals used were DNase-, RNase-, and protease-free.

Combining B1 and B3 RNAs by directional subcloning.

Directional subcloning of pT7B1 requires the plasmid to have overhangs; these were created by linearization with BamHI and EcoRI. The insert that encodes B3 was engineered from pT7B3Sn by PCR using the forward primer d(T TGCGCGGATCCAACTAATTCTCGTTCG) (the BamHI site is underlined) and the reverse primer d(TGGCCGGAATTCAACACTGTACGG TAC) (the EcoRI site is underlined). These primers give 8 and 28 fewer nucleotides at the 5' and 3' ends, respectively, than genomic BMV RNA 3. The resulting PCR product was double digested with EcoRI and BamHI and subcloned into pT7B1 previously digested with BamHI/EcoRI 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 shorter than BMV RNA 1 were prepared by PCR of pT7B1 plasmid; all truncations were carried out from the 3' end. We used the universal 5' primer d(TAATACGACTCACTATAGGTAGA CCACGGAACGAGGTTC) (the T7 promoter is underlined) and, as reverse primers, d(GCAATCAACTTCAGCAAATCG), d(CACATCCTCT CCTCATGTC), d(GTCTTCAAACCATACACAGTG), d(CTTGCTCAA ATTCTTCAACG), and d(GGATACAACCAGTTACCGTTG) to obtain the corresponding DNA templates for transcribing RNAs with lengths of 141, 499, 999, 1,498 and 1,960 nt, respectively.

RNA transcription.

Plasmids pT7B3Sn and pT7B1 were linearized with BamHI, plasmids p30B and pTE12 were linearized with XhoI, and plasmid pEYFP-Rep was linearized with SacI. pRDCT7B1B3 was linearized with BglII and EcoRI to produce RNA transcripts of 4,452 and 5,319 nt, respectively.

17

All of the templates were purified by standard procedures[39]. All DNA templates and plasmids encoding for RNAs between 140 and 6,395 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-h dialysis against disassembly buffer (0.5 M $CaCl_2$, 50 mM Tris [pH 7.5], 1 mM EDTA, 1 mM dithiolthreitol [DTT], 0.5 mM phenylmethanesulfonylfluroide [PMSF]) at 4°C and then by ultracentrifugation at 100,000 rpm for 100 min ($4 \times 10^5 \times g$) 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 (1 M NaCl, 20 mM Tris [pH 7.2], 1 mM EDTA, 1 mM DTT, 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 and matrix-assisted laser desorption ionization–time of flight mass spectrometry were used 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 (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 mass ratios into a constant RNA concentration (30 ng/µl). After overnight assembly at 4°C, 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 h at 70 V (electrophoresis buffer) or 4 h at 50 V (virus buffer) in a horizontal gel apparatus (Fisher) at 4°C. The samples were then stained in a solution of 5 μg of ethidium bromide/ml and visualized with an Alphaimager system.

In vitro assembly of VLPs.

Dissociated CCMV CP subunits and de- sired RNA (30 ng/μl) transcripts were mixed in a ratio of 6:1 (wt/wt) and then dialyzed for 24 h 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 h. For assemblies with RNAs longer than BMV RNA1, 100 to 300 μl of the assembly reactions were subjected to sucrose gradient (10 to 40% [wt/wt]) centrifugation in virus buffer using an ultracentrifuge SW 50.1 rotor at 33,000 rpm ($1.3 \times 10^5 \times g$) for 2 h. 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 $3,000 \times g$ for 5 min; this last step was repeated four times. All of the procedures were carried out at 4°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-μl 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 the 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 structures of the cDNA templates for the RNAs whose self-assembly with CCMV CP is studied here are shown schematically in Fig. 1. The shorter molecules consist of BMV RNA1 (3,200 nt) and RNAs (140, 500, 1,000, 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 2,117-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 mammalian Sindbis virus -- an 8,600-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 virus itself. Although 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 minimized the possible effects of any specific CCMV packaging signals, allowing us to focus on the generic effects of
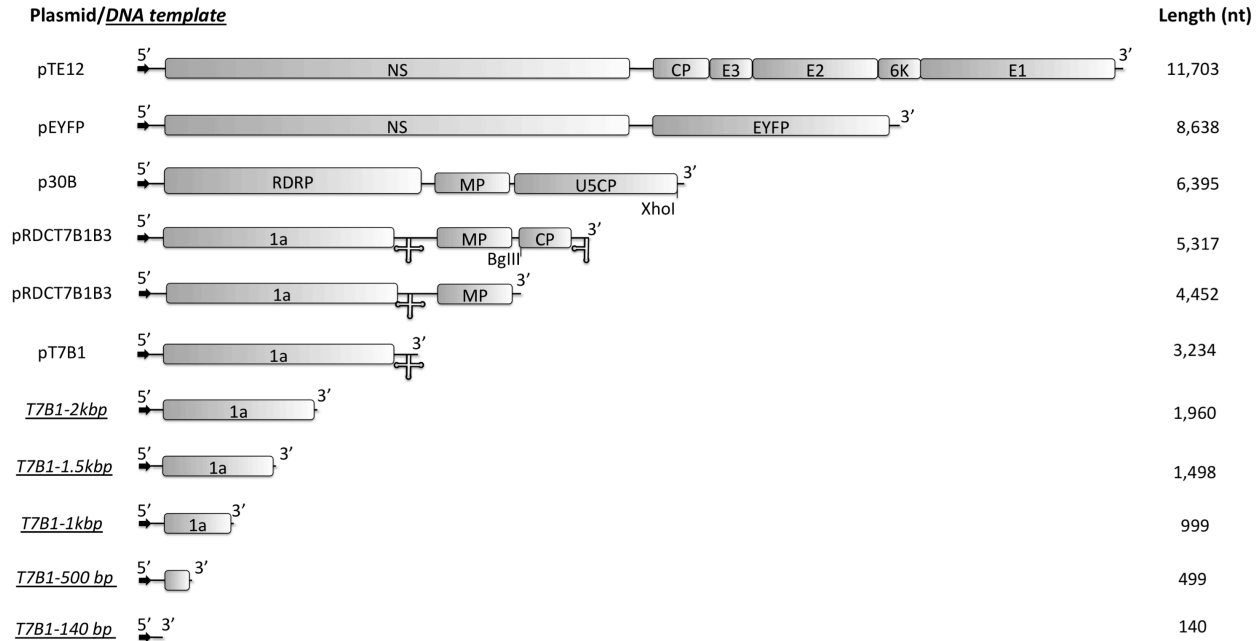
size and charge.

FIG 1 Schematic representation of the plasmids and DNA templates used for RNA syntheses. The arrows denote the transcription promoter, and the cloverleaf at the 3= end represents the highly conserved tRNA-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 nonstructural 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.

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 h and then dialyzed against virus suspension buffer before being analyzed by agarose gel retardation assays. The results in each case, as illustrated in Fig. 2 by the assemblies with BMV RNA1 (3,200 nt) and TMV RNA (6,400 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

21

wild-type (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 6,400-nt RNA the band migrates more slowly than that of CCMV; the reason for this is discussed below.)
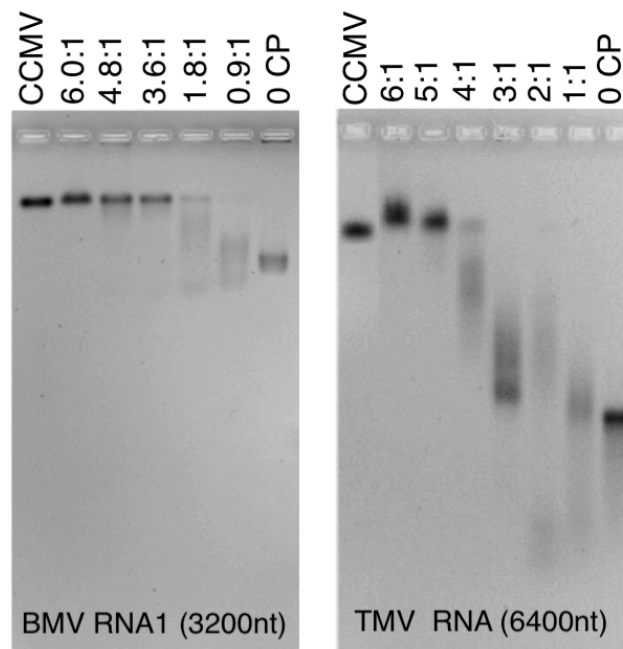

FIG 2 CP-RNA assembly titrations: gel retardation assays. Shown are 1% agarose gels stained with ethidium bromide. At the left is a titration of 3,217-nt BMV RNA1 with various amounts of CCMV CP ranging from 0 (right-most lane, RNA) to the *"magic ratio"*, 6:1 (lane second from left). CP/RNA ratios are wt/wt ratios of CP to RNA. This gel was run in electrophoresis buffer. On the right is a similar gel retardation assay carried out with 6,395-nt TMV RNA and run in virus buffer. In both gels, the leftmost 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.

The results of the assembly studies show RNAs of very different lengths can be completely packaged as long as the protein/RNA weight ratio is sufficiently high. It is remarkable, however, that we observe in every case that 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

22

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. 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 $M_{CP} = n_{CP}MM_{CP}$, where n_{CP} is the total number of moles of CP and MM_{CP} is the CP

molecular mass ($\sim 20,000$ Da). Similarly, for the total mass of RNA, we write $M_{RNA} =$

$n_{RNA}L_{RNA}MM_{NT}$, where n_{RNA} is the total number of moles of RNA, L_{RNA} is the RNA length in

nucleotides, and MM_{nt} is the average molecular mass of a nucleotide (~ 330 Da). It follows that

the special ratio of the number of CP subunits (n_{CP}) to the number of nucleotides ($n_{nt} =$

$n_{RNA}L_{RNA}$) is given by $n_{CP}/n_{nt} = (MM_{nt}/MM_{CP}) (M_{CP}/M_{RNA}) = (330/20,000) (6/1) = 1/10$. 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 inter- acting with RNA or otherwise involved in the assembly process.

Nevertheless, from the fact that at the *magic ratio* one CP subunit is present for each 10 nt of RNA we infer (since each CP N terminus has 10 cationic residues) that 10 N-terminal cationic residues are present for each 10 nt. Although 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 (with k_B the Boltzmann constant). This is also consistent with charge matching of oppositely charged colloidal particles, e.g., 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 counter-ion 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 on the order of 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 wt 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 (4,051 nt), bushy stunt virus (4,776 nt), turnip yellow mosaic virus (6,318 nt), and tobacco mosaic virus (6,400 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 used.

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 3,200 nt long, and around Xenopus elongation factor RNA, 1,900 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 ~ 4.

VLP products as a function of RNA length.

Figure 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 upper left-hand picture is WT CCMV. The numbers of capsids sharing a single RNA molecule (in the case of the large nucleotide lengths), the sizes of VLPs formed in each instance (Fig. 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 electron micrographs for the TMV VLPs makes clear the reason for this ambiguity. The majority of the capsids are paired -- see, for example, the 6,400-nt frame in Fig. 3 and the upper right- hand image ("image B") in Fig. 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 were not mentioned. 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].
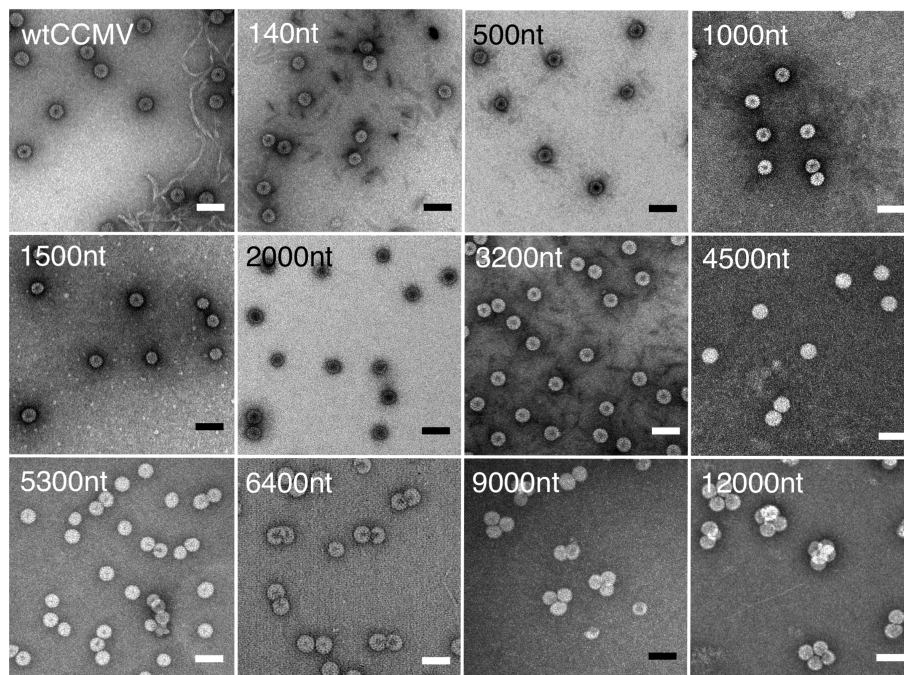

FIG 3 VLP products as a function of RNA length. Negative-stain transmission electron microscopy images of the assembly products of various lengths of ssRNA mixed with CCMV CP at the magic ratio (6:1 [wt/wt]) in RNA assembly buffer. The length of the RNA in each assembly is labeled in the upper left corner of each image. The upper leftmost image shows wt CCMV capsids. Grids were stained with uranyl acetate. Scale bars, 50 nm.

Formation of multiplets.

As shown in Fig. 4B, the multiplets first appear as a small fraction of doublets for 3,200 nt RNA. The fraction of doublets observed in the micrographs increases markedly at 4,500 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 trip- lets rise and go through maxima (cf. Fig. 4B, bottom). Assemblies with the 11,700-nt Sindbis virus RNA lead primarily to quadruplets and higher-order multiplets. (Note that each of the average diameters plotted in Fig. 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 ca. 18% over a period of 4 h, 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 the present study appeared to be aggregates of virus and proteinaceous 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 defective interfering ssDNA, the number of multiplets increases with DNA length [21].

The multiplet VLPs that we observed in CCMV VLPs are unlikely to have structures akin to those of geminiviruses, 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 (Fig. 5A and B) in the case of quadruplets forming from Sindbis virus 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 nanometer-sized holes in the capsids. Indeed, in some electron microscopy images (Fig. 5C) a bit of RNA can be seen poking out between some of them.
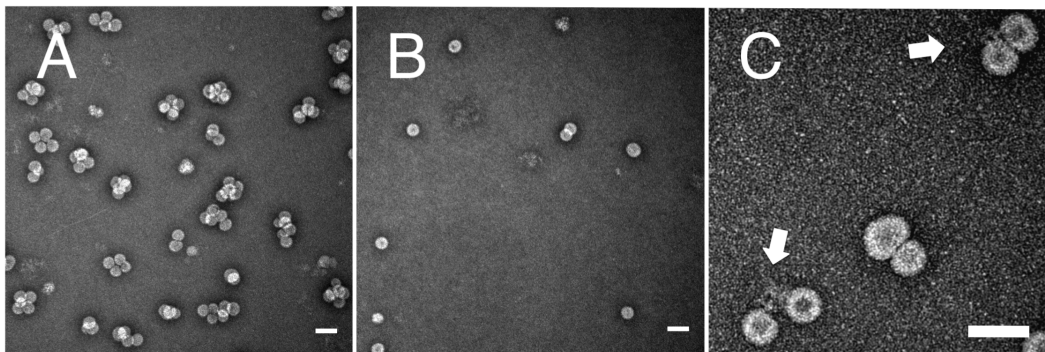

FIG 5 Exposure of RNA in Multiplets. (A) Typical transmission electron microscopy image of 12,000-nt RNA assembled with CCMV CP. (B) Same assembly mixture after exposure to RNase A in virus buffer for 1 h. Note the reduction in the number of multiplets, as well as an overall reduction in the total number of capsids present. (C) Shared RNA between capsids in a multiplet. The arrows point to RNA seen linking two capsids. Scale bars, 50 nm.

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, e.g., tetrahedral for the quadruplets (Fig. 4B, micrograph D), 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 virus RNA, show that the fraction of particles of each multiplicity is independent of dilution, indicating further that the multiplets are not simply physical aggregates (Fig. 6A).

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

28

compensated for 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 desorption of protein with reorganization of the remaining protein as a closed, icosahedral, capsid, including the polymer.
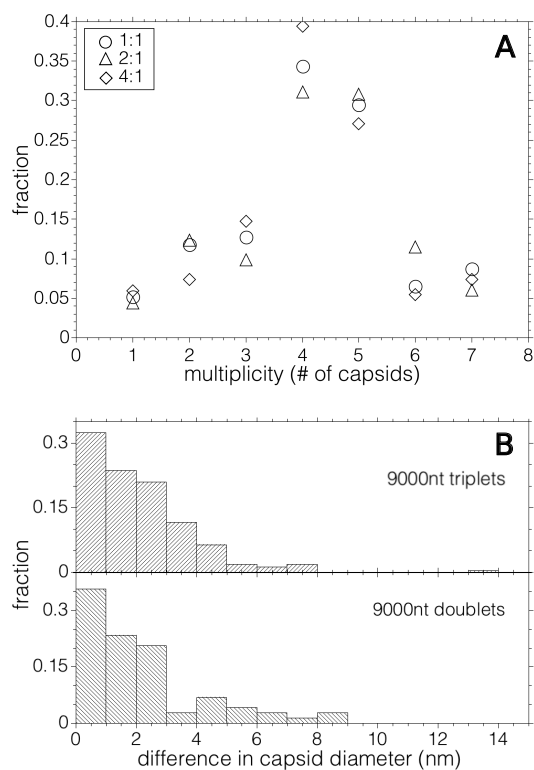

FIG 6 Properties of multiplets. (A) Absence of effect of 2- and 4-fold dilutions on the frequency of multiplets. The different symbols show the effect of sample dilution on fractions of multiplets of each type. The sample represented by the circles has been diluted by factors of 2 (△) and factor of 4 (◇). There is no systematic effect on the populations of multiplets with dilution. (B) Size differences between capsids within doublets and triplets formed from the assembly of 9,000-nt RNA and CCMV capsid protein.

This scenario is consistent with measurements of the differences in diameter between

partners in a multiplet (Fig. 6B). Although 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 wild-type length or longer.

Distributions of diameters determined from electron micrographs are shown by the histograms in Fig. 7 for the 11 RNAs studied. Also included in the figure for comparison is a measurement for WT CCMV, for which the average number of packaged nucleotides is 3,000. 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-electron microscopy to be 28 nm in diameter in solution[20], was found by transmission electron microscopy 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 structure 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, the 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 ~ 24 nm in diameter fall into a size range that has been identified with a 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, a value consistent with a T = 1 triangulation number, as has been observed in assemblies of BMV[44] and of VLPs formed from CCMV protein lacking the N terminus[40].
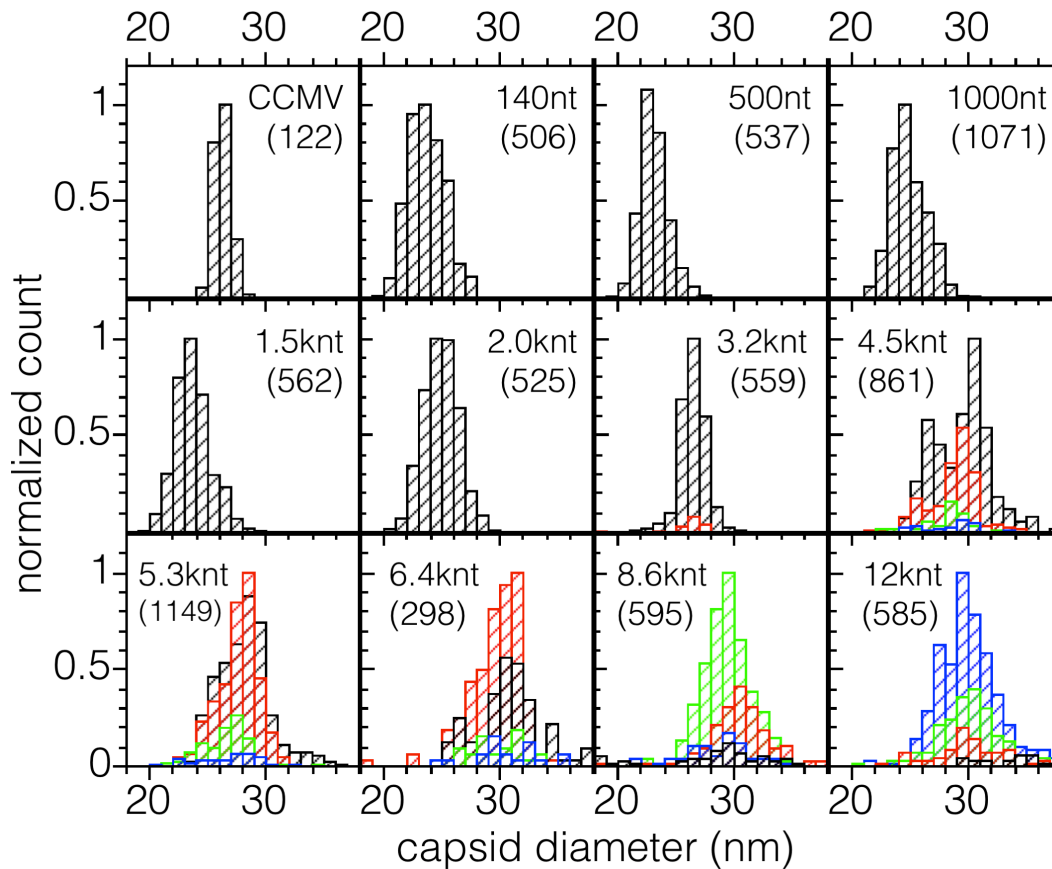

FIG 7 Distribution of capsid diameters appearing in singlets and multiplets for each length of RNA packaged. Contributions from singlets, doublets, trip- lets, and quadruplets are shown in black, red, green, and blue, respectively. The distribution at the upper left is for wt CCMV. In each distribution the counts have been normalized to that of the most abundant size. The total numbers of particles measured are given in parentheses.

Although 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 seven 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 observed a population of particles with diameters distinctly larger than those that have been identified as T = 3. However, 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 observed, however, are homogeneous in their distribution, well-formed and closed, as evidenced by the protection they afford against RNase to the pack- aged 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 26 nm $(4/3)^{1/2}$ = 30 nm, in accordance 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 ~ 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. 3 and the accompanying size distributions shown in Fig. 7, RNAs longer than the 3,000 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). However, 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 (3,200 nt) the scenario is different. VLPs containing 2,000, 1,500, 1,000, and 140 nt RNAs are a mixture of T = 2 and T = 3 capsids, whereas the 500-nt RNA is predominantly packaged into T = 2. The electron microscopic data for almost all of these samples (Fig. 7) show a broad VLP size distribution with a mean diameter that is intermediate between what we would expect for T = 2 and T = 3 capsids, indicating that both populations are significantly represented. Although 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 T = 3 populations. 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 (Fig. 7). From the gel retardation assays, we know that at the magic ratio all of the RNA is packaged, suggesting that for RNAs between 140 and 1,500 nt the capsids contain more than one molecule. 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 nm over that at 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 to 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 3,200-, 2,000-, 1,000-, and 500-nt RNAs are indistinguishable from that of CCMV, from which we conclude that there is only one 3,200- and 2,000-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 1,000-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 1,500-nt RNA (1.5 versus 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 RNAs (140 nt). 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), fluorescence (rhodamine)-labeled, PSS molecules by CCMV CP, we found T = 1 VLPs to be the predominant product[12]. However, 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 a rigid-ring, rhodamine label, several nanometers in length, 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 (< 2,000 nt), the RNA is packaged into $T = 2$ and $T = 3$ capsids with one to four molecules of RNA in the capsid. For sufficiently large RNAs (> 4,500 nt), single molecules of RNA are packaged by two or more $T = 3$ or $T = 4$ capsids. Finally, for intermediate lengths, single RNAs are pack- aged, 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 CP-CP interactions.

In vitro self-assembly of single capsids of CCMV having RNAs of ~ 3,000 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 (3,171 nt) and RNA2 (2,774 nt) are each packaged independently into $T = 3$ capsids, whereas RNA3 (2,173 nt) and RNA 4 (824 nt) are copackaged 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, i.e., the assembly is mediated by capsid protein that is translated from replication-derived mRNA[5]. 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.

Although 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]. These authors found that 13 residues other than those on the protein tail interact with the RNA, 7 of them basic, and that the RNA "forms three distinct bulges directly under the shell near the quasi-3-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 two hexamers and a pentamer and, at the stage in assembly that we 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. These findings 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: (i) nucleation of a capsid *"embryo"* involving a small number of proteins, followed by growth and (ii) 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 3,000 nt -- the length of WT 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 3,000 nt, a single molecule is packaged in a 26-nm capsid. Moreover, for RNA lengths > 4,500 nt, two or more capsids are involved in the packaging of a single molecule.

ADDENDUM IN PROOF

The existence of multiplets has also recently been reported in the case of self-assembly of CCMV CP around a conjugated polyelectrolyte (M. Brasch and J. J. M. Cornelisse, Chem. Commun. 48: 1446 –1448, 2012).

REFERENCES

1. **Adolph KW, Butler PJG.** 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 RF, Janda M, Ahlquist P.** 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, Rao ALN.** 2005. Dispensability of 3' tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. Virology **332**:650 – 658.

4. **Annamalai P, Rao ALN.** 2006. Packaging of brome mosaic virus sub- genomic RNA is functionally coupled to replication-dependent transcription and translation of coat protein. J. Virol. **80**:10096 –10108.

5. **Annamalai P, Rofail F, DeMason DA, Rao ALN.** 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 –1490.

6. **Bamunusinghe D, Seo J-K, Rao ALN.** 2011. Subcellular localization and rearrangement of endoplasmic reticulum by brome mosaic virus capsid protein. J. Virol. **85**:2953–2963.

7. **BancroftJB.** 1970. The self-assembly of spherical plant viruses. Adv. Virus Res. **16**:99 –134.

8. **Bancroft JB, Hiebert E.** 1967. Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. Virol- ogy **32**:354 –356.

9. **Bancroft JB, Hiebert E, Bracker CE.** 1969. The effects of various poly- anions on shell formation of some spherical viruses. Virology **39**:924 –930.

10. **Basnak G, et al.** 2010. Viral genomic single-stranded RNA directs the pathway towards a T

= 3 capsid. J. Mol. Biol. **395**:924 –936.

11. **Briddon RW, Markham PG.** 1995. Family Geminiviridae, p 158 –165. In Murphy FA, et al. (ed), Virus taxonomy: archives in virology. Springer-Verlag, New York, NY.

12. **Cadena-Nava RD, et al.** 2011. Exploiting fluorescent polymers to probe the self-assembly of virus-like particles. J. Phys. Chem. B **115**:2386 –2391.

13. **Caspar DLD, Klug A.** 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol. **27**:1–24.

14. **Choi YG, Rao ALN.** 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 YG, Rao ALN.** 2000. Packaging of tobacco mosaic virus subgenomic RNAs by brome mosaic virus coat protein exhibits RNA controlled polymorphism. Virology **275**:249 –257.

16. **De Haseth PL, Lohman TM, Record MT, Jr.** 1977. Nonspecific interactions of lac repressor with DNA: an association reaction driven by counter-ion release. Biochemistry **16**:4783– 4790.

17. **Dreher TW, Rao ALN, Hall TC.** 1989. Replication in vivo of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. **206**:425– 438.

18. **Dzianott A, Bujarski J.** 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 OM, Hagan MF.** 2010. Encapsulation of a polymer by an icosahedral virus. Phys. Biol. **7**:045003.

20. **Fox JM, et al.** 1998. Comparison of the native CCMV virion with in vitro assembled virions by cryoelectron microscopy and image reconstruction. Virology **244**:212–218.

39

21. **Frischmuth T, Ringel M, Kocher C.** 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, Bancroft JB, Bracker CF.** 1968. The assembly in vitro of some small spherical viruses, and other nucleoproteins. Virology **34**:492–508.

23. **Hu Y, Zandi R, Anavitarte A, Knobler CM, Gelbart WM.** 2008. Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. Biophys. J. **94**:1428 – 1436.

24. **Johnson JM, Willits DA, Young MJ, 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. 25.

25. **Johnson MW, Wagner GW, Bancroft JB.** 1973. A titrimetric and electrophoretic study of cowpea chlorotic mottle virus and its protein. J. Gen. Virol. **19**:263–273. 26.

26. **Jung B, Rao ALN, Anvari B.** 2011. Optical nanoconstructs composed of genome-depleted Brome mosaic virus doped with a near infrared chromophore for potential biomedical applications. ACS Nano. **5**:1243–1252. 27.

27. **Kobayashi A, Ehara Y.** 1995. In vitro encapsidation of cucumber mosaic virus RNA species. Ann. Phytopathol. Soc. Jpn. **61**:99 –102.

28. **Kroll MA, et al.** 1999. RNA-controlled polymorphism in the in vivo assembly of a 180-subunit and 120-subunit virions from a single capsid protein. Proc. Natl. Acad. Sci. U. S. A. **96**:13650 –13655.

29. **Lavelle L,et al.** 2009. Phase diagram of self-assembled viral capsid protein polymorphs. J. Phys. Chem. B **113**:3813–3820.

30. **Lustig S, et al.** 1988. Molecular basis of Sindbis virus neurovirulence in mice. J. Virol.

62:2329 –2336.

31. **Mascotti DP, Lohman TM.** 1990. Thermodynamic extent of counter-ion release upon binding oligo-lysines to single-stranded nucleic acids. Proc.Natl. Acad. Sci. U. S. A. **87**:3142–3146.

32. **Nugent CI, Johnson KL, Sarnow P, Kirkegaard K.** 1997. Functional coupling between replication and packaging of poliovirus replicon RNA. J. Virol. **73**:427– 435.

33. **Obenauer-Kutner LJ, et al.** 2002. The use of field emission scanning electron microscopy to assess recombinant adenovirus stability. Hum. Gene Ther. **13**:1687–1696.

34. **Pfeifer P, Herzog M, Hirth L.** 1976. Stabilization of brome mosaic virus. Philos. Trans. R. Soc. Lond. B Biol. Sci. **276**:99 –107.

35. **Porterfield JZ, et al.** 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, van der Schoot P, Gelbart WM, Knobler CM.** 2010. Shell structures of virus capsid proteins. J. Phys. Chem. B **114**:5522–5533.

37. **Qu F, Morris TJ.** 1997. Encapsidation of turnip crinkle virus is defined by a specific packaging signal size RNA. J. Virol. **71**:1428 –1435.

38. **Rao ALN.** 2006. Genome packaging by spherical plant RNA viruses. Annu. Rev. Phytopathol. **44**:61– 87.

39. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

40. **Sikkema FD, et al.** 2007. Monodisperse polymer-virus hybrid nanoparticles. Org. Biomol. Chem. **5**:54 –57.

41. **Sorger PK, Stockley PG, Harrison SC.** 1986. Structure and assembly of turnip crinkle

virus. II. Mechanism of reassembly in vitro. J. Mol. Biol. **191**:639 – 658.

42. **Speir JA, Munshi S, Wang G, Baker TS, Johnson JE.** 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, et al.** 2007. Core-controlled polymorphism in virus-like particles. Proc. Natl. Acad. Sci. U. S. A. **104**:1354 –1359.

44. **Tang J, et al.** 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, Scheek RM, van der Linden CC, Hemminga MA.**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 PA, Krishna NK, Schneemann A.** 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 BJM, Bancroft JB.** 1969. The infectivity of tobacco mosaic virus RNA in coat proteins from spherical viruses. Virology **37**:501–506.

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

49. **Zandi R, van der Schoot P.** 2009. Size regulation of ssRNA viruses. Biophys. J. **96**:9 –20.

50. **Zhang W, et al.** 2001. Structure of the maize streak virus geminate particle. Virology **279**:471– 477.

51. **Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ.** 2000. Mechanism of capsid

assembly for an icosahedral plant virus. Virology **277**:450 – 456.

Chapter III

In vitro quantification of the relative packaging efficiencies of

single-stranded RNA molecules by viral capsid protein[*]

INTRODUCTION

The remarkable capacity of the capsid protein (CP) of cowpea chlorotic mottle virus (CCMV) to self-assemble in vitro into nanometer-size capsids around a broad range of anionic materials -- its own single-stranded RNA (ssRNA) genome[2,6], heterologous ssRNAs[3,4,9,24], organic polymers such as polystyrene sulfonate[11,16,23], metal oxide particles1[9,28], functionalized gold nanoparticles[5], and negatively charged nanoemulsion droplets[13] -- as well as into a wide range of structures in the absence of any polyanions[1,8,26] has spurred interest in the mechanism of assembly. But a fundamental question has rarely or only inadequately been addressed: *how efficient are the assembly processes?*

To answer this question, it is first necessary to formulate a precise definition of efficiency of assembly, along with a procedure for determining it. In some cases, efficiency has been quantified as the fraction of filled capsids in the mixture of filled and empty capsids observed in electron micrographs of the assembly mixture[14], but this depends on the formation of empty capsids under the conditions of filled-capsid assembly. In other cases, the relative efficiency of packaging an RNA molecule has been defined as the ratio of the number of in vivo capsids

containing it to the number formed with the genomic RNA[5], when both are present in the same CP-expressing cell; however, this requires additional knowledge about the relative levels of replication of both RNAs.

Several assembly experiments, both *in vivo* and *in vitro*, have been reported[29,31,32,34] in which there have been qualitative measurements of the relative amounts of different RNAs packaged by a CP, often with the goal of determining the location of a pack- aging signal. In particular, Qu and Morris[32] studied the length dependence of packaging efficiency by carrying out head-to-head competitions in which protoplasts were coinfected with the wild-type (WT) RNA of turnip crinkle virus along with a series of mutant and chimeric RNAs. Also, it has been shown by in vitro head-to-head competitions that hepatitis B virus capsid protein has a slightly higher preference for its pregenomic RNA than for a heterologous one[29].

While head-to-head comparisons are informative, they do not by themselves allow an unambiguous determination of efficiency. It is necessary that an additional constraint be imposed: the CP has to be a limiting reagent. The way in which limiting-reagent conditions can be precisely achieved has become clear from recent assembly studies on CCMV carried out in our laboratory[12]. More explicitly, we have demonstrated that for RNAs ranging in length from 140 to 12,000 nucleotides (nt), a 6:1 mass ratio of CCMV CP to RNA is the minimal amount of CP necessary to package all the RNA (we have called this the *"magic ratio"*). Thus, if equal masses of two RNAs are mixed with CP at an overall ratio of 3:1, the RNAs must compete for CP because there is sufficient CP to package only one of them completely. The results of such competition experiments can be expressed in terms of a relative packaging efficiency E of an

RNA x, which is defined as $E = w_x/w_R$, the ratio of the mass of RNA x (competitor) packaged to the mass of the reference RNA; this is equivalent to the ratio of the number of packaged nucleotides of x to the number of packaged nucleotides of the reference molecule.

In this study, we used competition experiments of this kind to quantitatively compare the packaging efficiencies of a wide variety of RNAs by CCMV CP with respect to that of the brome mosaic virus (BMV) RNA1 (B1), which is 3.2 kb in length. We found that RNAs shorter than 2.0 kb cannot compete at all with B1, i.e., have relative packaging efficiencies of zero; sufficiently short RNAs, e.g., 0.5-kb molecules, are copackaged with B1; RNAs slightly shorter than 3.2 kb (e.g., 2.0 and 2.8 kb) and slightly longer (3.5 and 4.0 kb) are able to compete, but with relative efficiencies of only 30 to 80%, and involve the formation of "pseudo T = 2" and T = 4 capsids, respectively; and homologous RNA (i.e., CCMV RNA1) does not necessarily have the highest packaging efficiency of equal-length RNAs. By carrying out the two-step *in vitro* assembly process with different orders of mixing of the competing RNAs, we also found that at neutral pH the RNAs are bound reversibly by CP, whereas the formation of mature, RNase-resistant nucleocapsids occurs irreversibly under acidic pH conditions.

MATERIALS AND METHODS

Materials.

Restriction enzymes were obtained from New England BioLabs. T7 RNA polymerase was a gift from Feng Guo (Department of Biological Chemistry, University of California, Los Angeles). Enzymes were used as recommended by the manufacturer. The plasmids were grown, purified,

and analyzed by standard methods[36]. Plasmids pT7B1, pT7B2, and pCC1 contain full-length cDNA clones corresponding to BMV RNA 1 (B1) (3,234 nt), BMV RNA 2 (B2) (2,864 nt), and CCMV RNA 1 (C1) (3,170 nt)[20]; plasmid pRDCT7B1B3 is a cDNA clone corresponding to the ligated product of B1 and B3[12]. All other chemicals used were DNase, RNase, and protease free.

PCR amplification of DNA templates to transcribe RNAs.

DNA templates shorter than B1 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(CACATCCTCTCCT CATGTC), d(GTCTTCAAACCATACACAGTG), d(CTTGCTCAAATTCTTCAACG), d(GGATACAACCAGTTACCGTTG), and d(GTGAACA ACGTCACGGCG) to obtain the corresponding DNA templates for transcribing RNAs with lengths of 499, 999, 1,498, 1,960, and 2,507 nt, respectively. RNAs longer than B1 were transcribed from DNA templates obtained by PCR of pRDCT7B1B3 plasmid. For the 3,586- and 4,065-bp DNA templates the reverse primers d(CGTTGTAAAACGACGGCCAGTG) and d(CGCGCGGTCATCTTACCAGTT), respectively, were used. The forward primer was the same as for the pT7B1 plasmid-derived templates. All the templates were purified by standard procedures[36].

RNA transcription and fluorescent labeling.

Plasmids pT7B1 and pT7B2 were linearized with BamHI, pCC1 was linearized with XbaI, and pRDCT7B1B3 was linearized with EcoRI. All DNA templates and plasmids were transcribed with T7 RNA polymerase. For the quantitative competition experiments, RNAs were labeled

with either ChromaTide Alexa Fluor 488-5-UTP (AF488) or ChromaTide Alexa Fluor 546-13-UTP (AF546) (henceforth referred to as modified rUTPs) from Invitrogen. These modified rUTPs are randomly incorporated during transcription. To achieve an average label density between 2 and 5 modified rUTPs per RNA, the molar ratios of rATP to rGTP to rCTP to rUTP to modified rUTP during transcription were modified to 600:600:600:5.32:1. The label density for each labeled RNA was calculated by linear regression of a calibration curve of pure modified rUTPs. The fluorescence measurements were carried out with a QuantaMaster spectrofluorimeter (Photon Technology International) in a 40-μl quartz cuvette (Sterna Cells). Fluorescent RNAs were always kept in the dark in amber tubes, and room light was minimized when possible during handling.

CCMV CP purification.

CCMV was purified from infected California cowpea plant (*Vigna ungiculata* cv. Black Eye)[9], and CP was isolated largely as described previously[6,12]. SDS-PAGE and matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF) were employed to ascertain that the purified protein was intact.

In vitro **assemblies for the qualitative competition experiments.**

In carrying out our *in vitro* packaging competition assays, we distinguish between qualitative and quantitative measurements of relative efficiencies. In the qualitative case, we extract RNA from RNase-treated assembly mixes and simply run the RNA in native and denaturing gels to determine whether one or both are present. In the quantitative case, the RNAs are labeled with different fluorophores so that their relative numbers in RNase-treated assembly mixes can be

assayed quantitatively by measuring the fluorescence spectra of the purified nucleocapsids. Five controls were always used in each qualitative competition experiment: reference RNA plus CCMV CP, competitor RNA plus CCMV CP, WT CCMV, reference RNA, and competitor RNA. The RNA concentration in every control was 30 ng/µl, and the CP/RNA ratio (wt/wt) for the assemblies was 6:1; the WT CCMV concentration was 120 ng/µl (corresponding to about 30 ng/µl RNA). In every competition experiment, equal masses of the competitor and reference RNAs -- each at a concentration of 30 ng/µl -- were mixed in buffer B (BB; 1 M NaCl, 20 mM Tris-HCl [pH 7.2], 1 mM EDTA, 1 mM dithiolthreitol [DTT], and 1 mM phenylmethanesulfonylfluroide [PMSF]), with dissociated CCMV CP subunits at a CP/total RNA ratio (wt/wt) of 3:1 (here total RNA is competitor RNA plus reference RNA). Samples and controls were dialyzed for 20 h at 4°C 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]), followed by a 4-h dialysis against virus suspension buffer (VSB; 50 mM sodium acetate, 8 mM magnesium acetate [pH 4.5]). All controls and samples were processed identically. Aliquots (8-µl) of each sample and control were set aside for electrophoretic mobility assays, while the rest was treated with RNase A (Invitrogen) (4 ng/µl) for 45 min. RNase A was inactivated by incubating the samples with RNase inhibitor (Roche) for 15 min. All samples and controls were purified and concentrated by washing with VSB using a 100-kDa Amicon centrifuge filter (0.5 ml; Millipore) at 3,000 × g for 5 min; this step was repeated 4 times. Two 6-µl aliquots of each sample and control were set aside for UV-visible (UV-Vis) spectroscopy and transmission electron microscopy (TEM), and the rest was used to analyze the RNA content by gel electrophoresis. All the procedures were carried out at 4°C.

RNA extraction from VLPs.

Prior to RNA extraction, samples were incubated for 15 min with 10 U of RNase inhibitor (Roche) at 4°C. Virus-like particles (VLPs) (~ 750 ng of RNA) were disrupted by adding 10X RNA extraction buffer (0.5 M glycine, 0.5 M NaCl, 0.1 M EDTA [pH 9.0]) and 5 μl of proteinase K (New England BioLabs); samples were incubated for 30 min at 37°C prior to addition of 5 μl of 20% SDS. The samples were then vortexed, microcentrifuged, incubated for 1 h at 37°C, diluted to 150 μl with TE buffer (10 mM Tris [pH 8.0], 1 mM EDTA), heated at 98°C for 4 min, and chilled on ice for 5 min. Finally, the RNA was extracted by phenol-chloroform and ethanol precipitation[33]. The recovered RNA was analyzed by native (in Tris-acetate-EDTA [TAE]) and denaturing (8 M urea in TAE) gel electrophoresis in 1% agarose.

In vitro **assemblies for quantitative competition experiments.**

Each quantitative competition consisted of two sets of studies, each performed in triplicate or quadruplicate. In the first, samples of the reference and competitor RNA were labeled with AF488 and AF546, respectively. In the second set, the labeling was reversed. Nine controls were included in each quantitative competition experiment: AF488-reference RNA plus CCMV CP, AF546-reference RNA plus CCMV CP, AF488-competitor RNA plus CCMV CP, AF546-competitor RNA plus CCMV CP, WT CCMV, and all four kinds of RNAs. The RNA concentration in every control was 30 ng/μl, the CP/RNA ratio (wt/wt) was 6:1, and the concentration of WT CCMV was 120 ng/μl. As in the qualitative experiments, equal masses of the competitor and reference RNAs at a CP/total RNA ratio (wt/wt) of 3:1 were employed. The assembly was also carried out by a 20-h dialysis against RAB followed by a 4-h dialysis against VSB; we then treated the sample with RNase and purified it exactly as in the qualitative

50

experiments. All controls were treated exactly as in the competition samples.

Quantitative measurements by UV-Vis and fluorescence spectroscopy.

After purification of the VLPs containing fluorescent rUTPs, all samples were diluted to 52 μl. A 2-μl aliquot was used to measure the UV-Vis absorbance (Fisher Nanodrop spectrophotometer). It has been observed that the ratio of the absorbances at 260 and 280 nm for concentrated virus solutions may be strongly affected by light scattering[30], requiring a correction. Given the low concentration of our samples (~ ng/μl), the uncorrected UV spectrum differed by less than 1% from the corrected one. The samples containing AF488 and AF546 were excited at 490 and 555 nm, respectively, and the fluorescence spectra were recorded between 500 and 550 nm (AF488) and 560 and 610 nm (AF546). The fluorescence intensity of each sample was normalized with respect to its absorbance at 260 nm and the label density of the excited RNA species.

Electrophoretic mobility analysis (EMA).

After the two-step dialysis, 8-μl aliquots of each sample and control (before and after RNase A treatment) were mixed with 3 μl of 100% glycerol (RNase, DNase, and pro- tease free) and loaded into a 1% agarose gel (EMD OmniPur) in virus electrophoresis buffer (0.1 M sodium acetate, 1 mM EDTA [pH 6]). The samples were electrophoresed for 1.25 h at 60 V in a horizontal gel apparatus (Fisher) at 4°C. For the qualitative experiments, the gels were stained in a solution of 5 μg/ml ethidium bromide and visualized with an Alphaimager system (ProteinSimple). For the quantitative experiments, electrophoresis was carried out in complete absence of light. The gels were visualized in an FX Pro plus Fluorimager/PhosphorImager (Bio-Rad) by exciting AF546, followed by staining with ethidium bromide and visualization as

already described.

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 μl-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 micro- scope 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 the geometric mean of two orthogonal measurements determined with ImageJ (U.S. National Institutes of Health) software from recorded images.

RESULTS

We chose to minimize any possible specific RNA-CP interactions by using BMV RNA1 (B1) instead of CCMV RNA1 (C1) as the reference. We designed a series of small competitor RNAs by deriving them from the 5' end of B1; the lengths for the truncated RNAs were 0.5, 1.0, 1.5, 2.0, and 2.5 kb. For competitor RNAs longer than B1, we used 3.6- and 4.0-kb RNAs that are truncations -- maintaining the 5' end -- of the product of the ligation of BMV RNAs 1 and 3 (5.3 kb)[12]. Finally, we also studied two viral RNAs: BMV RNA 2 (2.8 kb [B2]) and CCMV RNA1 (3.1 kb [C1]). A schematic representation of the plasmids and DNA templates used for the transcription of the RNAs employed can be found in Fig. 1.
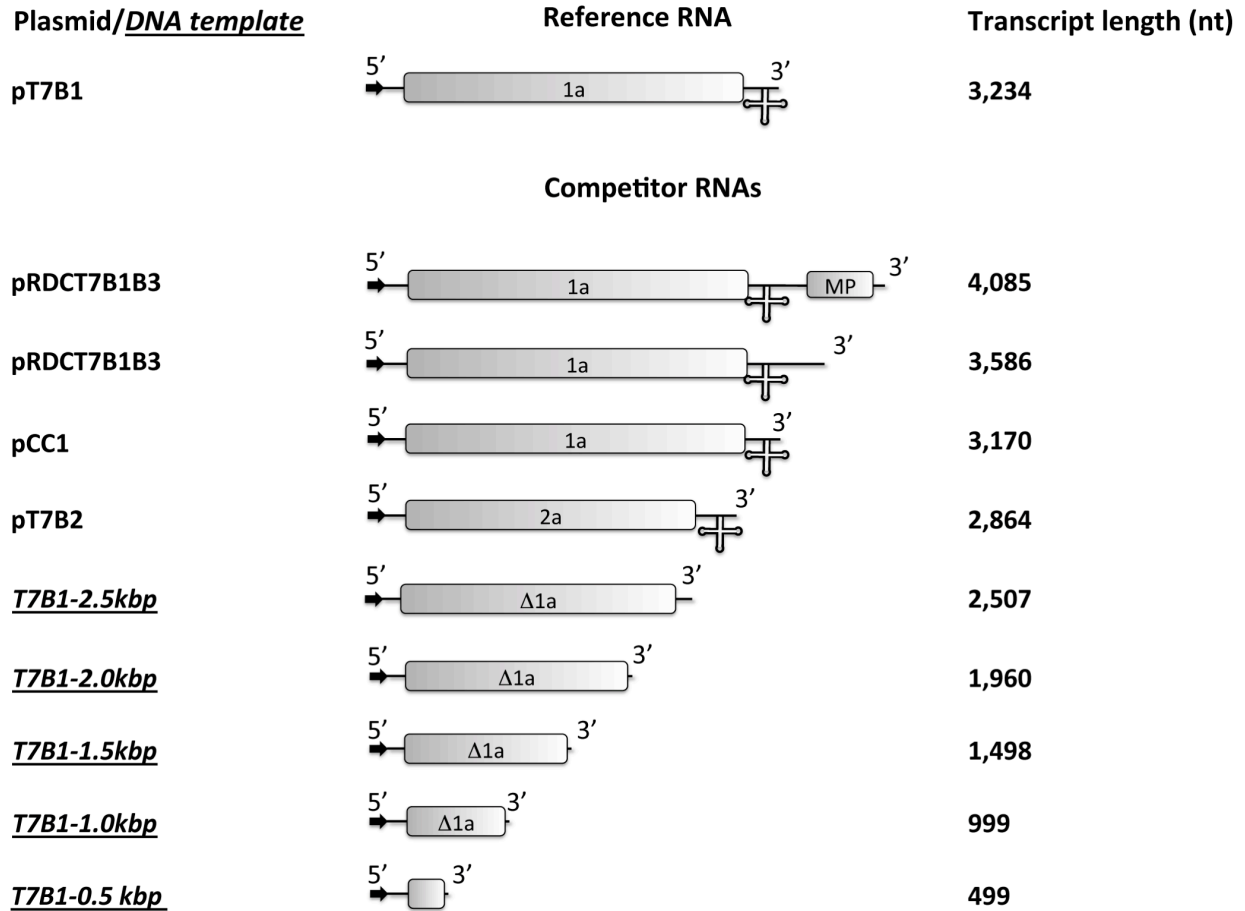
Plasmid/_DNA template_	Reference RNA	Transcript length (nt)
pT7B1	5' [1a]✝ 3'	3,234

Competitor RNAs

pRDCT7B1B3	5' [1a]✝[MP] 3'	4,085
pRDCT7B1B3	5' [1a]✝ 3'	3,586
pCC1	5' [1a]✝ 3'	3,170
pT7B2	5' [2a]✝ 3'	2,864
T7B1-2.5kbp	5' [Δ1a] 3'	2,507
T7B1-2.0kbp	5' [Δ1a] 3'	1,960
T7B1-1.5kbp	5' [Δ1a] 3'	1,498
T7B1-1.0kbp	5' [Δ1a] 3'	999
T7B1-0.5 kbp	5' [ ] 3'	499

FIG 1 Schematic representation of the plasmids and DNA templates used for RNA syntheses. The arrows indicate the T7 transcription promoter, and the cross at the 3' end represents the highly conserved tRNA-like structure in the corresponding RNA transcripts. The boxes represent the open reading frames (ORFs) of the RNAs, 1a and 2a are viral replicases, and MP is the movement protein. "Δ1a" indicates that the viral replicase ORF was truncated. The representations of the plasmid templates are scaled to their relative lengths.

Briefly, a competition experiment consists of mixing equal masses of the competitor and reference RNA with CCMV CP at a CP/total RNA mass ratio of 3:1 in buffer B (BB). The samples are then dialyzed against RNA assembly buffer (RAB) for 20 h, followed by 4-h dialysis against virus suspension buffer (VSB). Samples are kept at 4°C throughout the experiment.

Both qualitative and quantitative assays were performed. The goal of the qualitative assay was simply to identify which RNA species was packaged. This was achieved by analyzing the reaction mixture in an electrophoretic mobility assay (EMA) before and after RNase A

treatment; an EMA under denaturing conditions was also employed to examine RNA extracted from purified VLPs after treatment with RNase A. The resolution of these qualitative experiments was sufficiently high for cases in which only one of the RNAs was packaged. In principle, when both species were packaged, the relative amount of each RNA could also be determined by the latter procedure, but it is not reliable because of possible degradation of RNA during the extraction process. Quantitative assays were performed by labeling the RNAs with fluorescent rUTPs (AF488 and AF546), which were incorporated during in vitro transcription.

Fluorescent labeling has no effect on packaging efficiency.

To determine if the incorporation of the modified rUTPs had any effect on the packaging process, we assembled CCMV CP (at the *magic ratio*) with different ratios of unlabeled to AF488-labeled B1 (0:1, 1:1, and 1:0). As shown in Fig. 2a, the EMAs of the three samples are identical. The *magic ratio* was also not altered; i.e., the minimum CP/RNA ratio required for complete packaging was 6:1 in each case (data not shown). The normalized fluorescence intensity of the VLPs that were assembled in a 1:1 unlabeled-labeled RNA mixture was precisely half that of the VLPs assembled in a 0:1 unlabeled-labeled RNA mixture (Fig. 2b). Finally, negative-stain transmission electron microscopy (TEM) showed that there are no discernible differences between VLPs containing different amounts of labeled and unlabeled RNA (data not shown).

RNAs of 1.0 kb and 1.5 kb cannot compete with B1, but 0.5-kb RNA is copackaged.

The EMA for competition experiments between B1 and the 1.0- and 1.5-kb RNAs showed similar behavior. Figure 3 shows that in a competition between 1.0-kb and B1 there are three

main species: VLPs, high-molecular-mass RNA/CP complexes, and free RNA. Denaturing gels of the RNA extracted from purified VLPs after RNase A treatment revealed no RNA other than B1 (Fig. 4). The size distributions determined by TEM for VLPs containing a 1.0-kb RNA and that from the competition between this RNA and B1 are shown in Fig. 5e. The distributions for VLPs containing only 1.0-kb RNAs show mixed populations of capsids with diameters between 23 and 26 nm. We have previously argued that it is most likely that capsids with a diameter around 23 nm correspond to "pseudo T = 2" (henceforth referred to simply as "T = 2") and that those with diameters close to 26 nm correspond to T = 3 capsids[12, 23]. Here, and throughout this article, we assign T numbers to VLPs on the basis of comparing the VLP radii to that of WT CCMV T = 3 virions (26 nm). More explicitly, from the fact that radii of Caspar-Klug structures scale as $T^{1/2}$, we associate radii of 22 and 30 nm with T = 2 and 4 VLPs, respectively[23].
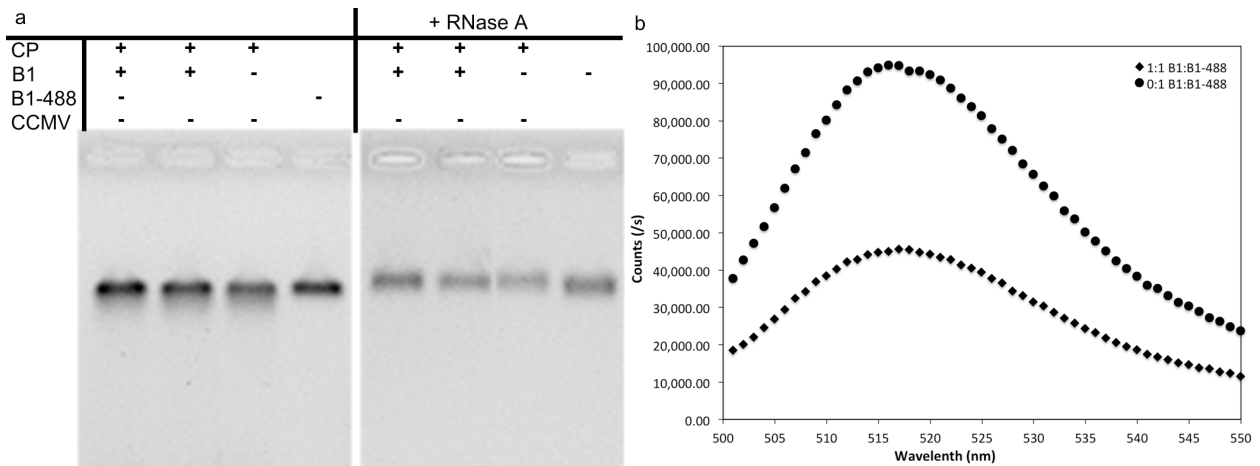

FIG 2 The incorporation of AF488 into B1 RNA (B1-488) has no effect on the assembly process. (a) The electrophoretic mobility analysis (EMA) shows that the electrophoretic mobilities of VLPs containing different ratios of unlabeled to labeled B1 are identical. (b) The fluorescence intensity of VLPs in which half the B1 is labeled is half that of VLPs containing only labeled B1.

On the other hand, the size distribution of capsids containing only B1 is centered around 26 nm (dotted curves in Fig. 5). The capsids obtained from the competition experiments

involving B1 and 1.0-kb RNA have diameter distributions that are indistinguishable from those for capsids containing only B1 (Fig. 5e). Similarly, the competition with 1.5-kb RNAs leads to diameter distributions around 26 nm, while the distributions for each of the shorter RNAs alone are bimodal (Fig. 5a to d). Along with the data from the RNA extraction (Fig. 4B and C), these observations indicate that essentially all of the competitor RNAs were excluded from packaging; hence, their relative packaging efficiency is zero. Finally, it is worth noticing that the amount of high-molecular-mass complexes is small compared to the situation in which the competitor RNA is packaged, which is possible only if most of the competitor RNA is excluded from forming stable RNA/CP complexes.
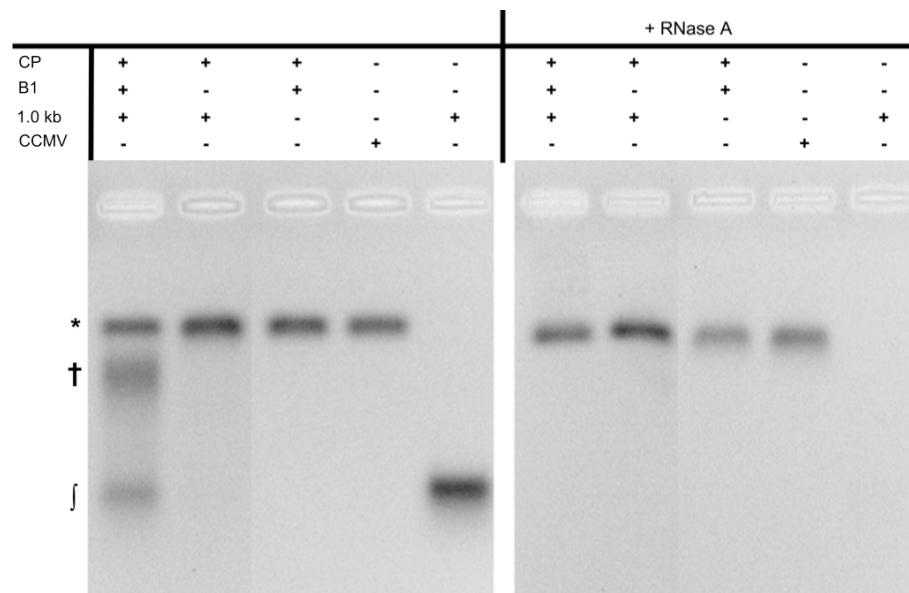

FIG 3 EMA of a competition experiment between B1 and a 1.0-kb RNA. The left panel shows the three main species in a competition experiment: virus-like- particles (VLPs) (*), high-molecular-mass RNA/CP complexes (†), and free RNA (∫). The right panel shows the samples after treatment with RNase A; only the bands corresponding to the VLPs remain. The gel was stained with ethidium bromide.

Although the EMA for the competition experiments with a 0.5-kb RNA was almost identical to those for the 1.0- and 1.5-kb competitions, the denaturing EMA of the RNAs extracted from the purified capsids (after RNase A treatment) showed two bands, corresponding

to B1 and the 0.5-kb RNA (Fig. 4A). Figure 6a compares the normalized UV-Vis spectra of WT CCMV and VLPs containing either only B1, only 0.5-kb RNA, and both (from this competition experiment). The spectra of VLPs containing a single RNA species overlap that of WT CCMV, but the spectrum of the VLPs from the competition experiment has a significantly higher ratio of absorbances at 260 and 280 nm (2.13 versus 1.74), demonstrating that the concentration of nucleotides per protein is higher than for the other samples. The capsid diameter distribution for VLPs containing only 0.5-kb RNA strongly suggests that they are predominately T = 2 (average diameter, 23 nm), but the capsids from the competition between B1 and 0.5-kb RNA have a diameter distribution centered at 27 nm (Fig. 6b). Based on the UV-Vis spectra and the capsid diameter distribution, we can conclude that the 0.5-kb RNA is copackaged with B1. The copackaging cannot be attributed to dimerization of the RNAs as shown by native gels of mixtures of the two RNAs, which showed two distinct bands under assembly conditions in the absence of CP (data not shown).
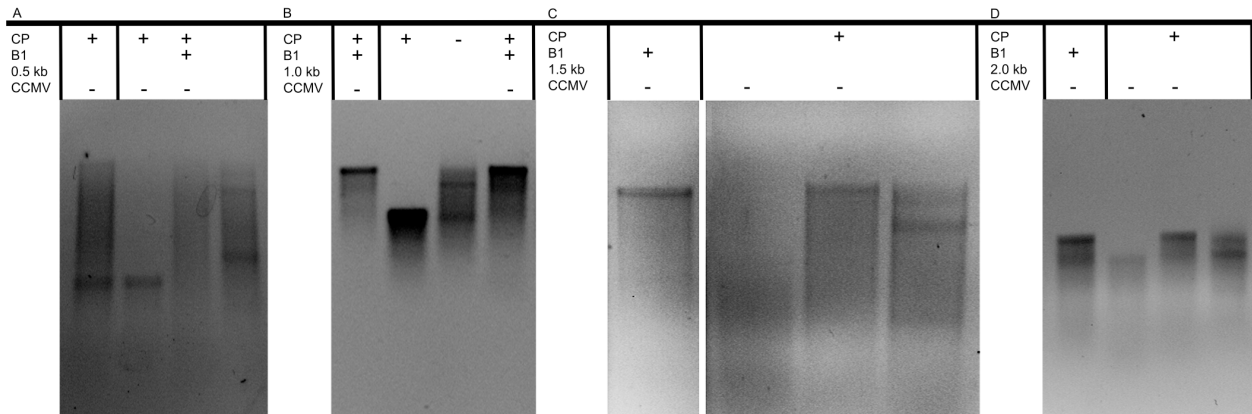

FIG 4 EMAs of RNAs extracted from competitions between B1 and 0.5-kb RNA (A), 1.0-kb RNA (B), 1.5-kb RNA (C), and 2.0-kb RNA (D). Panels A and D show that both competitor RNAs were packaged in the presence of B1, while panels B and C show that neither the 1.0- nor 1.5-kb RNA was packaged. These EMAs were carried out under denaturing conditions: 8 M urea in TAE.
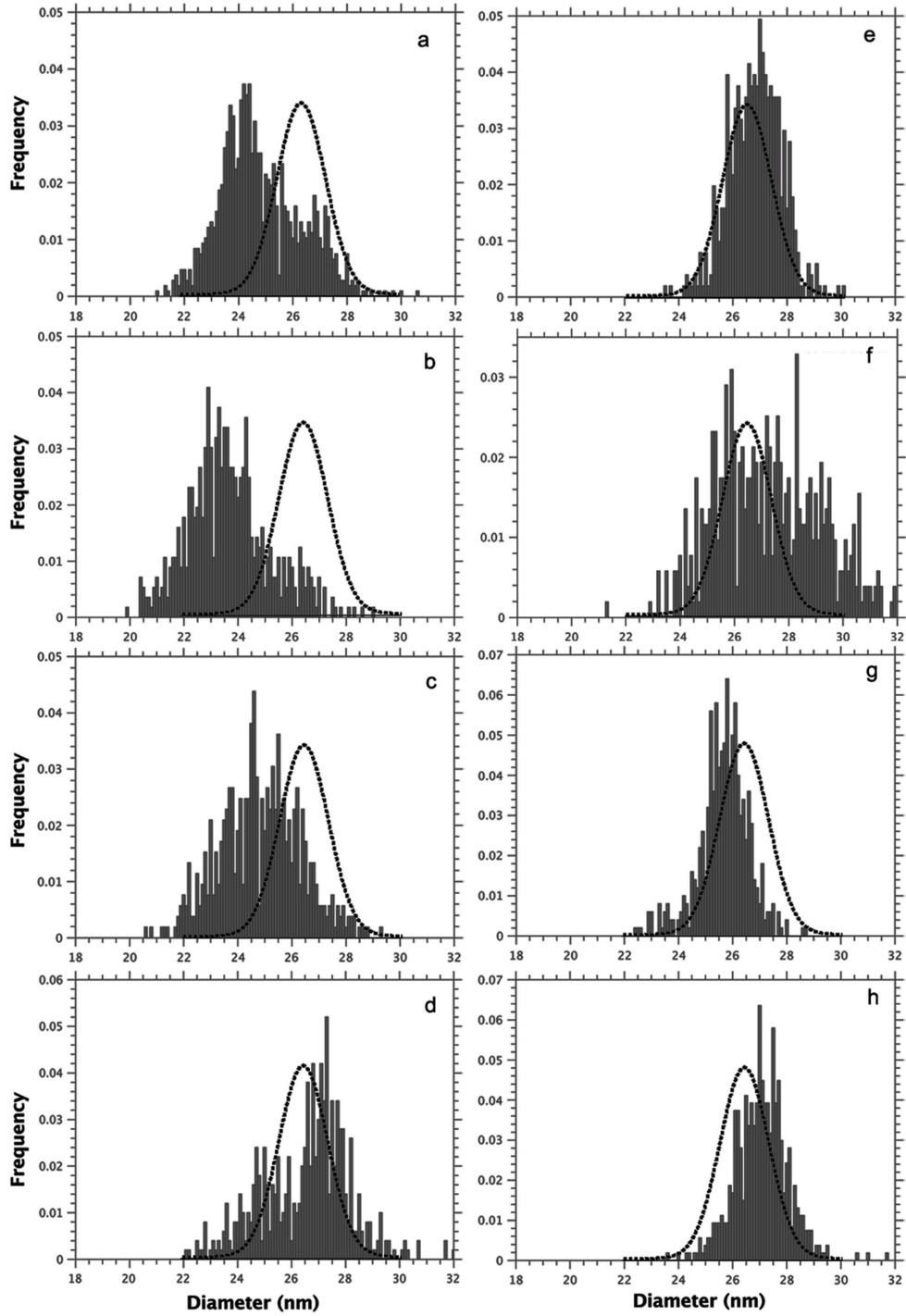
FIG5 (a, b, c, and d) Diameter distributions of VLPs containing a single RNA species: RNAs of 1.0 (a), 1.5 (b), 2.0 (c), and 2.5 (d) kb. (e, f, g, and h) Same as panels a to d, but for VLPs from competition experiments between B1 and RNAs of 1.0 (e), 1.5 (f), 2.0 (g), and 2.5 (h) kb. The dotted line in each panel corresponds to the diameter distribution of VLPs containing B1 RNA. Except for the competition between B1 and a 2.0-kb RNA, which contains a higher population of VLPs with a diameter smaller than 26 nm, the diameter distributions for all competition experiments are almost identical to those for VLPs containing only B1 RNA.

RNAs of 2 kb and longer compete with B1.

We found that the shortest RNA that competes with B1 but is not copackaged is 2.0 kb in length (Fig. 4D). The EMA of the assembly mix in this case showed that the RNA was present in the form of either VLPs or high-molecular-mass RNA/CP complexes; essentially no free RNA was detected (data not shown). From an examination of the fluorescence from AF546-labeled RNA in the EMA it is obvious that both RNA species were assembled into VLPs. As we have noted, this technique does not allow us to reliably determine the relative populations of each RNA species. To obtain more quantitative estimates, the VLPs were purified after RNase A treatment and their fluorescence intensity was then measured and normalized with respect to the total RNA content in each and the average label density in each RNA species. To calculate the relative packaging efficiency, the normalized intensity of VLPs containing AF546-labeled 2.0-kb RNA was divided by the normalized intensity of VLPs labeled with AF546-labeled B1; the same was done for the AF488-labeled RNAs. Because the reaction mixture had equal masses of the 2.0-kb RNA and B1, there were 1.6 times more moles of the smaller RNA than of the full-length RNA, and accordingly, the measured normalized intensities were corrected for this difference. The results for both fluorophores agree: the relative pack- aging efficiency of the 2.0-kb RNA was estimated to be 0.74 ± 0.16. The size distribution of VLPs with a 2.0-kb RNA indicated a mixed population of $T = 2$ and $T = 3$ capsids. This is the first of our competition experiments in which a capsid population was not exclusively associated with $T = 3$ capsids (Fig. 5g). We conclude that the $T = 2$ capsids contain 2.0-kb RNA, while only about 2/3 of the $T = 3$ capsids contain B1 (the other 1/3 contain 2.0-kb RNA).

Figure 5d shows that when 2.5-kb RNA alone is packaged, roughly equal numbers of $T =$

2 and T = 3 capsids are formed; in contrast, the diameter distribution for the capsids arising from the competition between B1 and 2.5-kb RNA showed a single population centered at 27 nm, corresponding to T = 3 (Fig. 5h). And yet the corresponding EMA gel (not shown) looks qualitatively similar to that for the competition with 2.0-kb RNA plus B1 (Fig. 4D), so we know that the 2.5-kb RNA is able to compete with B1, even though B1 is preferred. More quantitatively, we found from bulk fluorescence measurements that the relative packaging efficiency of the 2.5-kb RNA is 0.33 ± 0.06 (see Fig. 9). It is not clear why we did not observe any T = 2 capsids in this competition experiment but rather only when the 2.5-kb is at the magic ratio in the absence of B1.
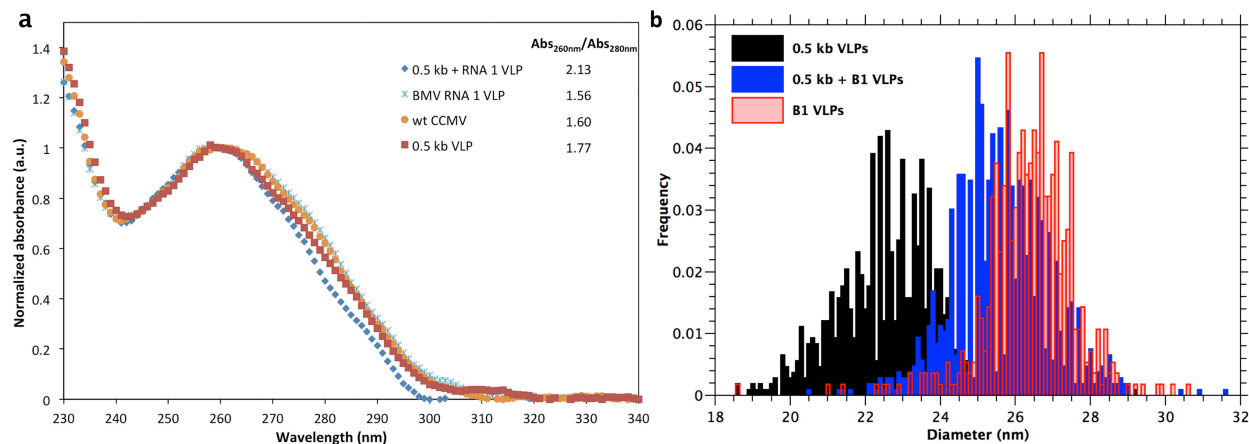

FIG 6 (a) The normalized UV-Vis spectra of VLPs from a competition experiment between 0.5-kb RNA and B1 show that the ratio of absorbances at 260 and 280 nm ($Abs_{260\,nm}/Abs_{280\,nm}$) is considerably higher than it is for WT CCMV or VLPs containing either 0.5-kb RNA or B1. This ratio reflects the relative concentration of nucleotides in a capsid. (b) The size distribution of VLPs from this competition experiment is almost identical to that for VLPs containing only B1, while the distribution for those containing 0.5-kb RNA is considerably smaller. The higher RNA content in T = 3 VLPs indicates that the 0.5-kb RNA was copackaged with B1.

More biologically relevant competition experiments test B1 versus B2 (2.8 kb) and B1 versus C1 (3.1 kb). The EMAs for these samples are almost indistinguishable from those of the 2.0- and 2.5-kb competitions. Figure 7 shows the EMA for the competition between B1 and C1;

almost all of the unpackaged RNA was found as RNA/CP complexes, and most of the

fluorescence signal of the VLPs came from B1. B2 and C1 alone each form T = 3 capsids with

an average diameter of 27 nm; the VLPs resulting from the B1 versus B2 and B1 versus C1

competitions have identical diameter distributions (data not shown). The bulk fluorescence

measurements showed that the relative packaging efficiencies of B2 and C1 are 0.56 ± 0.11 and

0.35 ± 0.09, respectively, corresponding to about 1/3 and 1/4 of the VLPs containing B2 and C1,

respectively. Finally, Fig. 7 shows that the electrophoretic mobilities of C1 and B1 are the same

and that the presence of either fluorophore does not affect the result of the competitions.
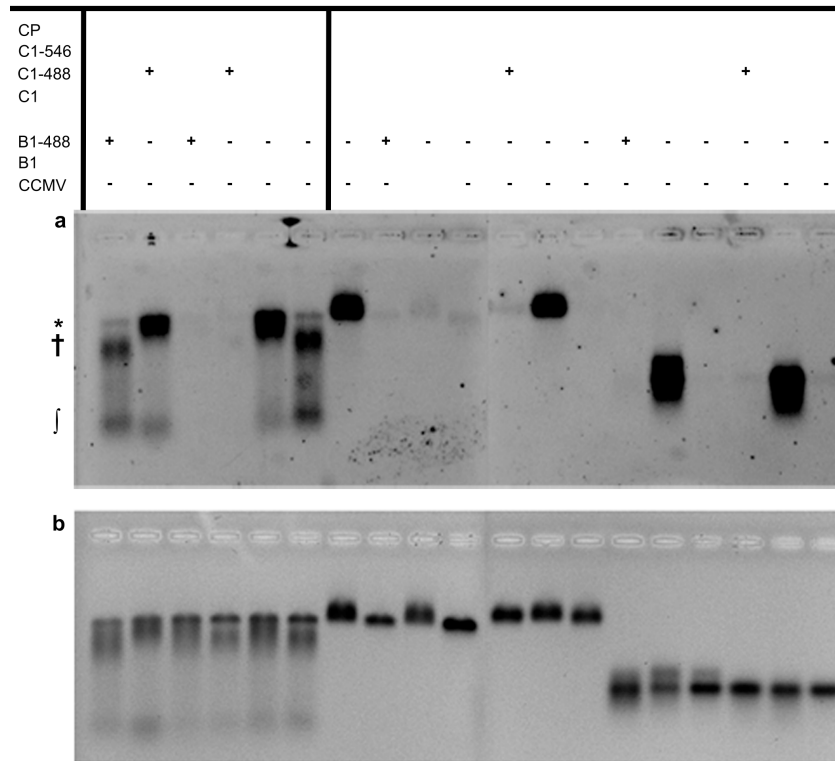

FIG 7 Competition between B1 and C1. (a) This EMA, visualized by exciting Alexa Fluor-546 UTPs, shows that B1 is packaged more efficiently than C1 into VLPs (*). Most of C1 is in the form of high-molecular-mass RNA/CP complexes (†), and there is a small fraction of free RNA (∫). B1-488 (C1-488) and B1-546 (C1-546) correspond to BMV RNA 1 (CCMV RNA 1) labeled with AF488 and AF546. (b) The same gel stained with ethidium bromide showing total RNA content. This EMA also confirms that neither the presence nor the nature of the fluorophore perturbs the assembly process.

Longer (3.6- and 4.0-kb) RNAs can also compete.

The last two competition experiments involved RNA molecules that are 12.5 and 25% longer than B1. As mentioned earlier, these two molecules (3.6- and 4.0-kb long) are truncations of the ligation product of B1 and B3 and, as in the case of the short RNAs, the 5' end of B1 was conserved. The 3.6-kb molecule was the RNA with the third- highest relative packaging efficiency (0.65 ± 0.18) among all of those examined. TEM micrographs show that VLPs containing only a 3.6-kb RNA have the same average diameter (26 nm) as VLPs containing B1 and those arising from the competition experiment between B1 and this RNA. The 4.0-kb RNA has a relative packaging efficiency of 0.81 ± 0.24; 95% of the VLPs are spherical (with a small fraction of them having a diameter consistent with $T = 4$) and the remaining 5% are asymmetric. These asymmetric capsids are ellipsoidal, with an axial ratio of ≥ 1.5, and constitute as much as 25% of all particles when the 4.0-kb RNA is packaged alone (Fig. 8). While ellipsoidal CCMV VLPs of this size have not been reported before, Bancroft et al.[8] observed smaller (28-nm-long) ellipsoidal particles with a similar axial ratio when CCMV was digested with pancreatic RNase at pH 7.3. A summary of the relative packaging efficiencies for RNAs between 1.0 and 4.0 kb in length is shown in Fig. 9. The curve drawn there, which is a guide to the eye, pertains only to cases in which $T = 3$ capsids form (triangles) and shows that there is a maximum packaging efficiency, which, not surprisingly, occurs at the length corresponding to B1.

Reversibility of RNA/CP binding at neutral pH; irreversibil- ity of nucleocapsid formation at low pH.

To examine if the relative efficiencies of the RNAs that we have observed reflect differences in the kinetics of assembly or differences in stability, we have carried out a series of experiments in

which we examined how the order of addition of the RNAs affects the assembly products. The experiments probe the degree to which reversibility exists under assembly conditions.
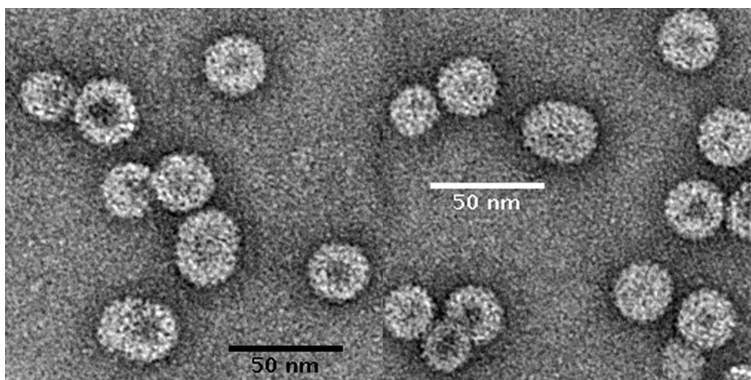

FIG 8 Negative-stain electron micrograph of VLPs containing a 4.0-kb-long RNA. There are two different sizes of spherical capsids and elongated particles. In competition experiments between B1 and a 4.0-kb RNA, about 5% of the particles have axial ratios of ≥ 1.5; in the absence of B1, 25% of the particles have ratios of ≥ 1.5. The particle at the right end of the white scale bar in the right panel has an axial ratio of 1.63.
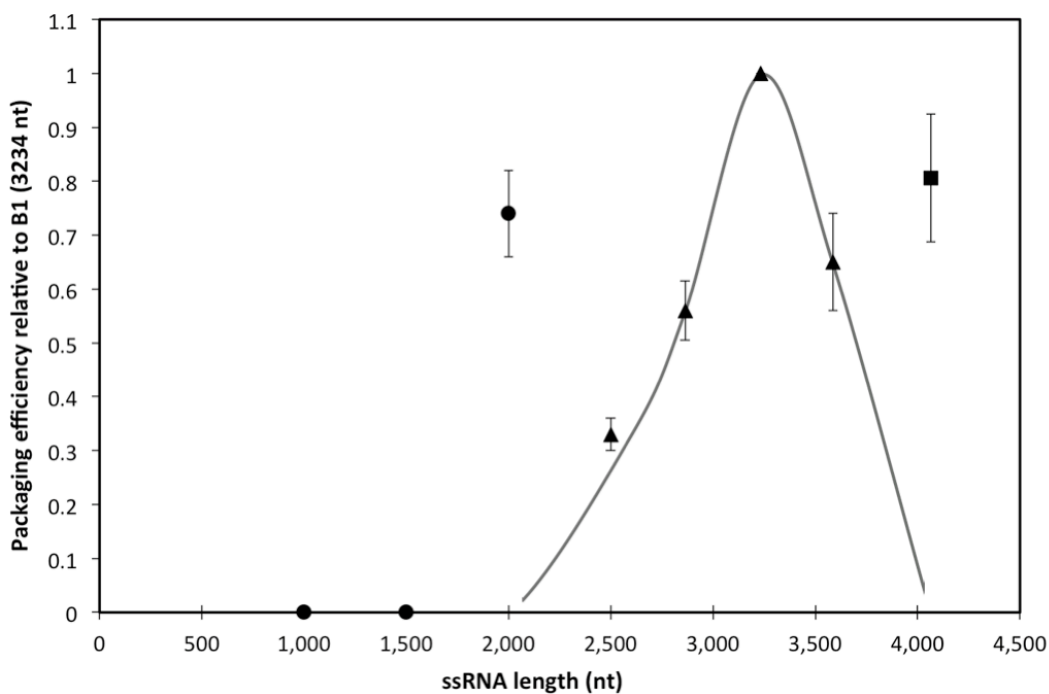

FIG 9 Relative packaging efficiency of ssRNAs as a function of length. The vertical bars represent the standard deviations. All measurements are relative to B1 (3,224 nt). The triangles represent competitions where the competitor RNAs form $T = 3$ capsids only. The circles correspond to cases where $T = 2$ and $T = 3$ capsids form, and the square corresponds to the case involving $T = 3$ and $T = 4$. The curve, which is drawn to guide the eye, illustrates that when the RNAs can only form $T = 3$ capsids, the relative packaging efficiency has a maximum at 3.2 kb.

We performed a series of competition experiments between B1 and B2. In the first set of reactions both B1 and B2 were mixed with CP at the same time (in BB) and then dialyzed at neutral pH just as in the previous experiments, with a CP/total RNA ratio of 3:1. In the second set, either B1 or B2 was added to the CP solution and dialyzed against RAB for 6 h before the other RNA was added, followed by dialysis against RAB for 12 h more before dialysis against VSB for 4 h. The result of this experiment is shown in the EMAs in Fig. 10. In the upper gel the signal comes only from RNA labeled with AF546, while the lower EMA shows the total RNA (ethidium bromide staining). In lanes 1 and 2 (left to right), B1 and B2 were added at the same time. Lanes 3 and 4 show the result when B2 was added 6 h after B1, and lanes 5 and 6 show the result for the reverse order. It is evident from the gels that qualitatively the result is independent of the order of addition or the time between the additions; i.e., B1 is preferentially packaged over B2. We carried out similar studies for a case in which the competition for the CP was markedly different, i.e., involving as a competitor the 1.0-kb RNA that is not packaged at all when going head-to-head against B1 at a CP/total RNA ratio of 3:1. As with the case with B1 plus B2, the results of these RNA experiments with B1 plus 1.0-kb RNA were found to be independent of the order of the mixing of the two RNA molecules in the neutral-pH assembly reaction; only B1 was packaged. The effect of pH on the reversibility of binding was examined in another two sets of experiments, and the results are shown in the four right-most lanes of the EMA in Fig. 10. In these experiments, one of the RNA species and CP were dialyzed against RAB for 20 h and against VSB for 2 h, after which the other RNA was added and the mixture dialyzed against VSB for an additional 2 h. In contradistinction to the experiments described above in which changes of order are carried out entirely at neutral pH, it is evident that if the second RNA -- regardless of whether it is the less or more efficiently packaged one -- is added after the pH has been lowered

64

to pH 4.5 (VSB), it is not incorporated into capsids; i.e., the reversibility is lost once the pH is lowered.

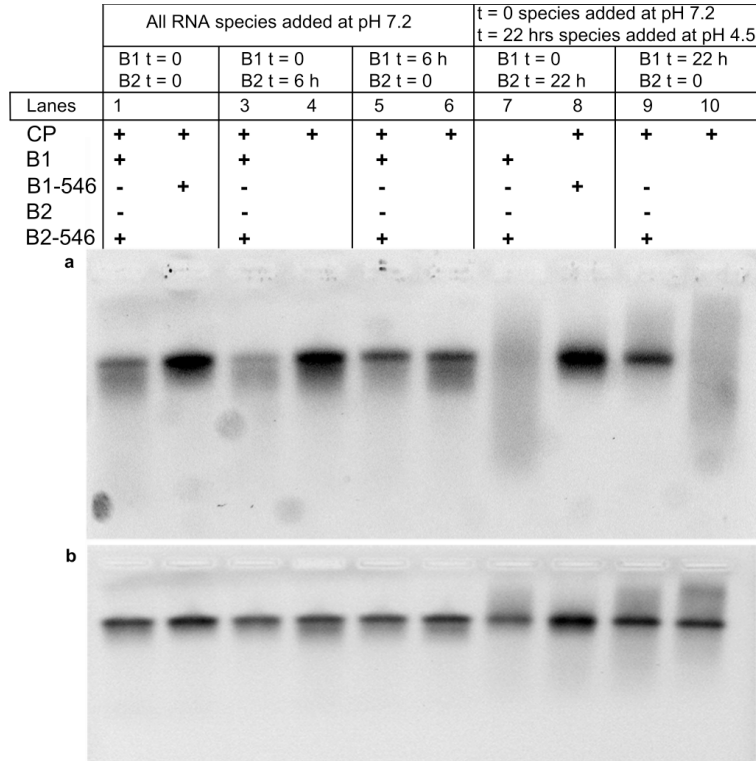|	All RNA species added at pH 7.2						t = 0 species added at pH 7.2 t = 22 hrs species added at pH 4.5			
	B1 t = 0 B2 t = 0		B1 t = 0 B2 t = 6 h		B1 t = 6 h B2 t = 0		B1 t = 0 B2 t = 22 h		B1 t = 22 h B2 t = 0	
Lanes	1		3	4	5	6	7	8	9	10
CP	+	+	+	+	+	+		+	+	+
B1	+		+		+		+			
B1-546	-	+	-		-		-	+	-	
B2	-		-		-		-		-	
B2-546	+		+		+		+		+	

FIG 10 EMAs of competition experiments between B1 and B2 at different times and for different buffer and gel-staining conditions. (a) The signal comes only from AF546-labeled RNA. (b) Same gel stained with ethidium bromide showing the total RNA content. From left to right: the first two lanes show competition experiments in which both RNAs were added at the same time to RAB (pH 7.2); lanes 3 and 4 show competition experiments in which B2 was added after B1, at pH 7.2; and lanes 5 and 6 show the opposite case, again at pH 7.2. The last four lanes show the effect of adding a second RNA after lowering the pH to 4.5 (VSB).

These results demonstrate that the assembly process of CCMV CP around RNA at neutral pH is reversible: CP, RNA, and RNA/CP complexes are in equilibrium with their surroundings. It is only upon pH acidification that true -- RNase-resistant -- capsids are formed, which are no longer at equilibrium, e.g., whose CPs cannot be exchanged upon addition of new RNA. Similar conclusions were reached recently by Tsvetkova et al.[37], who studied the assembly of BMV CP around gold nanoparticles functionalized with carboxylates. They concluded that at neutral pH

the CP saturated the anionic particles, adsorbing noncooperatively and forming noncompact poorly organized structures. The formation of a well-defined, compact structure occurred in a co-operative fashion only after the pH was lowered to 4.5.

DISCUSSION

A key to interpreting the results we have found for the relative packaging efficiencies of different RNA molecules -- by means of their head-to-head competition for CP with a reference RNA under conditions where the protein is in short supply -- can best be presented in terms of the two-step assembly reaction involved in nucleocapsid formation. More explicitly, it is useful to treat the reversible and irreversible behavior characterizing, respectively, the RNA/CP complex formation in the first (neutral-pH) step and the RNase-resistant nucleocapsid formation in the second (acidic- pH) step.

Our picture of RNA/CP complex formation at neutral pH is consistent with the *"en masse"* scenario suggested on the basis of simulations by Elrad and Hagan[21] in which strong, nonspecific, CP/RNA interactions[25] lead to RNA molecules "saturated" with capsid proteins. A similar scenario was proposed by Devkota et al.[18] in their molecular mechanics interpretation of high-resolution structural data for Pariacoto virus, and earlier in a general *"micellization"* discussion by McPherson[27] of simple self-assembling RNA viruses. The basic idea is that nonspecific electrostatic interactions between the basic residues of the proteins and the phosphate backbone drive the binding of the CP to the RNA and induce its condensation into a template for nucleocapsid formation.

66

At neutral pH, in the case of CCMV CP, we propose that in a competition experiment between two different RNA species at a CP/total RNA mass ratio of 3:1, essentially all the CP is bound roughly half on each of the RNAs. On average, however, the (smaller number of) longer RNAs will have more CPs bound to them, at the same density as on the shorter RNAs; this will lead, upon pH lowering, to a higher rate of nucleation on the longer molecules. At neutral pH, thermal fluctuations drive the CPs from one RNA molecule to another, and the structure of the RNA/CP complexes will be polydisperse, including various intermediates that involve some cooperative lateral interactions between CPs but that remain in thermal equilibrium with their surroundings. These *"procapsid"* states are therefore still labile and will be ordered ("fixed") into RNase-resistant, *"mature"* capsids only upon pH lowering.

The kinetic and thermodynamic factors affecting the formation of a mature VLP from a reversibly bound *procapsid* can be largely understood using the framework of classical homogeneous nucleation theory[17] (homogeneous rather than heterogeneous nucleation applies to the extent that no specific origin of assembly plays a dominant role in the formation of CCMV capsids around RNA). The classical nucleation rate is the number of critical nuclei formed per unit time per unit system volume, with the critical size varying inversely with the free-energy difference ($\Delta\mu$) per molecule between the initial and final states. The nucleation rate is proportional to $\exp(-\Delta G^*/k_B T)$, where $k_B T$ is the thermal energy and ΔG^* -- the barrier to forming a critical nucleus -- is a decreasing function of $\Delta\mu$. In this context, then, we can immediately identify two factors that will affect the nucleation rate of VLPs and hence the relative efficiency of packaging. First, the number of nucleation events per unit time increases with system size: all other things being equal, nucleation is more probable in longer than in

shorter RNAs. Second, the more stable the VLP formed the larger will be the magnitude of $\Delta\mu$ (i.e., the energy difference per molecule between the *procapsid* and *mature capsid*) and the greater the nucleation rate as well. We apply these considerations in discussing the measured relative packaging efficiencies.

Consider, for example, the case of competition between 3.2-kb (B1) and 2.5-kb RNA (or some other RNA, e.g., 2.0-kb RNA, that can compete at least a bit with B1). As discussed above, to a good approximation each molecule competes equally well for binding of CPs and is, on average, half-covered with protein, the 2.5-kb RNA with about 125 CPs and B1 with about 160 CPs. At neutral pH, the RNA/CP complexes arising from 2.5-kb RNA and B1 mixtures are in thermal equilibrium in the sense that they have the same composition and structures -- whatever those structures may be (e.g., RNA decorated with bound CP, or loosely formed procapsids, or anything in between) -- independent of the order in which the RNAs are added to CP. When the pH is lowered, the protein-protein interactions become stronger, leading to a greatly enhanced nucleation rate for capsid formation, with the longer molecules enjoying a higher rate of nucleation than the shorter ones. As a result, unlike the case at neutral pH where the CP binding and unbinding rates are essentially identical for all of the RNAs, once nucleation occurs, the longer RNA has a smaller off-rate for protein unbinding and fluctuations will lead to increasingly complete, partially formed, mature capsids. Nucleation can also occur on the shorter molecules, of course, but at a lower rate than for the longer ones, thereby giving rise to their smaller packaging efficiency.

Recall that molecules shorter than 2.0 kb, when mixed with CP at a 6:1 mass ratio in the absence of competitor, are packaged completely but only in multiple copies, e.g., two 1.0-kb RNAs in a T = 2 VLP or two 1.5-kb RNAs in a T = 3 VLP. A short molecule like this cannot

compete effectively against B1 in being packaged because, in addition to the nucleation rate being lower on a shorter molecule, it has to wait -- after a critical nucleus begins to grow -- for another short-RNA/CP complex to come along in order for VLP formation to be complete. (We cannot yet rule out the possibility that the shorter molecules have already come together at neutral pH.) In the case of the 1.0-kb RNA, nucleation rates are still smaller because $T = 2$ capsids have higher energy (per CP) than $T = 3$; i.e., $\Delta\mu$ is smaller than it is for $T = 3$ capsids with longer RNA.

While the order of adding RNAs at high pH is not of consequence for competition outcomes because of thermal equilibration at neutral pH, the order does matter when one of the two RNAs is added at acidic pH. For example, consider the case where only 1.0-kb molecules are mixed with CP and the solution dialyzed first against neutral pH (RAB) for 20 h and then against acidic pH (VSB) for 2 h. Even though the nucleation rate for fully saturated 1.0-kb molecules at acidic pH is lower than that for fully saturated 3.2-kb RNAs, and even though a second 1.0-kb RNA must find the nucleated complex, these processes occur well before the 2-h point when the 3.2-kb molecule is added. Accordingly, the short RNA is completely packaged.

The packaging efficiency of a 3.6-kb RNA differs little from that of a 2.8-kb RNA, indicating that the CCMV CP is not very dis- criminating when packaging RNAs close to the average viral length ($\approx 3,000$ nt). This agrees with the results of Porterfield et al.[29], who found that hepatitis B virus capsid protein has no strong preference for its viral RNA when the assembly is done in the presence of a heterologous RNA of comparable length. The surprising result is that a 4.0-kb RNA has the highest relative packaging efficiency of all the competitor

69

RNAs tested. When B1 competes against a 4.0-kb RNA at an overall CP/RNA mass ratio of 3:1, both RNAs have, on average, half the amount of protein bound at the magic ratio. The longer RNA complex will have a higher rate of nucleation of CPs into T = 3 capsids, but on the other hand, the resulting VLP has a higher energy (a smaller $\Delta\mu$) than that of the 3.0-kb molecule when it is confined in the T = 3 capsid, thereby partially offsetting the effect of length.

In all samples and control assemblies, with the exception of a small fraction of elongated particles that arise when CCMV CP is assembled around a 4.0-kb RNA, only spherical VLPs were observed. We have shown earlier that for RNAs ≥ 4.5 kb, one RNA is shared by two or more capsids (multiplets)[12]; accordingly, it is not surprising that, when approaching this length (e.g., for 4.0-kb RNA), intermediate structures between doublets (two capsids sharing one RNA) and singlets (one RNA per capsid) can be formed. A simple calculation shows that if two T = 3 half-capsids are connected by a ring of hexamers, the axial ratio of the resulting particle will be 1.5, in agreement with that observed. Moreover, the absence of elongated capsids when B1 competes against a 4.0-kb RNA shows that their free energy is considerably higher (hence smaller $\Delta\mu$) than those of T = 3 and T = 4 capsids.

So far, all of the head-to-head competition experiments we have discussed involve two RNA molecules of different length. The competition between RNAs of identical length but different sequences revealed a striking result: even though C1 is the homologous RNA to CCMV CP, its packaging efficiency is lower than that for B1. The two RNAs share 80% sequence homology and B1 is only 67 nt longer, so they are essentially of identical length. They also appear to have approximately the same three-dimensional (3D) size, as evidenced by their having

similar electrophoretic mobilities (data not shown) and hence similar hydrodynamic radii, even though in general long RNAs of identical length can have very different mobilities and sizes[22, 38]. *Why, then, is C1 less efficiently packaged by its capsid protein than the heterologous B1?* It is likely that there are some features of its secondary/tertiary structure that give rise, locally, to a lower barrier to nucleation -- in this sense suggesting some heterogeneous contributions to the nucleation process that occurs on the CP-bound RNAs at acidic pH.

The difference in packaging efficiency between RNAs of essentially identical length requires us to acknowledge that such structural contributions are likely to play a role in addition to length, which has been our focus. Further studies comparing the efficiencies of RNAs of identical lengths but different sequences will be necessary in order to separate these effects. Finally, we remark that one should be able to identify an equal (3.2-kb)-length nonviral sequence RNA that competes even more efficiently against C1 RNA for CCMV capsid protein than does B1. After all, the viral genome has a large number of evolutionary pressures on it beyond needing to be efficiently packaged by its CP; e.g., it needs to be replicated and transcribed efficiently to code properly for its several gene products and to be able to disassemble efficiently from the capsid once it has penetrated into the cell. A competitor for CP, on the other hand, needs only to have secondary and tertiary structure and overall size and shape that are better optimized for packaging that those of the viral genome.

Most *in vitro* assemblies of RNA viruses are performed at neutral pH, and CCMV is no exception. What we have shown is that under these conditions RNA and CP are in equilibrium and RNA/CP complex formation is affected by the addition of more RNA (even 6 h after they were initially mixed). It is well known that CCMV CP-CP interactions increase as pH

decreases[2]; moreover, as pH is lowered sufficiently and in the absence of RNA, CCMV CP can self-assemble into tubes, multishells, and empty capsids[1,8,26]. So by taking into account that RNA and CP strongly interact at neutral pH[12] and that CP-CP interactions at acidic pH are stronger than at neutral pH, it is natural for us to propose that CCMV CP self-assembly around RNA involves at least two steps.

While the focus of this research has been on *in vitro* assembly, it is appropriate to ask if a pH-driven two-step assembly has any relevance for *in vivo* assembly. It is known that in the case of flaviviruses[35], *in vivo* capsid maturation is driven by pH acidification. For these viruses, immature capsids are assembled at the endoplasmic reticulum (pH 7.2) and are transported through the secretory pathway to the trans-Golgi network, where maturation occurs upon pH lowering (pH 5.7)[35]. The assembly pathway of bromoviruses *in vivo* has been less well studied, but recent fluorescence experiments[7] have pointed to the role of subcellular localization of CP in vesicles in which there is a colocalization with sites of replication; the relative amounts of CP and RNA in these sites are unknown. We cannot rule out the possibility that cytopathology induced by virus infection could alter the normal pH in these sites, making the environment more congenial for virus assembly.

In summary, we have shown that the in vitro packaging efficiency of CCMV CP depends nonmonotonically on the length of ssRNA. When a competitor RNA is significantly shorter than the reference RNA (< 2.0 kb), it is not packaged, but if it is still smaller (≤ 0.5 kb), it is copackaged with the reference RNA. This is the case in BMV and CCMV, in which RNA3 and its subgenomic RNA4 (whose sequence is identical with the 3' half of RNA3) are copackaged[33].

72

For RNAs larger than 2.0 kb, their relative packaging efficiencies increase with RNA length, although none of them are more efficiently packaged than the reference RNA. We have also shown than when two RNAs of the same length, but different sequences, have to compete against each other for CP, it is likely that their secondary and tertiary structures play a crucial role. These results suggest that the overall 3D sizes of the RNAs affect the thermodynamics of their nucleocapsid assembly, while the secondary (and tertiary) structures affect the kinetics of nucleation, especially when the two RNAs have the same charge and overall size. In particular, we have shown that B1 has a higher packaging efficiency than C1, which is the homologous RNA for CCMV CP. Finally, we have shown that at neutral pH, RNA and CCMV CP form complexes that are in thermal equilibrium with respect to exchange of CP, while after lowering of the pH to acidic conditions, RNase-resistant nucleocapsids arise irreversibly from these complexes.

REFERENCES

1. **Adolph KW, Butler PJG.** 1974. Studies on the assembly of a spherical plant virus. I. States of aggregation of the isolated protein. J. Mol. Biol. **88**:327–338.

2. **Adolph KW, Butler PJG.** 1976. Assembly of a spherical plant virus. Phi- los. Trans. R. Soc. Lond. B Biol. Sci. **276**:113–122.

3. **Adolph KW, Butler PJG.** 1977. Studies on the assembly of a spherical plant virus. III. Reassembly of infectious virus under mild conditions. J. Mol. Biol. **109**:345–357.

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

5. **Aniagyei SE, et al.** 2009. Synergistic effects of mutations and nanoparticle templating in the self-assembly of cowpea chlorotic mottle virus capsids. Nano Lett. **9**:393–398.

6. **Annamalai P, Rao ALN.** 2005. Dispensability of 3' tRNA-like sequence for packaging cowpea chlorotic mottle virus genomics RNAs. Virology **332**:650 – 658.

7. **Bamunusinghe D, Seo J-K, Rao ALN.** 2011. Subcellular localization and rearrangement of endoplasmic reticulum by brome mosaic virus capsid protein. J. Virol. **85**:2953–2963.

8. **Bancroft JB, Hills GJ, Markham R.** 1967. A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro. Virology **31**:354 –379.

9. **Bancroft JB.** 1970. The self-assembly of spherical plant viruses. Adv. Virus Res. **16**:99 –134.

10. **Belyi VA, Muthukumar M.** 2006. Electrostatic origin of the genome packaging in viruses. Proc. Natl. Acad. Sci. U. S. A. **103**:17174 –17178.

11. **Cadena-Nava RD, et al.** 2011. Exploiting fluorescent polymers to probe the self-assembly of virus-like particles. J. Phys. Chem. B **115**:2386 –2391.

12. **Cadena-Nava RD, et al.** 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.

13. **Chang CB, Knobler CM, Gelbart WM, Mason TG.** 2008. Curvature dependence of viral protein structures on encapsidated nanoemulsion droplets. ACS Nano **2**:281–286.

14. **Chen C, Kwak Stein B, Kao CC, Dragnea B.** 2005. Packaging of gold particles in viral capsids. J. Nanosci. Nanotechnol. **5**:2029 –2033.

15. **Choi YG, Rao ALN.** 2010. Molecular studies on bromovirus capsid protein. VII. Selective packaging of BMV RNA4 by specific N-terminal arginine residue. Virology **275**:207–217. 919.

16. **Comellas-Aragonès M, et al.** 2009. Controlled integration of polymers into viral capsids.

Biomacromolecules **10**:3141–3147.

17. **Debenedetti PG.** 1996. Metastable liquids: concepts and principles, p 148 –159. Princeton University Press, Princeton, NJ.

18. **Devkota B, et al.** 2009. Structural and electrostatic characterization of Pariacoto virus: implications for viral assembly. Biopolymers **91**:530 –537.

19. **Douglas T, Young MJ.** 1998. Host-guest encapsulation of materials by assembled virus protein cages. Nature **393**:152–155.

20. **Dreher TW, Rao ALN, Hall TC.** 1989. Replication in vivo of mutant brome mosaic virus RNAs defective in aminoacylation. J. Mol. Biol. **206**: 425– 438.

21. **Elrad OM, Hagan MF.** 2010. Encapsulation of a polymer by an icosaheral virus. Phys. Biol. **7**:045003– 045020.

22. **Gopal A, Zhou HZ, Knobler CM, Gelbart WM.** 2012. Visualizing large RNA molecules in solution. RNA **18**:284 –299.

23. **Hu Y, Zandi R, Anavitarte A, Knobler CM, Gelbart WM.** 2008. Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. Biophys. J. **94**:1428 – 1436.

24. **Johnson JM, Willits DA, Young MJ, 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.

25. **Katen S, Zlotnick A.** 2009. The thermodynamics of virus capsid assembly. Methods Enzymol. **455**:395– 417.

26. **Lavelle L, et al.** 2009. Phase diagram of self-assembled viral capsid protein polymorphs. J. Phys. Chem. B **113**:3813–3820.

27. **McPherson A.** 2005. Micelle formation and crystallization as paradigms for virus assembly. Bioessays **27**:447– 458.

28. **Minten IJ, et al.** 2011. Metal-ion-induced formation and stabilization of protein cages based on cowpea chlorotic mottle virus. Small 7:911-919.

29. **Porterfield JZ, et al.** 2010. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. J. Virol. **84**:7174 –7184.

30. **Porterfield JZ, Zlotnick A.** 2010. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. Virology **407**:281–288.

31. **Pyne JW, Hall TC.** 1979. Efficient ribosome binding of brome mosaic virus (BMV) RNA4 contributes to its ability to outcompete the other BMV RNAs for translation. Intervirology **11**:23–29.

32. **Qu F, Morris TJ.** 1997. Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. J. Virol. **71**:1428 –1435.

33. **Rao ALN.** 2006. Genome packaging by spherical plant RNA viruses. Annu. Rev. Phytopathol. **44**:3.1–3.27.

34. **Reusken CBEM, Neeleman L, Bol JF.** 1994. The 3'-untranslated region of alfalfa mosaic virus RNA 3 contains at least two independent binding sites for viral coat protein. Nucleic Acids Res. **22**:1346 –1353.

35. **Rodenhuis-Zybert IA, Wilschut J, Smit JM.** 2011. Partial maturation: an immune-evasion strategy of dengue virus? Trends Microbiol. **19**:248 –254.

36. **Sambrook J, Fritsch EF, 36. Maniatis T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

37. **Tsvetkova I, et al.** 2012. Pathway switching in templated virus-like particle assembly. Soft

Matter **8**:4571– 4577.

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

Characterization of viral capsid protein self-assembly around short single-stranded RNA[*]

INTRODUCTION

In single-stranded RNA (ss-RNA) viruses genome packaging is unequivocally coupled to capsid assembly. This process is spontaneous, driven by capsid protein-capsid protein (CP) and CP-RNA interactions. As a result ssRNA viruses do not need a virally-encoded enzyme (molecular motor) to package their genome, and with only a few exceptions empty capsids are not formed during a viral infection. Understanding how RNA and CP interact *in vivo* to build a capsid during a viral infection has been show to be a challenging task, mostly because of the multiple roles of the CP beyond genome packaging, i.e., suppression of RNA silencing or the control of minus- and plus-strand synthesis (1). Hence extensive work has been done on *in vitro* packaging studies of viral and non-viral RNAs by capsid protein. These rather simple approaches have highlighted the importance of both specific and non-specific interactions that give rise to virus assembly (2-7).

In an effort to understand how such interactions direct cowpea chlorotic mottle virus (CCMV) virion capsid assembly, we have previously focused on how the relation between the size of the packaged RNA affects the size of the capsid and its efficiency of formation (7,8). We have shown that CCMV CP can spontaneously self-assemble around viral (homologous and

[*] This work was carried out in collaboration with Rees F. Garmann from UCLA and will be submitted to the *Biophysical Journal.*

heterologous) and virally-derived RNAs ranging in length from 140 to 12,000 nucleotides (nt) (7). In particular, if the RNA is significantly shorter than the 3200-nt wild-type (WT) (≤ 1000 nt) several RNAs are packaged in a single capsid, such that the RNA content is almost the same as that of the WT. For RNAs with lengths close to the WT (≥ 2000nt, ≤ 4000nt), only one molecule is packaged and the capsids vary in size (7) from "pseudo $T = 2$" (i.e., $T = 1$ where the subunits are CP dimers rather than monomers) to $T = 3$ to $T = 4$. However, if the RNA is too long (≥ 5000 nt) one RNA molecule is shared by two or more capsids.

The *in vitro* results for short RNAs -- indicating that two or more molecules are packaged -- are consistent with the *in vivo copackaging* of RNAs 3 and 4 in the case of several species of the *Bromovirus* genus. In fact, among multipartite ssRNA viruses, copackaging of two or more RNA molecules is a very common strategy, which in some cases leads to packaging of subgenomic RNAs that are not necessary for infection (9,10). However, to date there are no conclusive studies on the mechanism(s) that lead to multiple RNAs being packaged in a single capsid, mainly because during the course of an infection it is almost impossible to uncouple RNA replication from capsid assembly (9-11). In the present study we examine RNA copackaging *in vitro* by using CCMV CP and a 500-nt long RNA. Our measurements are aimed at understanding, in the simplest experiments and in the absence of any specific RNA-CP interactions, how short equal-length RNAs can be copackaged by CCMV CP.

We find that, unlike the case for longer RNAs ($\geq$ 1000-nt) (7,12), the self-assembly of CCMV CP around 500 nt RNAs is cooperative; at CP:RNA mass ratios smaller than the minimum needed to package all of the RNA there are only two species in solution – CP-bound single molecules of RNA and virus-like-particles (VLPs) consisting of several RNAs bound to CP. By using electrophoretic mobility assays, velocity sedimentation, negative-stain and cryo-

electron microscopy, and single-molecule fluorescence correlation spectroscopy (sm-FCS) we show that at neutral pH, RNA and CP form amorphous complexes that are smaller than a WT viral capsid. We also show that the number of such RNA-CP complexes that aggregate to form *protocapsids* depends on the absolute RNA concentration and that fully-formed nucleocapsids (VLPs) are formed only upon acidification. Finally, we find from that four 500-nt RNAs are packaged into each VLP. On the basis of these results we argue that the cooperativity shown by short RNAs is a consequence of the equilibrium between multiple-RNA CP/RNA complexes and single-RNA CP/RNA complexes, i.e., by the aggregation of single-RNA CP-bound complexes into multiple-RNA complexes.

METHODS AND MATERIALS

Materials. Restriction enzymes were obtained from New England Biolabs. T7 RNA polymerase was a gift from Prof. Feng Guo (Dept. of Biological Chemistry, UCLA). Enzymes were used as recommended by the manufacturer. All other reagents used were DNase-, RNase- and Protease-free.

PCR amplification of DNA templates to transcribe RNAs. The cDNA for the 500-nt RNA was obtained by PCR of the first 500 bp of the pT7B1 plasmid (13). B1 refers here to the first gene of the brome mosaic virus (BMV); we used a universal 5' primer d(TAATACGACTCACTATAGGTAGACCACGGAACGAGGTTC) (T7 promoter is underlined) and d(CACATCCTCTCCTCATGTC) and d(GTCTTCAAACCATACACAGTG) as reverse primers. The templates were purified by standard procedures (14).

RNA transcription and fluorescence labeling. The DNA template was transcribed with T7 RNA polymerase and labeled with ChromaTide Alexa Fluor 488-5-UTP (AF488) (Invitrogen). This modified rUTP is randomly incorporated during transcription by adjusting the molar ratios of rATP:rGTP:rCTP:rUTP:modified-rUTP to 600:600:600:5.32:1, giving a density of labeling around 0.5. The label density for the RNA was calculated by linear regression of a calibration curve of pure AF488-rUTPs. The fluorescence measurements were carried out with a QuantaMaster Spectrofluorimeter (Photon Technology International) in a 40-μL quartz cuvette (Sterna Cells). Fluorescent RNAs were always kept in the dark in amber tubes and room light was minimized when possible during handling.

CCMV CP purification. CCMV was purified from infected California cowpea plant (*Vigna ungiculata cv* Black Eye) (15) and CP was isolated as described previously (4). SDS-PAGE and MALDI-TOF were employed to ascertain that the purified protein was intact.

CCMV CP labeling. Alexa Fluor-647 (AF647) succinimidyl ester (Invitrogen) was covalently linked to solvent-exposed lysines on the exterior surface of CCMV virions (16). CCMV was concentrated to 5 mg/mL and then dialyzed overnight at 4ºC against 0.1 M HEPES, 5 mM MgCl$_2$ pH 7.2. The dye-conjugation reaction was carried out at room temperature for 2 h by mixing the virus with AF647 (10 mg/mL) at an AF647:CCMV (virion) mass ratio of 0.005 (AF647:CCMV (virion) molar ratio of 18.4:1) and then dialyzing against VSB overnight at 4ºC. A sucrose cushion was used to purify the sample from the unreacted dye: 2.5 mL of sample plus 0.5 mL of 10% sucrose in VSB were centrifuged for 2 hours at 100,000 rpm and 4 ºC in a TLA 110 rotor, and the supernatant was then discarded and the pellet resuspended with VSB. If CP

was not needed immediately the virus was frozen at -80ºC; otherwise it was purified as described above. The density of labeling (DOL), defined as the fraction of CPs labeled by AF647, was determined before and after CP purification by UV-Vis spectroscopy according to the manufacturer. More explicitly, $DOL = \frac{Abs_{650} \times Mw}{[Protein] \times \varepsilon_{AF647}}$, where Abs_{650nm} is the absorbance maximum of AF647, M_W is the molecular weight of the CCMV CP (20.3 KDa), [Protein] is the CP concentration in mg/mL, and ε_{AF647} is the molar extinction coefficient of the dye (239000 M^{-1} cm^{-1}).

In vitro **assemblies**. Unless otherwise stated the RNA concentration in every reaction was 30 ng/μL (150 nM RNA). We followed two different assembly protocols: one- and two-step dialysis. For both protocols the first step was to mix the CP and RNA in Buffer B (1M NaCl, 20 mM Tris-HCl pH 7.2, 1 mM EDTA, 1 mM DTT and 1 mM PMS). For the one-step assembly, the samples were then dialyzed for 24 h at 4 °C against 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). For the two-step assembly a 12-h dialysis against RAB was followed by a 12-h dialysis against virus suspension buffer (*VSB*: 50 mM sodium acetate, 8 mM magnesium acetate pH 4.5). All reactions were carried out at 4ºC and protected from the light.

Electrophoretic mobility analysis (EMA). After each assembly, a 10-μL aliquot of each sample and control was mixed with 3 μl of glycerol and loaded into a 1% agarose gel (EMD OmniPur) in virus electrophoresis buffer (0.1 M sodium acetate, 1 mM EDTA, pH 5.5). The samples were electrophoresed at 4 °C for 1.25 h at 65 V in a horizontal gel apparatus (Fisher)

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 AF546, AF488 and ethidium bromide.

Single-molecule Fluorescence Correlation Spectroscopy (sm-FCS). Sm-FCS measurements were performed using a confocal microscope (Olympus IX71). Emission lines at 488 and 637 nm of a continuous-wave argon ion laser (IonLaser Technology, Frankfort, IL) were used as excitation sources. The details of the experimental setup are described elsewhere (17). The excitation volume was of the order of 1 fL, and before and after every experiment it was determined by measuring the diffusion coefficient of a purified sample of fluorescently labeled 500-nt RNA VLPs (AF488-RNA and AF647-CP) at the same concentration as the samples. The fluorescence signal was detected by two avalanche photodiodes (APDs, AQR-14, Perkin-Elmer Inc.). Detected photon pulses were sent to a hardware correlator card (ALV-60010, ALV GmbH, Langen, Germany), which computed the cross-correlation of the two channels with a temporal resolution of 6.5 ns.

FCS uses fluorescence fluctuations within the confocal volume to extract properties such as the diffusion coefficient of a fluorescently tagged molecule, as well as fluorescence kinetics due to dark states (triplet-state and photon-bunching). Fluorescence fluctuations, *I(t)*, were analyzed using an autocorrelation function:

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \qquad \text{Eq. (1)}$$

The contribution of diffusion, *G(τ)$_D$*, to the *G(τ)* can be described, under the assumption of an elliptical Gaussian-shaped confocal volume, by (17):

$$G(\tau)_D = \left[\langle N\rangle(1 + \frac{\tau}{\tau_D})\right]^{-1} \qquad \text{Eq. (2)}$$

Here <N> is the average number of particles in the confocal volume and τ_D is the characteristic diffusion time, which depends on the diffusion coefficient, D, and the beam waist, ω_{xy} in the x,y-dimensions of the laser focus:

$$\tau_D = \omega_{wy}{}^2/4D \qquad \text{Eq. (3)}$$

The effective hydrodynamic radius, R_h, was determined by using the Stokes-Einstein relation:

$$D = \frac{k_B T}{6\pi\eta R_h} \qquad \text{Eq. (4)}$$

where k_B, T and η are the Boltzmann constant, the temperature and solvent viscosity, respectively.

The measured G(τ) also includes contributions from intra- and/or intermolecular dynamics that give rise to fluorescence fluctuations on time scales much faster than the characteristic diffusion time, τ_D. Such fluctuations can be described by a simple exponential decay for the dark states:

$$\chi(\tau)_{kinetics} = 1 - B + B \cdot e^{-\frac{\tau}{\tau_{triplet}}} - A + A \cdot e^{-\frac{\tau}{\tau_{bunching}}} \qquad \text{Eq. (5)}$$

Here B and $\tau_{triplet}$ are the fraction and characteristic lifetime of fluorophores in their triplet state, and A and $\tau_{bunching}$ are the amplitude and characteristic lifetime pf the photon-bunching, with $\tau_{triplet} < \tau_{bunching} < \tau_D$. The typical values for the lifetime of the triplet-state, photon-bunching and the diffusion time are 10^{-3}, 10^{-1} and 10^1-10^2 ms, respectively. The measured $G(\tau)$ can be written as:

$$G(\tau) = G(\tau)_D \cdot \chi(\tau)_{kinetics} \qquad \text{Eq. (6)}$$

When there are two diffusing species with two different quantum yields the autocorrelation curve can be re-expressed as:

$$G(\tau)_D = \frac{\alpha_1^2 N_1 G(\tau)_{D,1} + \alpha_2^2 N_2 G(\tau)_{D,2}}{(\alpha_1 N_1 + \alpha_2 N_2)^2} \qquad \text{Eq. (7)}$$

84

where $G(\tau)_{D,1}$ and $G(\tau)_{D,2}$ are the diffusive parts of the autocorrelation correlation function for species 1 and 2, respectively, α_2, N_1 and N_2 are the ratio of the quantum yield of species 2 with respect to that of species 1, and the number of particles of species 1 and 2, respectively. By definition $\alpha_1 = 1$.

sm-FCS sample preparation and measurements. For CP:RNA titration curves the samples where diluted 15-fold in a single step (from 150 nM RNA to 10 nM) while for the dilution series the samples where diluted in serial steps from 150 nM to 100, 50, 25, 10, 5, 1, 0.5, 0.25, and 0.05 nM RNA. To avoid non-specific adsorption of the samples at these low concentrations both buffers (either RAB or VSB) used for dilutions contained 0.05% (v/v) Tween-20. All measurements were performed at room temperature and each sample was measured 40 times with a 10-s acquisition time per run. The data were analyzed by two methods: by non-linear least squares fitting for each run, taking the average of the fitting parameters for the 40 runs; and by first averaging all the $G(\tau)$s and then by a non-linear least squares fitting the data and using the standard deviation of the average $G(\tau)$ as the instrumental error. Both methods gave the same results and the quality of the fitting procedures was comparable. The data were analyzed by using QtiPlot data analysis software (ProIndep Serv).

Velocity centrifugation in sucrose gradients. Each 10-40% sucrose gradient was made by adding 5.2 mL of a 20% sucrose solution in the desired buffer (RAB or VSB) to a thin-wall tube for a SW 50.1 rotor (Beckman-Coultier). The tubes were balanced (0.001 g precision), parafilmed and frozen at -80 ºC for at least 1 h, then thawed at 4 ºC for 2 h and refrozen. This freezing-thawing cycle was repeated two more times, resulting in very reproducible 10-40%

85

sucrose gradients. The gradients were loaded with 0.2 mL of the sample, balanced again and centrifuged for 3 h at 30,000 rpm at 4 ºC, and then were manually fractionated by taking 0.2-mL aliquots from top to bottom. The pipet tip was changed after collecting each fraction. The 26 extracted fractions were loaded into a 96-well flat-black bottom plate. The fluorescence for the CP and the RNA (AF647 and AF488) were measured for each fraction using a Tecan M1000 plate reader.

Negative-stain and cryo-EM analysis of VLPs. For negative staining, purified VLPs were applied to glow-discharged 400-mesh copper grids (Ted Pella) that previously had been coated with Parlodion (SP1-Chem) and carbon (Ted Pella). A 6 µl 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. Each reported average diameter of VLPs is the geometric mean of two orthogonal measurements from 1000 particles, determined with ImageJ (U.S. National Institutes of Health) software from recorded images.

Cryo-EM samples were prepared by depositing 3uL of assembly reaction 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 a liquid ethane bath cooled to liquid nitrogen temperature. Micrographs were acquired using the FEI Tecnai G2 TF20 microscope operated at an accelerating voltage of 200 kV. Total beam exposure was maintained between 30-50 e/$Å^2$.

Images were recorded at 3-4 microns underfocus with a TIETZ F415MP 4k X 4k pixel CCD camera.

RESULTS

RNA and CP labeling does not affect *in vitro* assembly. The average DOL of the CP turned out to be extremely important. At a virion DOL of 0.7 (corresponding to about 130 labels per capsid) the degree of capsid disassembly at pH 7.5 (even at $I = 2$ M) was one order of magnitude lower than that for the WT virions. A DOL ten times smaller was found to be optimum for efficient capsid disassembly and labeled-CP recovery. The electrophoretic mobility of the AF657-labeled virions was greatly increased by the labeling (data not shown), because covalent bonding of each singly-charged anionic label to a lysine results in a decrease of one positive charge and hence an *increase* in overall *negative* outer-surface charge of the capsid. It has been shown that under the conditions at which the labeling reaction takes place the lysines that are more likely to react are the solvent-exposed ones on the outer surface of the capsid (16).

To examine if the change in the CP charge affected the assembly process, we carried out assemblies at different ratios of labeled:unlabeled CP around AF488-labeled 500-nt RNA (AF488-RNA), as well as unlabeled RNA. Electrophoretic mobility assays (EMA) and negative-stain electron micrographs (EM) showed that VLP formation is not affected by varying the ratios of labeled:unlabeled CP (with either labeled or unlabeled RNA, data not shown); further, the previously established *magic ratio* holds, and as in previous studies only "pseudo-T=2" icosahedral capsids were formed with 500-nt RNA (7,8). In our previous study we had found that independent of length and sequence all of the ssRNA in solution can be packaged by CCMV CP if the CP:RNA mass ratio is 6:1, and we denoted this threshold value as the *magic ratio* (7).

87

However, the electrophoretic mobility of AF647-labeled-VLPs was indistinguishable from that of unlabeled-VLPs only when the CP DOL ≤ 0.05 (data not shown). It has been reported (16) that chemical modification of WT CCMV virions can result in an increase in their diameter, but the negative-stain EMs show that even at a CP DOL ~ 0.7 the diameter of the modified virions is indistinguishable from that of WT (data not shown).

Assembly of CCMV CP around 500-nt RNA is cooperative. It has been previously demonstrated that the assembly of CCMV CP around RNAs with lengths close to or greater than the wild-type (≈ 3200 nt) is non-cooperative (7,12); surprisingly, the EMA profile of the CP:RNA titration curve with a fixed amount of 500-nt RNA showed that this is not the case for 500-nt RNA. This titration curve was obtained by following the two step-assembly protocol: 12-h dialysis against RAB (pH 7.2) followed by 12 h dialysis against VSB (pH 4.5). As evident from Fig. 1, at CP:RNA mass ratios < 3 there are two main species -- RNA, and CP/RNA complexes that run more slowly than RNA; at CP:RNA > 3 there are mostly VLPs and a small fraction of free RNA. Only at CP:RNA = 3 are there comparable amounts of both. Fig. 1b shows that most of the CPs are associated with CP/RNA complexes and VLPs. In accordance with our previous results (7) this titration curve shows that the minimum CP:RNA mass ratio needed to package 100% of the RNA into VLPs is 6:1.

To determine if this cooperative behavior could be caused by a rapid change in pH that occurs when the assembly mixture is loaded in the gel (the EMA is run at pH 5.5), we dialyzed the assemblies only against RAB. This titration curve shows the same cooperative behavior at CP:RNA > 3, but at lower mass ratios there are more than two species (see Fig. 2), suggesting that at low CP:RNA mass ratios a rapid change in pH is not effective in reorganizing the CP.

88

These results show that the degree of cooperativity during capsid assembly depends not only on the nature of the CP (18) but, at least for CCMV, it depends on the length of the packaged RNA as well.
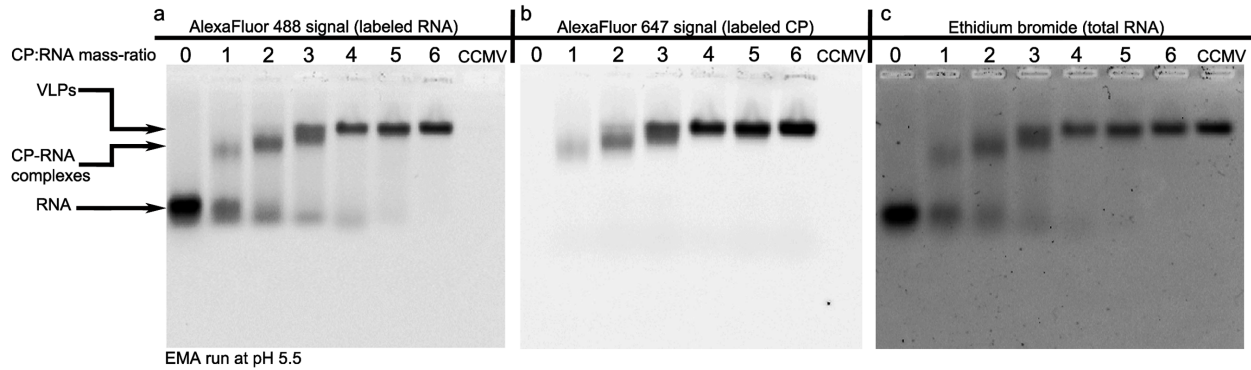

FIG 1. Electrophoretic mobility assay of 500-nt RNA with different amounts of CCMV CP. 1% agarose electrophoretic mobility assay run at pH = 5.5 and visualized by exciting either a) fluorescently labeled RNA (AF488), b) CP (AF647) or c) stained RNA (ethidium bromide). A constant amount of AF488-500-nt RNA was titrated with AF647- CP, and the samples were assembled with the two-step protocol described in the text. Going from left to right, the mass ratio of CP:RNA in each lane increases from 0:1 to 6:1. These EMAs show that the 500-nt RNA capsid assembly is highly cooperative and that the minimum CP:RNA mass ratio at which all of the RNA is packaged is 6:1 (*magic ratio*).
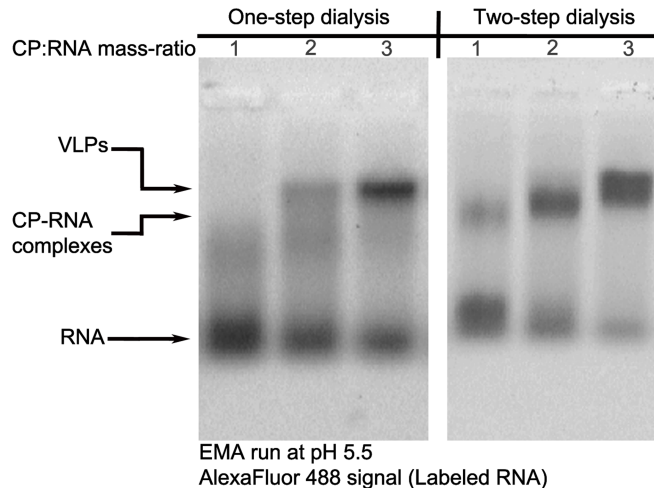
FIG 2. Rapid acidification does not promote cooperativity at low CP:RNA mass ratios. The EMAs show the difference between samples dialyzed only against RAB (one-step dialysis) and those against RAB and VSB (two-step dialysis, see Fig. 1a). For CP:RNA mass ratios ≤ 3, rapid pH acidification due to the conditions at which the EMA are run was not sufficient to promote a two-state system. By comparing both EMAs it is clear that low-molecular weight CP/RNA complexes formed during pH quenching disappear to form higher-molecular-weight complexes if the sample is dialyzed against a low-pH buffer.

CP/RNA complexes contain multiple RNAs at neutral pH and VLP formation is highly cooperative. To better characterize these complexes -- avoiding the pH lowering associated with

gel electrophoresis -- we carried out a new set of titration curves (CP:RNA mass ratios of 0:1, 1:1, 2:1, 3:1, 4:1, 5:1 and 6:1) with each sample run in a 10-40% sucrose gradient. For simplicity, Fig. 3 shows only the samples at a CP:RNA mass ratio of 0:1, 1:1, 3:1 and 6:1. Fig. 3a shows that at neutral pH (1-step assembly) at a CP:RNA mass ratio = 3 the assembly of 500-nt RNA VLPs is a two-state process and the CP/RNA complexes contain several RNAs (hence multiple-RNA CP/RNA complexes). It also shows that at the magic ratio almost all of the RNA is in the multiple-RNA CP/RNA complexes. In agreement with Fig. 1 all the CP:RNA mass ratios greater between zero and six exhibit a two-state system (data not shown). We see from Fig. 3b that at CP:RNA mass ratios < 6, the CP preferentially binds to multiple-RNA CP/RNA complexes rather than to complexes containing a single RNA. This behavior is even more pronounced for the two-step assemblies (compare Figs. 3b and d). These gradients show that at neutral pH there is still a little RNA at the top of the gradient even at the magic ratio and it is only upon pH acidification (2-step assembly) that all of the RNA is packaged into VLPs. More precisely, in the 1-step assembly almost 10% of the RNA remains in the monomeric form even at the magic ratio, while only 1% remains in the top fractions for the 2-step assembly mix. Figs. 3a and c show that at a CP:RNA mass ratio = 1 most of the RNA remains as monomeric, however the multiple-RNA CP/RNA complexes are found at fraction 8, unlike the VLPs which are always observed in fractions 9 and 10. This is in agreement with Fig. 1, which shows that at this mass ratio the slow moving species do not have the same electrophoretic mobilities as a WT virion. These also suggest that the multiple-RNA CP/RNA complexes (1-step assembly) contain the same number of packaged RNA as the VLPs (2-step assembly).
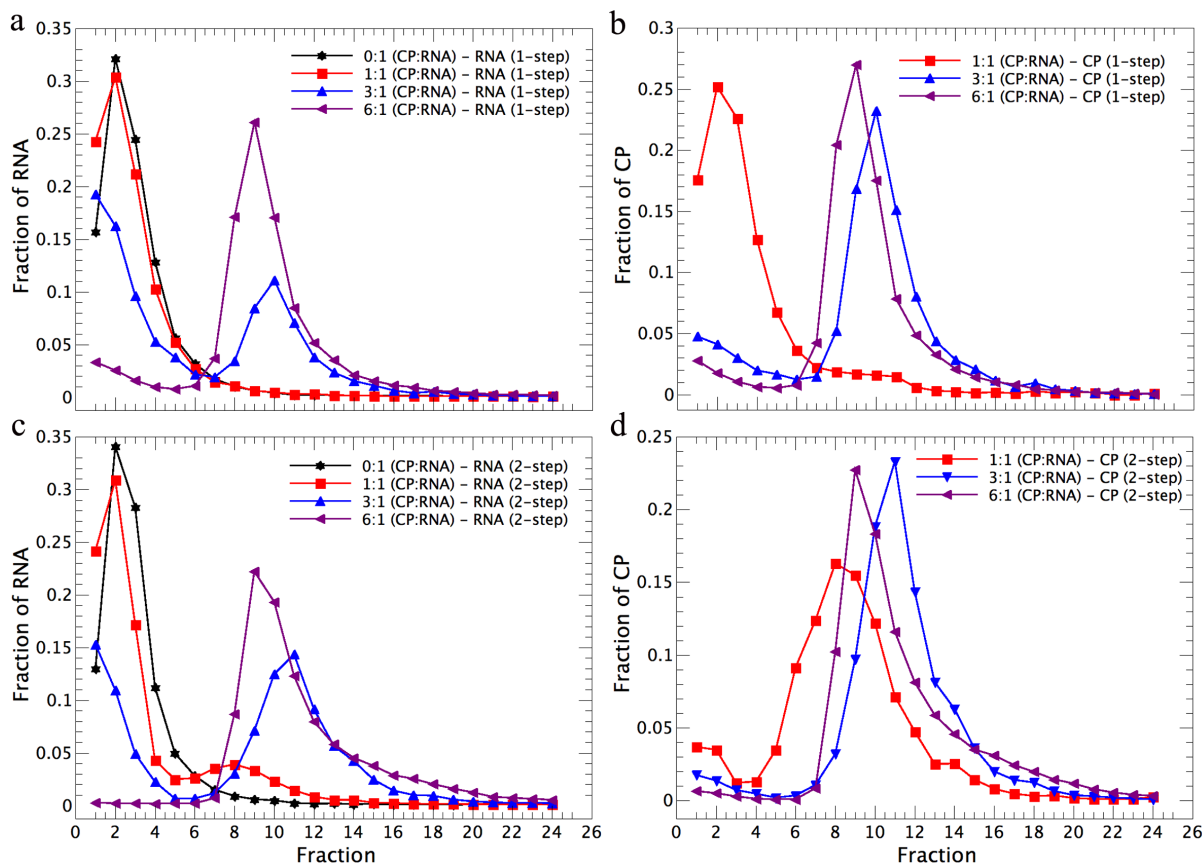
FIG 3. 500-nt RNAs form multiple-RNA complexes at neutral pH. Sedimentation profiles in a sucrose gradient of bi-labeled assemblies (AF647-CP and AF488-RNA) at CP:RNA mass ratios ranging from 0:1, 1:1, 3:1 and 6:1. a) and b) Profiles of the RNA (a) and CP (b) for the one-step assembly protocol. c) and d) Sedimentation profiles of the RNA (c) and CP (d) for the two-step assembly protocol. These sedimentation results show that, even at neutral pH, CP-RNA equilibrium involves only two species: single-RNA CP/RNA complexes and multi-RNA CP/RNA complexes.

To better understand the distribution of CP and RNA in solution we calculated the fraction of CP and RNA bound to the multiple-RNA complexes (fractions 7 − 16 from Fig. 3) as a function of the added CP. From Fig. 4 it is clear that the binding of CP and RNA to the multiple-RNA complex is highly cooperative in both (1-step and 2-step) assembly protocols and that only upon pH acidification can all of the RNA be packaged if the CP:RNA mass ratio is 6:1. To calculate the degree of cooperativity we fitted the data to a Hill equation for cooperative binding, $f = (1 + (K_D/[CP])^n)$ where f is the fraction of CP (blue and green curves) and RNA (red and black curves) bound in the multiple-RNA complexes, K_D is the dissociation constant, $[CP]$ the CP concentration and n the Hill coefficient. We note that while this particular model is

91

commonly used to describe binding of a small ligand to a substrate, it also applies to this series of cooperative binding events. At neutral pH (1-step assembly) the Hill coefficient for the equilibrium of the CP (RNA) is 3 (2.8) and the K_D is 2.06 μM (3.85 μM). This K_D for the CP is not associated with the equilibrium between free and bound CP but rather with the equilibrium between CP bound to a single RNA and to a multiple-RNA complex, while that for the RNA describes the equilibrium between the single and multiple RNAs. If, however, these data are fit assuming n-independent sites (a non-cooperative process) rather than cooperative binding events, then the K_D is larger than any of the experimental concentrations at which the assemblies were carried out (data not shown). This would imply that all of the CP/RNA complexes should be found as monomers rather than as multiple RNA complexes, which is obviously not the case. Also, a log-log plot of the ratio of fraction of bound CP to that of unbound CP, as a function of the CP concentration, shows a slope that is greater than one for each data set in Fig. 4: moreover it is exactly the same as the Hill coefficient in Fig. 4 (data not shown). These two facts reinforce the idea that this is a cooperative equilibrium rather than the result of n-independent binding events.

The K_D's for the RNA (2.58 μM) and CP (0.50 μM) after the pH was lowered are smaller than at neutral pH (3.85 and 2.06 μM respectively), in agreement with the fact that CP-CP attractive interactions are stronger at low pH. Note further that at neutral pH the cooperative binding equilibrium between the monomeric and multimeric RNAs, as well as the bound CP to each RNA species, suggests that there are at least three binding events and that the number of packaged RNAs is four. The difference in the equilibrium constants of the RNA and the CP suggests that the binding of several RNAs is driven by the amount of CP bound to them. To test

this idea we plot the ratio of the fraction of CP bound to the multiple-RNA complexes to the fraction of RNA bound to the multiple-RNA complexes, for each of the assembly protocols (see Fig. 5). This plot shows that there is an asymmetry in the distribution of the total amount of CP and RNA in the multiple-RNA complexes; at low mass ratios most of the CP in solution is bound to the complexes while most of the RNA remains as single molecules. The ratio of CP to RNA in the multiple-RNA complexes is seen to decrease even as the CP:RNA mass ratio increases (and therefore also as the total amount of CP increases), indicating that the amount of RNA bound in the multiple-RNA complexes must be increasing.
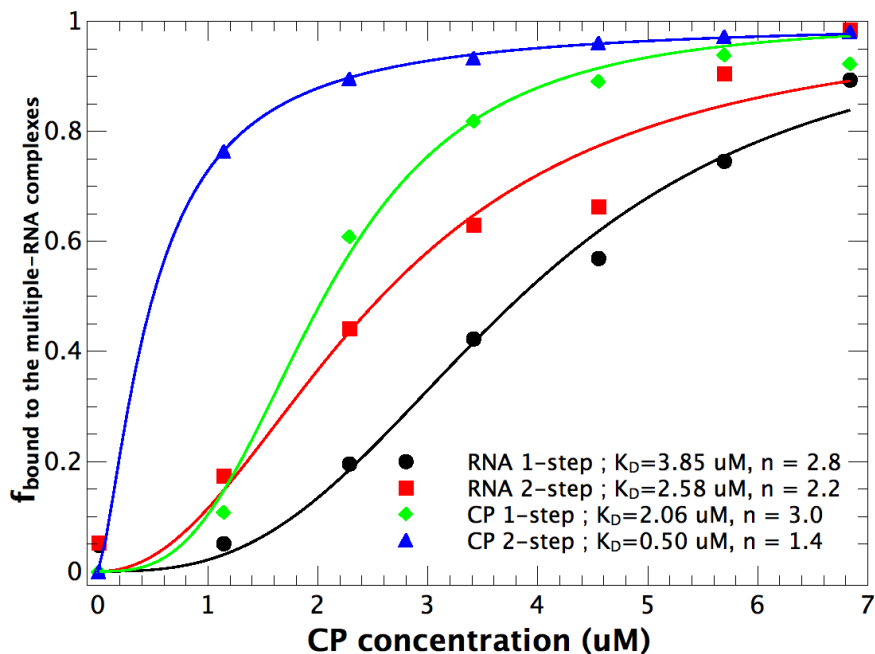
FIG 4. Equilibrium of CP between single-RNA and multiple-RNA complexes is more cooperative at neutral pH. By plotting the data from Fig. 3 to calculate the fraction of CP (or RNA) bound in the multiple-RNA complexes as a function of the total CP concentration. These fractions are then fitter to the Hill equation, to obtain (see the text) the dissociation constant (K_D) and the Hill coefficient (n) for the equilibrium of the CP (or RNA) between the single- and multiple-RNA complexes. The results are consistent with the fact that CP-CP interactions are stronger at acidic pH (VSB) than at neutral pH (RAB). The Hill coefficient indicates not only the degree of cooperativity but also the minimum number of subunits of the substrate; in this case $n \approx 3$ for the 1-step assembly and the number of RNAs in a VLP for the 2-step assembly (and in a multiple-RNA CP-RNA complex at 150 nM RNA) is 4. The solid lines show the Hill equation fits.
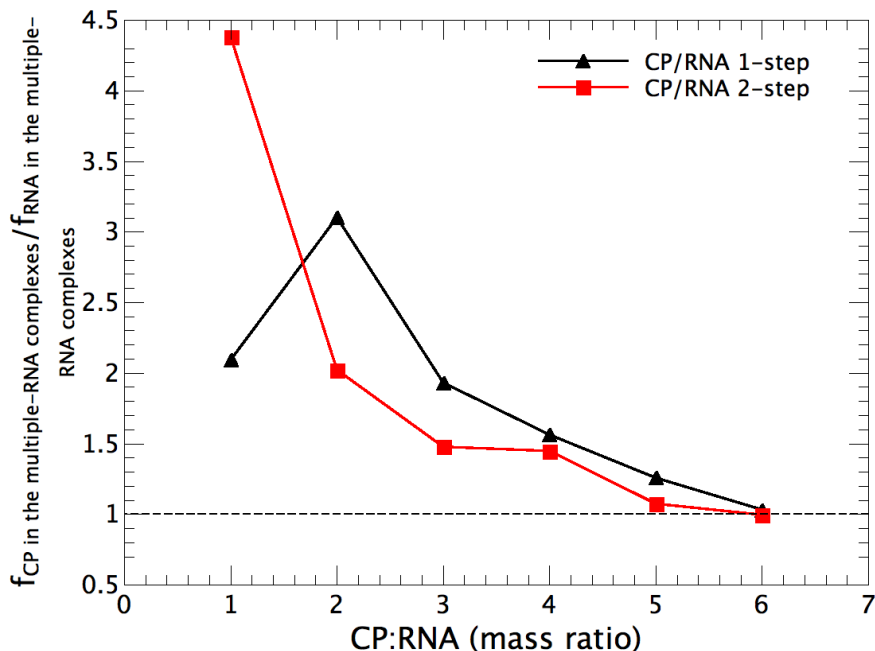
FIG 5. The formation of multiple-RNA complexes is driven by CP. Shown is the ratio of the fraction of CP bound in the multiple-RNA complexes to the fraction of RNA bound in the multiple-RNA complexes, at acidic (2-step) and neutral (1-step) pH. The only way this ratio can decrease, as the amount of added CP increases, is if the amount of complexed RNA increases, demonstrating that it is the binding of the CP to the RNA that promotes the interaction of several CP/RNA complexes.

The size of the capsid does not depend on the absolute RNA concentration. To investigate if the final size of the capsid depends on the initial absolute concentration of 500-nt RNA, we carried out assemblies at the magic ratio by the two-step dialysis pathway at RNA concentrations of 50, 150 and 500 nM. The distributions in Fig. 6 show that there are no significant differences between the VLPs formed at these three concentrations.

sm-FCS confirms that acidification is required for VLP formation. The velocity sedimentation experiments, EMA and EM required high RNA absolute concentrations (150 nM RNA). Accordingly, we used sm-FCS to estimate the size of the CP/RNA complexes at single-molecule concentrations (10 nM RNA) that might be more appropriate for comparison to the *in*

vivo scenario. As discussed in the METHODS AND MATERIALS section the autocorrelation function, $G(\tau)$, the size of a particle (hydrodynamic radius [R_h]), the number of particles in the confocal volume, and the photophysics of the fluorophores can be extracted from the sm-FCS data. Since at any given time there are only a few (~1-5) molecules in the confocal volume (5-10 nM RNA with an RNA DOL ≈ 0.5) any artifacts due to aggregation are suppressed. Moreover, by labeling either the RNA or the CP with a fluorophore, the size of each can be selectively measured in the presence of the other.
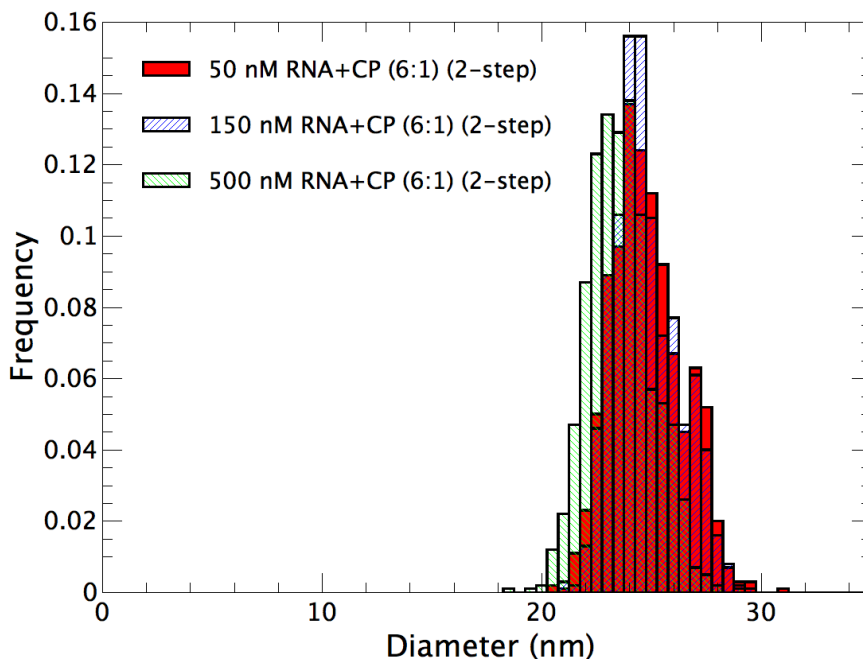
FIG 6. The size of the capsids is independent of the absolute concentration of the RNA. 500-nt RNA assemblies at three RNA concentrations at the magic ratio were carried out by the 2-step protocol and then analyzed by negative-stain EM. The absolute RNA concentration is seen to have no effect on the size of the capsid.

Fig. 7a shows the average R_h of AF488-RNA as a function of the amount of CP added to a fixed amount of RNA. To understand the role of pH we followed two different assembly protocols, both performed at 150 nM RNA and then diluted 15-fold. Fig. 7a shows that the R_h of AF488-RNA increases as the CP:RNA mass ratio increases; at neutral pH the R_h goes from ≈ 2.5

to ≈ 6 nm (the expected R_h for a "pseudo T = 2" capsid is 11.5 nm) while the two-step dialysis sample yields capsids with an R_h ≈ 11.0 nm at a mass ratio of 3 and above. The 2-step data are in agreement with the gels shown in Fig. 1, which show that even at such low ratios most of the RNA has formed VLPs. The titration curve in Fig. 1 shows that, for the 2-step assemblies there are only two species at any given mass ratio smaller than 6. We therefore tried to analyze the sm-FCS data for both assembly protocols by assuming that there were either one or two species in solution.
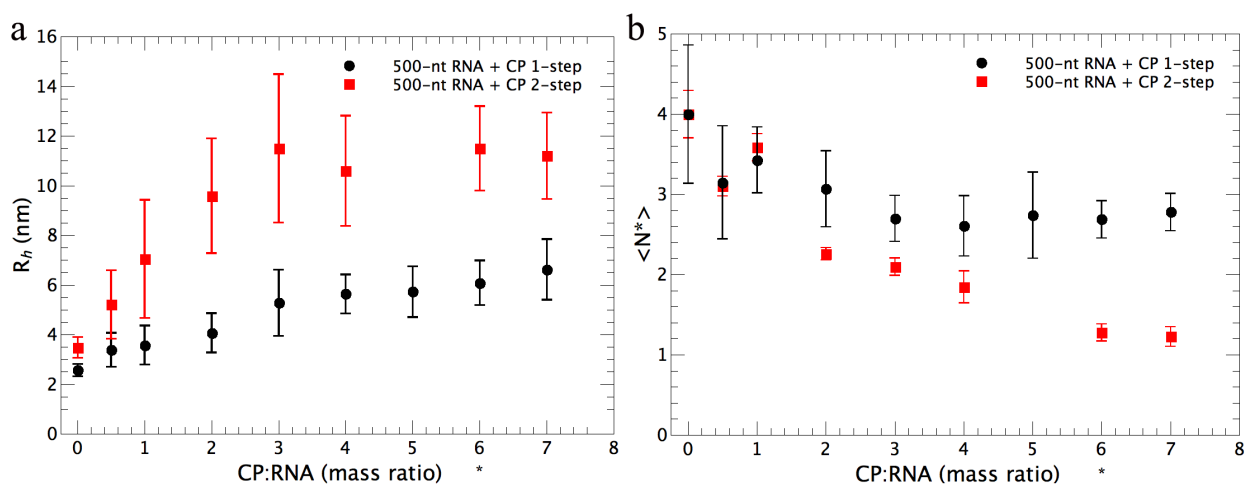
FIG 7. AF488-labeled 500-nt RNA hydrodynamic radii (R_h) and average number of particles <N*> as a function of added CP. a) The R_h was obtained by fitting the sm-FCS auto-correlation curve for AF488-labeled RNAs. Each data point consists of 40 measurements and the error bars represent the standard deviation of the mean. The black circles and red squares indicate, respectively, samples prepared by a 1- and 2-step protocol. The small increase in R_h at neutral pH demonstrates that CP/RNA complexes only associate weakly at neutral pH (1-step) and that acidification (2-step) is required to produce VLPs. The star on the x-axis indicates the magic ratio. The dotted line is the R_h associated with the 500-nt VLPs that was used to calibrate the confocal volume. b) The normalized average number, <N*>, of particles of RNA decreases by a factor of four when VLPs are formed at the magic ratio (star) after lowering the pH. In other words, four 500-nt RNA are packaged into a capsid.

The choice of the fitting assumptions (one or two species), however, did not lead to significant changes in the R_h distributions (data not shown). We were also unable to show, by either keeping the R_h of one species constant while fitting the R_h and the mole fraction of the other species, or by fitting both R_h while keeping the mole fraction of the other species constant, that samples at a mass ratio > 3 had a bimodal distribution around 2.5 and 11.5, nm as one would

expect from the EMA. Meseth and co-workers have shown that, in order to measure a bimodal distribution in a two-component system by sm-FCS, the ratio of the R_h for both species has to be greater than 1.6 and their mole fractions and quantum yields have to be comparable (19). From Figs. 1-3 we know that at a CP:RNA mass ratio = 3 the mole fractions of single- and multiple-RNA complexes in the 2-step assembly mixes are comparable, and from Fig. 7a we see at least a 3-fold difference in R_h's for the single- and multiple RNA/CP complexes. However, there is a 5-fold decrease in the quantum yield of fluorescently labeled RNA upon encapsidation (data not shown). This difference in quantum yields is most likely the reason why we are unable to show that there are two species in solution for CP:RNA mass ratio > 2 and at low pH.

The average number of particles in the confocal volume, <N>, can be calculated – see Eq. (2) – by extrapolating the autocorrelation function to $\tau = 0$. Fig. 7b shows the normalized average number of particles, <N*>, as a function of added CP. <N*> was calculated by normalizing <N> for samples at a non-zero CP:RNA mass ratio by their <N> value ($<N>_0$) at CP:RNA = 0:1 (for each of the two dialysis pathways), so that at a CP:RNA = 0:1 <N*> = 4 (= $4<N>/<N>_0$). This figure shows that when the mass ratio increases from 0:1 to 6:1 (magic ratio) for the 1-step assemblies, <N*> decreases from ≈ 4 to ≈ 3, whereas for the two-step assemblies it decreases to ≈ 1. This result indicates that at the magic ratio and low pH four RNAs are packaged per VLP. Only the <N*> at CP:RNA = 0:1, at both pH's, and at 6:1 at low pH, can be fitted by a one-diffusing-species model; any other point on both titration curves (i.e., for 1- and 2-step assemblies) involves two species in solution (see Fig. 3). While we were not able to show in the distribution of R_h's that there are two diffusing species, it was straightforward to obtain a <N*> at low pH for both species. As mentioned above the quantum yield of the VLPs is five-fold lower than that of pure labeled-RNA; by taking this into account in Eq. 7 and extrapolating to $\tau = 0$ we

were able to show that 45% (55%) of the RNA remains in single-(multiple-) RNA complexes. At neutral pH and at any mass ratio ≥ 3, $\langle N^* \rangle$ remains constant. If we assume that both species have the same quantum yield and that the multiple-RNA CP/RNA complexes contain two RNAs (based on their R_h's) we infer 65% of the RNAs as 2 RNA/CP complexes. In the other limiting case, when the quantum yield of the multiple-RNA complexes is five-fold lower than that of single-RNA CP/RNA complexes, then only 40% of the RNAs are present as 2-RNA/CP complexes.

Taken together these results demonstrate that in order to form VLPs the pH has to be lowered and that indeed four 500-nt RNAs are packaged into a VLP. They also suggest that at neutral pH and at sm-FCS concentrations (here, 10 nM RNA) CP/RNA complexes with an $R_h \approx$ 6 nm contain two RNAs. Note that even though at low pH the R_h does not change after a mass ratio of 3 (see Fig. 7a), $\langle N^* \rangle$ reaches its minimum only at the magic ratio. This is possible only if at a CP:RNA ≤ 5 there is still unpackaged RNA, which is in agreement with Figs. 1-3. It also confirms that a CP:RNA mass ratio of 6:1 is the minimum mass ratio required to package all of the RNA in solution.

CP and RNA at the magic ratio and at neutral pH form amorphous aggregates. The structure of the multiple-RNA CP/RNA complexes at neutral pH was determined by negative-stain and cryo-electron microscopy. Figs. 8-10 show representative electron-microscope images of single RNAs, CP/RNA complexes at the magic ratio, and VLPs for the one-step and two-step assembly pathways. Fig. 8a shows cryo-EM images of a 500-nt RNA in RAB, exhibiting low electron density and irregular shapes (see Fig. 8a). Given that long ssRNA does not have a well-defined shape (20) we measured its size by determining the geometric mean of a rectangle that

contains all of the electron density of the RNA; the average size ("diameter") of the RNA –

determined in this way – was 21 nm. At neutral pH the multiple-RNA CP/RNA complexes have

a greater electron density than pure RNA and their overall size differs little from that of a capsid

(~ 22 nm). They are not spherical, however, and their electron density is not homogeneous (Fig.

8b). It is only after acidification that spherical capsids are formed (Fig. 8c). Although the

overwhelming majority of capsids are spherical we find a small fraction of them that have small

defects, i.e., protuberances (see Fig. 9). The meaning of these defects will be explained in detail

in the next section.
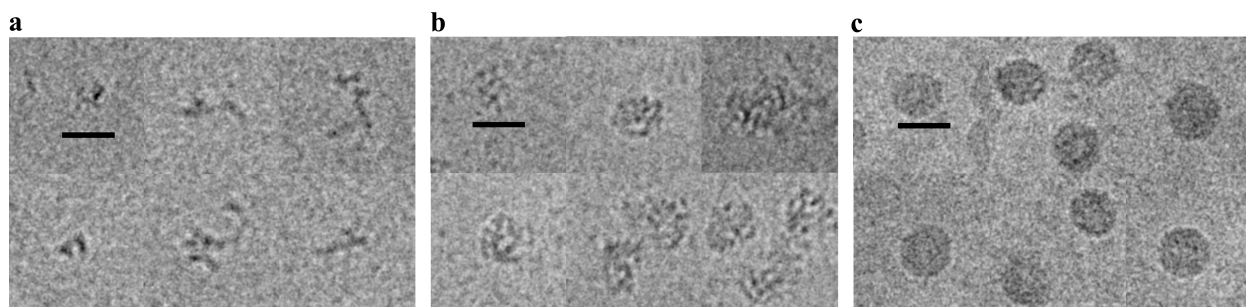
FIG 8. Cryo-electron micrographs of the assemblies at neutral pH and after acidification. Panels a-c show cryo-EM images of CP/500nt-RNA complexes. a) CP:RNA = 0:1. b) At a CP:RNA mass ratio of 6:1, neutral pH CP/RNA complexes are amorphous and their electron density is low compared to that of properly formed VLPs; however, their size and shape are close to those of a capsid. c) It is only upon pH acidification of the 6:1 solution of complexes that proper capsids (VLPs) are formed. The scale bar is 25 nm.
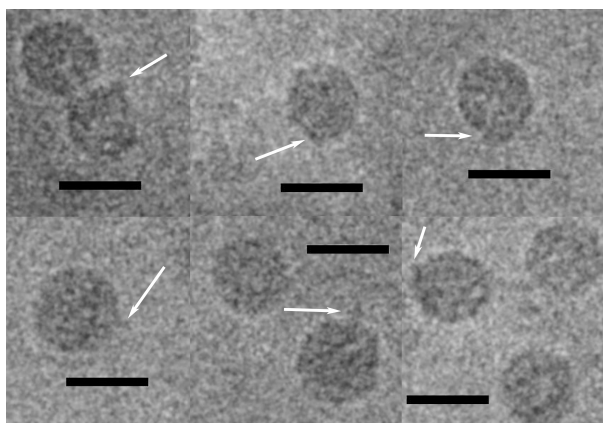

FIG 9. Cryo-electron micrographs of the assemblies after acidification. A small fraction of the capsids obtained by the 2-step assembly protocol and at a CP:RNA mass ratio of 6:1 show small defects (see white arrows). These defects are associated with the excess of CP in solution that is needed to package 100% of the RNA in solution. The scale bar is 25 nm.

We also analyzed these samples by negative-stain EM (Fig. 10). This technique is not as reliable as cryo-EM since it can induce artifacts due to the sample adsorption to the grid, inherent pH lowering, and drying. However, most of the CP-RNA structures found at neutral pH are amorphous; we observed only a very low frequency of capsids (Fig. 10a). In contrast, after the two-step dialysis, negative-stain EM shows (in agreement with the cryo-EM images, see Fig. 8c) that almost all of the structures are capsids, with only a few misassembled ones (Fig. 10b).
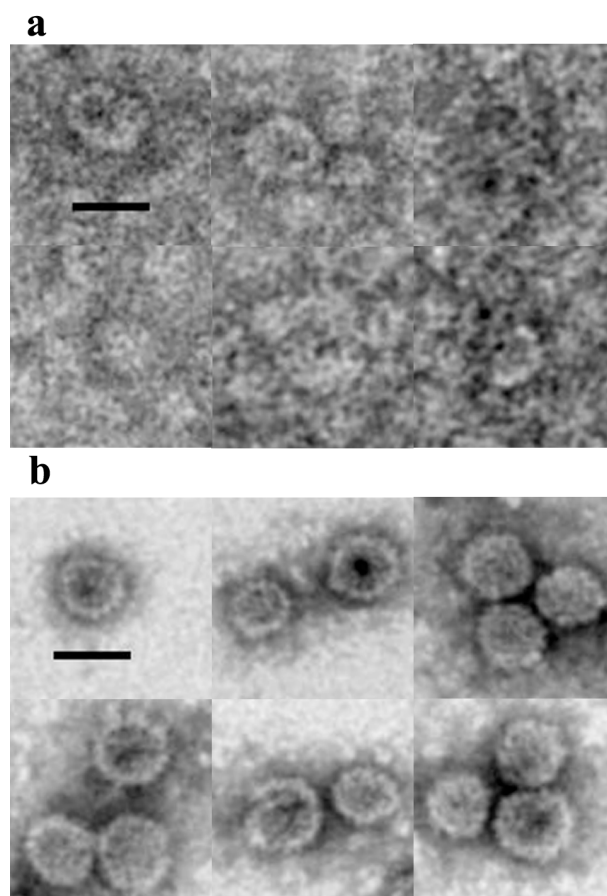
FIG 10. Negative-stain electron micrographs of the assemblies at neutral pH and after acidification. Panels a-b show negative-stain EM images of 6:1 CP:RNA assembly mixes. a) Even though staining with uranyl acetate involves a lowering of the pH, well-defined capsids are rare. b) In agreement with cryo-EM, capsids *are* formed if the samples are dialyzed against a low-pH buffer before being stained. The scale bar is 25 nm.

DISCUSSION

To characterize the assembly of CCMV capsid protein around a short ssRNA we have shown that: (i) at neutral pH and at the magic ratio, the number of RNAs in the multiple-RNA CP/RNA complexes depends on the absolute concentration of RNA -- 2 at 10 nM and 4 at 150 nM; (ii) four RNAs are encapsidated when VLPs are assembled in a two-step dialysis and at the magic ratio; (iii) the size of the capsid is independent of the absolute RNA concentration; (iv) at neutral pH, 150 nM RNA, and at CP:RNA mass ratios lower than the magic ratio, there are two species in solution (single-RNA and multiple-RNA CP/RNA complexes); (v) the fraction of CP bound to the single-RNA CP/RNA complexes depends on the CP:RNA mass ratio, with this fraction always small compared to the fraction bound to the multiple-RNA complexes; and (vi) at neutral pH the multiple-RNA CP/RNA complexes are amorphous even though their size is similar to that of the final VLPs (at pH 4.5).

Nucleocapsid assembly for ssRNA viruses can been classified as either cooperative or non-cooperative depending on whether at some CP concentration, lower than the minimum needed to package all of the RNA in solution, there are predominantly two (cooperative) or more species (non-cooperative). Earlier work has suggested that cooperativity during capsid assembly is determined by the strength of the CP-CP interactions (12,18,21). Here we show that a CP that exhibits non-cooperative assembly around "medium to large" RNAs (7,12) can assemble into capsids with high cooperativity around small RNAs. It has been proposed recently that the degree of cooperativity of nucleocapsid assembly is a function of the CP-CP interactions and of the "*work*" required to package nucleic acid (18), suggesting that assembly for a particular CP could be either cooperative or non-cooperative depending on the work needed to package the RNA. The precise meaning of the work term is not clear because by definition any spontaneous

process requires no work, and thus work as such plays no role in these self-assembly processes. Our results show that the degree of cooperativity depends not only on the relative strength of the CP-CP and CP-RNA interactions, but also on the *length* and number of packaged RNAs.

For our assemblies of CCMV CP and a 500 nt-RNA, both the EMAs and the sucrose gradients show that at CP:RNA mass ratios less than 6:1 there are two main species: single-RNA CP/RNA complexes (with a small amount of bound CP) and multiple-RNA complexes that have an excess of bound CP. By an excess of CP we mean that there is more CP bound than the 120 required to form a T = 2 capsid. This super-stoichiometric amount of bound CP has been shown to be essential for packaging all of the RNA in solution (7). We also have strong evidence that at neutral pH all of the CP is bound to the RNA and that there is no free CP in solution (Garmann, R.F. 2013 personal communication). The sucrose gradients show that even in the 2-step assemblies the excess CP is bound to the VLP (see Fig. 3); however the cryo- and negative-stain EM images (Fig. 8 and 9) show that these VLPs have the expected morphology of a capsid. Previous work on the self-assembly of CCMV CP in the absence of negatively charged cargo (22-24) has shown that CCMV CP spontaneously form *multi*-shells at low pH and medium ionic strength. This means that, during the pH-lowering step, the excess CP bound to the CP/RNA complexes first unbinds, allowing capsid formation, and then binds again to the exterior surface of the VLP. Although we do not find any multishells, extensive and careful analysis of the EM images show small patches (or protuberances) which could be extra hexamers or pentamers bound to the exterior surface of the VLP (see white arrows in Fig. 9).

The sucrose gradient, EMA and sm-FCS data for pure RNAs demonstrate that in the absence of CP the RNA does not form dimers or multimers; therefore, the association of two or

102

more RNAs must be promoted by the binding of CP. Since there are no specific RNA-RNA intermolecular interactions, then the formation of multiple-RNA complexes is due to association of CP/RNA complexes, as opposed to binding of multiple CP-free RNAs. This association is likely driven by weak lateral CP-CP interactions. However, this does not explain why at CP:RNA mass ratio less than 6:1 there are in addition to multiple-RNA CP/RNA complexes, single 500nt-RNAs with a small amount of bound CP and no intermediate species (2- or 3-RNA/CP complexes). This is different than the case of RNAs with lengths exceeding > 1000 nt, where at a mass ratio less than 6:1 there is -- in addition to capsids -- a distribution of different amounts of CP bound to RNA (7,12).

We propose that the difference in the distribution of the fraction of bound CP to both species (single- and multiple-RNA complexes) (see Figs. 4 and 5) is key to understanding the assembly mechanism when two or more RNA molecules are packaged, and that this is what gives rise to a two-state system. In fact, it is easier to understand how cooperative behavior arises from the length of the packaged RNA by comparing the results presented here with those for long RNAs (7,12). When CP and RNA are mixed at neutral pH, the average number of CPs bound per RNA is equal to the CP:RNA mole ratio in solution. Note that this does not mean that all RNAs have the same number of bound CPs, but rather that there is a distribution of bound CPs that peaks at the CP:RNA mole ratio at which the assembly experiment is carried out. If we compare the case for a 3.2 and a 0.5 kb RNAs at half the magic ratio, then there are on average 160 and 25 CPs per RNA, respectively. On the one hand, forming a T = 3 capsid requires 180 CPs, therefore a 3.2 kb RNA depends only on small CP fluctuations (~ 20 CP) to have at least 180 CPs. From this it follows that some fraction of RNAs will (have 180 CPs and) form VLPs,

while the rest will remain as CP/RNA complexes, all of which is in complete agreement with the experimental results. On the other hand, since the 500-nt RNAs form T = 2 capsids and the maximum number of CP binding sites per RNA (50) is less than 120, then the formation of multiple-RNA CP/RNA complexes is required to build a capsid. At half of the magic ratio these multiple-RNA complexes have on average 100 CPs, which is less than the amount required to form a T = 2 capsids. This suggests that in order to form complexes with the right number of CPs there has to be an equilibrium between the CP on the single- and multiple-RNA complexes. Moreover, when a CP is in a multiple-RNA complex it can form more CP-CP interactions than when in a single-RNA CP/RNA complex; hence the multiple-RNA complexes are lower free energy compared to that of single-RNA complexes.

A fundamental point is when comparing both assembly cases (3.2 vs 0.5 kb RNA) is that the change in the CP chemical potential, when going from a long RNA with 160 to one with almost 180 CPs, is smaller compared to that of a CP going from a short RNA with 25 CPs to a multiple-RNA complex with almost 120 CPs. In fact, given the small size of this experimental system (500 nt), finite system effects might play an important role. As the equilibrium of the CP favors binding to the multiple-RNA ones, the free energy of the small complex increases while that of the multiple-RNA complexes decreases. This decrease in the free energy is mainly due to the fact that, although weak compared to those at acidic pH (22-24), CCMV CP-CP lateral interactions are present at neutral pH, Moreover, these interactions depend also on the angle between two adjacent CP's. While for the large RNAs the curvature of CP/RNA complexes with 150 or 180 CP are expected to be similar (if not the same), this is not so for the single- and multiple-RNA case. Note that when the CP:RNA mass ratio approaches 6 the CP binding density of the single-RNA complexes increases, hence increasing the yield of capsid formation.

As mentioned before building an excess of CP in the multiple-RNA complexes is what drives the assembly. In a perfect positive cooperative model the first binding event has the lowest affinity, and therefore large changes in ligand concentration are needed to promote ligand binding. In other words, going from the low to the high affinity state requires a relatively large change in ligand concentration, while successive high affinity binding events require very small ligand concentration changes. By fitting the binding curves with a positive cooperative binding model we extracted a Hill coefficient of 3 for the equilibrium between single- and multiple-RNA complexes at neutral pH. Given that 4 RNAs are packaged and that the Hill coefficient is 3, it is likely that the dimerization of CP/RNA complexes is the low-affinity event, while the subsequent formation of 3- and 4-RNA/CP complexes are higher-affinity events. This model is consistent with the sm-FCS data at neutral pH, which show that even at the magic ratio -- but at a CP concentration lower than the K_D for either the CP or RNA -- the CP/RNA complexes are smaller in size than a VLP and have fewer than four RNAs. As shown in Fig. 7b <N*> decreases from 4 to 3 when the CP:RNA mass ratio goes from 0:1 to 6:1, suggesting that at this concentration and mass ratio the interaction of four CP/RNA complexes leads to an equilibrium between monomeric and dimeric CP/RNA complex rather than the formation of a VLP containing four RNAs. This is in agreement with the idea that the CP/RNA dimers have a low binding affinity at neutral pH.

A third important aspect of this assembly system is the final size of the VLPs. Our data show that even a ten-fold increase in the absolute concentration of RNA (while keeping the CP:RNA mass ratio constant) has no effect on the size of the capsid, demonstrating that the formation of the multiple-RNA complexes is not a simple aggregation process. First, it is well

105

known that CCMV CP has a preferred curvature involving a 14 nm radius (corresponding to a T = 3 capsid) (24). However we find that the fraction of capsids with a diameter consistent with T = 3 is small compared to that with a "pseudo T = 2" size, implying that the final size of the capsid depends on how the RNA affects the relative orientation of the bound CPs. In other words, the intrinsic curvature of the RNA (which acts as a scaffold) affects how CPs interact with each other, hence establishing a preferred curvature different from that of the CPs when assembled around long RNAs or in the absence of cargo at low pH and high ionic strength. Second, if the final state of these complexes were a result of the degree of aggregation then one would expect that their size will depend on the absolute concentration of RNA, while if the final size of the multiple-RNA CP/RNA complexes is controlled by the preferred curvature of the CP then the final size of a capsid would be independent of the absolute concentration.

Finally, our sm-FCS and Cryo-EM data are in agreement with previous results (7, 8, Garmann et al. 2013 personal communication) showing that at neutral pH CCMV CP lateral interactions are so weak that they cannot nucleate the formation of a spherical shell; nucleation of capsids requires a lowering of the pH. The need for at least a two-stage dialysis pathway reinforces the conclusion that successful capsid assembly requires tuning of the CP-CP and CP-RNA interactions to minimize kinetic traps.

CONCLUSIONS

By combining all these results we are proposing an assembly of VLPs around for short ssRNAs that highlights the role of the nucleic acid as a scaffold and an active component in capsid assembly; (i) when CP and RNA are mixed at neutral pH we find no free (unbound) CP; (ii) at

neutral pH the formation of the higher-order complexes is driven by CP; (iii) the difference in the CP chemical potential between the single- and multiple-RNA complexes results in a two-state system at CP:RNA mass ratios less is than 6:1, in which the single-RNA species have almost no bound CP; (iv) the formation of dimers of CP/RNA complexes is the critical step since it is the lowest affinity-binding step, while the formation of 3- and 4-RNA/CP complexes are highly cooperative; and (v) the multiple-RNA CP/RNA complexes are not yet capsids unless the CP-CP interactions are increased by pH lowering.

REFERENCES

1. Venter, P. A., N. K. Krishna, 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(10):6239-6248.

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

3. 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.

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

5. Kobayashi A., and Y. Ehara. 1995. In vitro encapsidation of cucumber mosaic virus RNA species. Ann. Phytopathol. Soc. Jpn. 61:99 –102.

6. Ford, R.J., A.M. Barker, S.E. Bakker, R.H. Coutts, N.A. Ranson, S.E.V. Phillips, A.R. Person, and P.G. Stockley. 2013. Sequence-specific, RNA-protein interactions overcome

electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. J. Mol. Biol. 425:1050-1064.

7. Cadena-Nava, R. D., M. Comas-Garcia, R.F. Garmann, A. L. N. Rao, C. M. Knobler, W. M. and 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(6):3318-3326

8. Comas-Garcia, M., R. D. Cadena-Nava, A. L. N. Rao, C. M. Knobler, W. M. and Gelbart. 2012. In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. J. Virol. 86(22):12271-12282

9. Rao, A. L. N. 2006. Genome packaging by spherical plant RNA viruses. Annu. Rev. Phytopathol. 44:61-87

10. Annamalai P., and A. L. N. Rao. 2006. Packaging of brome mosaic virus sub- genomic RNA is functionally coupled to replication-dependent transcription and translation of coat protein. J. Virol. 80:10096-10108.

11. Vento, P. A., and A. Schneemann. 2007. Assembly of two independent populations of flock house virus particles with distinct RNA packaging characteristics in the same cell. J. Virol. 81(2):613-619.

12. 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.

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

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

15. Bancroft J. B., G. J. Hills, and R. Markham. 1967. A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro. Virology 31:354 –379.

16. Gillitzer, E., D. Willits, M. J. Young, T. and Douglas. 2002. Chemical modification of a viral cage for multivalent presentation. Chem. Commun. 20:2390-2391

17. Tsay, J. M., Doose, S., and Weiss, S. 2006. Rotational and translational diffusion of peptide-coated CdSe/CdSZnS nanorods studied by fluorescence correlation spectroscopy. J. Am. Chem. Soc. 128(5):1639-1647.

18. Zlotnick, A., J. Z. Porterfield, and C-Y. Wang. 2013. To build a virus on a nucleic acid substrate. Biophys. J. 104:1595-1604.

19. Meseth, U., T. Wohland, R. Rigler, and H. Vogel. 1999. Resolution of fluorescence correlation measurements. Biophys. J. 76(3):1619-1631.

20. Gopal A., H. Z. Zhou, C. M. Knobler, and W. M. Gelbart. 2012. Visualizing large RNA molecules in solution. RNA 18:284 –299.

21. 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.

22. Adolph K. W., and P. J. G. Butler. 1974. Studies on the assembly of a spherical plant virus. I. States of aggregation of the isolated protein. J. Mol. Biol. 88:327–338.

23. Lavelle L, R. D. Cadena-Nava, M. Gingery, J. R. Vega-Acosta, M. Phillips, L.A. Pinedo-Torres, J. Ruiz-Garcia, W. M. Gelbart, and C. M. Knobler. 2009. Phase diagram of self-

assembled viral capsid protein polymorphs. J. Phys. Chem. B 113:3813–3820.

24. Prinsen, P., P. van der Schoot, W. M. Gelbart and C. M. Knobler. 2010. Multishell structures of virus coat proteins. J. Phys. Chem. B. 114:5522-5533.

25. Carrillo-Tripp M., C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks, III and V. S. Reddy. 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nuc. Acid. Research. 37, D436-D442.

Chapter V

Summary

This thesis is concerned with the *in vitro* self-assembly of a single-stranded RNA (ss-RNA) virus, cowpea chlorotic mottle virus (CCMV), and the effects of the length of the nucleic acid on capsid assembly. CCMV is a positive-sense ssRNA plant virus, which encodes only for three non-structural proteins (RNA-dependent RNA polymerase [proteins 1a and 2a] and a movement protein) and a structural protein (capsid protein [CP]). Single-stranded RNA viral assembly is a highly concerted process between CP and RNA, in which the latter is not a passive cargo but rather acts as a scaffold and a catalyst. Although ensuring packaging selectivity during *in vivo* assembly might involve specific mechanisms such as compartmentalization (viral factories), RNA-dependent RNA polymerase-CP interactions and/or RNA packaging signals, non-specific interactions are also responsible for some of the fundamental features that control the assembly. These non-specific interactions depend mostly on length of the RNA and its three-dimensional conformation and could explain why some viruses have evolved to segment their genome into several molecules (multipartite genomes) that are not necessarily all packaged into the same capsid.

In Chapter 2 we explore the effects of the length of RNA on virion assembly and genome packaging. We show that the size of a capsid depends not only on the preferred curvature of the CP, but also on the length and number of packaged RNA molecules. Under non-physiological conditions and in the absence of RNA, CCMV CP self-assembles into empty capsids with a 28

111

nm diameter, indicating that this capsid size corresponds to the spontaneous curvature of the CP. However, the size of the capsid can deviate from the "*optimal*" one (T = 3) if the length of the packaged RNA differs significantly from that of the wild-type (WT) RNA content (3.2 kb). Viral RNAs have evolved to be encapsidated by their cognate CP, suggesting that specific RNA-CP interactions greatly affect viral assembly. To avoid artifacts due to specific interactions we used the RNA1 (3.2 kb) from the closely related brome mosaic virus (BMV) rather than CCMV RNA1 (3.1 kb [C1]). [The lengths of the four RNA molecules of the bromoviruses are about 3.2 kb (RNA1), 2.8 kb (RNA2), 2.1 kb (RNA3), and 0.8 kb (RNA4).] We found three regimes: (i) for short RNAs ($\leq$ 1. kb) multiple copies are packaged per capsid such that the RNA content (relative to protein) of the virus-like-particles (VLPs) is almost identical to that of the WT ($\approx$ 2.0 kb in "pseudo T = 2" and $\approx$ 3.0 kb in T = 3 capsids); (ii) for medium length RNAs (between 1.5 and 5.0 kb) there is only one RNA molecule per capsid, hence the amount of packaged RNA depends on the length of the RNA molecule; and (iii) for long RNAs (> 5.0 kb) a single RNA molecule is shared by multiple capsids (multiplets).

Short RNAs, except for the 0.5 kb, are packaged into a mixed populations of "pseudo T = 2" and T = 3 capsids; however not even the shortest RNA packaged (140 nt) was able to form T = 1 capsids. In a previous study it was shown that a 38-kDa fluorescently-labeled poly(styrene sulfonate) (PSS) molecule (which has a radius of gyration comparable to that of a 140-nt long RNA) was preferentially packaged into a T = 1 capsid. This may be due to the fact that RNA is a branched polymer while PSS is a linear one, although the exact reason for this difference is not clear. For long RNAs there are only two preferred capsid sizes, T = 3 and 4, although in these cases the analysis is more complicated because it involves multiple capsids per RNA molecule. From all these experiments it is clear that even though CCMV CP has some "flexibility" to make

capsids that deviate from the size of the WT there is a limit, set by the how much the curvature of the capsid can be affected by the size of the RNA.

The *in vitro* packaging of non-viral RNAs by CCMV CP has long been known, nonetheless this is the first work in which it is shown that independent of RNA length and sequence CCMV CP is able to package 100% of the RNA in solution. By carrying out a series of titration curves of CP with RNAs of varying lengths we find that this high encapsidation yield (not efficiency) is possible if and only if the assembly reaction is carried out under very specific experimental conditions, namely at a CP:RNA mass-ratio of 6:1. As long as this criterion is met, at the right pH and ionic strength, any RNA can be fully packaged. We term this experimental condition the "*magic ratio*" and it turns out that it corresponds to equal amounts of negative charges from the RNA phosphate backbone and positive charges from the CP N-termini are present in solution. Below this value, the amount of packaged RNA decreases as the mass ratio decreases. This result demonstrates that in order to package all of the RNA in solution a superstoichiometric amount of CP (more than the one needed to make a capsid) has to bind to the RNA, leading to an assembly scenario that is in agreement with the "*en masse*" pathway proposed independently by Michael Hagan; more explicitly, the CPs saturate the RNA phosphates by non-specific binding, and then lateral CP-CP interactions promote capsid formation. This limit-case scenario is valid when the CP-RNA interactions are stronger than those between CPs.

In Chapter 3 we show that in order to measure relative packaging efficiencies in head-to-head competition experiments between two different RNA molecules, the amount of CP present

must be significantly less than that required to packaged both of them. Specifically, in order to measure relative packaging efficiencies we constrained the system by: (i) normalizing the packaging efficiency of a competitor RNA relative to that of a reference RNA with the same length as the viral one (BMV RNA 1, 3.2 kb [B1]); (ii) adding equal masses (therefore equal number of phosphates) of the competitor and the reference RNA; and (iii) having a CP:total-RNA RNA (competitor + reference) mass ratio = 3:1, ensuring that only one of the RNA species can be fully packaged but not both of them. Introducing these three constraints proved to be an effective way to systematically measure how the packaging efficiency of CCMV CP depends on the length of the RNA. We show that the relative packaging efficiency is a non-monotonic function of RNA length, having its maximum at the WT length.

Even though these experiments were carried out *in vitro* we observed some results reminiscent of the *in vivo* scenario in which the subgenomic RNA (RNA4) is co-packaged in a one-to-one mole ratio with RNA3: short RNAs (0.5 kb) are co-packaged along with the WT-length one (B1). RNAs ≤ 1.5 kb cannot compete to form capsids in the presence of B1, and with the exception of the 2.0 and 4.0 kb, RNA as the length of the RNA approaches that of the WT the packaging efficiency increases. The 2.0 and 4.0 kb RNAs had similar relative packaging efficiencies and did not follow this trend, but the interpretation of these results is complicated because a significant populations of "pseudo T = 2" and T = 4 capsids are formed, respectively. A surprising result was that CCMV RNA 1 (C1) has a lower efficiency of packaging by (its cognate) CCMV CP than does BMV RNA 1. The RNAs have the same length and are 80% homologous; hence one might expect they should have the same packaging efficiency. But C1 has evolved to be packaged by its cognate CP, suggesting from a biological point of view that C1 should have a greater packaging efficiency than B1. The experimental result implies that B1

could contain a series of short RNA sequences with higher affinity for CCMV CP that are either not present in C1 or are at a lower concentration.

Chapter 4 focuses on how multiple short RNAs can be co-packaged into a VLP. This is the first *in vitro* study in which an assembly pathway for RNA co-packaging has been examined for a (model) ss-RNA virus system. We show that the degree of cooperativity of virus assembly depends not only on the relative strength of the CP-CP and CP-RNA complexes but also on the assembly pathway. Since a single 0.5-kb RNA molecule cannot form a capsid by itself, it depends on collisions with other CP-RNA complexes to have the right overall size and CP concentration. As a consequence there are only two species – RNAs and VLPs – at CP:RNA mass ratio lower than the "magic ratio". A key result is that in the absence of strong RNA-RNA intermolecular interactions, association of multiple RNAs is mediated by the amount of bound CP. This assembly system follows a distinctive nucleation process that depends on the degree of association of multiple CP-RNA complexes and which might be associated with finite-size effects. Finally, by using a single-molecule Fluorescence Correlation Spectroscopy technique we show that four 0.5-kb RNAs are packaged per VLP, which is in agreement with our previous bulk measurements.

Chapter VI

Perspectives

We have shown that *in vitro* self-assembly of cowpea chlorotic mottle virus (CCMV) is a concerted process between its capsid protein (CP) and its RNA, and that this is affected by the length and number of packaged RNA molecules. The work presented in this thesis shows that RNA is not a passive cargo, but rather a key component that, along with the CP, determines the size of the capsid and its assembly mechanism. We found that there are three main RNA packaging regimes as a function of the length of packaged RNA; (i) for short RNAs (≤ 1000 nt) several molecules are packaged such that the average RNA content inside of the capsid is the same at that of the wild-type; (ii) for intermediate RNAs (~ 3000 nt) only one molecule gets packaged; and (iii) for long RNAs (> 4000 nt) a single RNA molecule is shared by two or more capsids. We also showed that when comparing the RNA packaging efficiencies of RNAs the meaningful quantity to be measured is the relative packaging efficiency of a competitor RNA with respect to a reference one under conditions in which only one of them can be fully packaged. By doing so we were able to determine that the relative packaging efficiency of ssRNA increases as it approaches the wild-type RNA length, at which it is maximum. Surprisingly, we found that when doing an equal-length competition experiment between CCMV and BMV RNAs 1, the latter one was packaged more efficiently by CCMV CP than CCMV RNA 1. Finally demonstrated that a capsid protein well know for non cooperative assembly around long RNAs ($\sim$3000 nt) can assembly cooperatively when several short RNAs (500 nt) are packaged.

This thesis focused on RNA-sequence-independent equilibrium properties of viral assembly *in vitro*, which are consistent with the fact that *in vitro,* CCMV CP is well known for its ability to self-assemble around almost any negative charged cargo with the appropriate size. However, some viruses like bacteriophages MS2 and R17 and the plant virus tobacco satellite necrosis virus exhibit a remarkable *in vitro* RNA packaging selectivity. Bacteriophage MS2 is one of the better-studied cases of RNA selectivity due to the presence of high-affinity RNA sequences (packaging signals [PS's]) across its genome. These PS's are known to be responsible for allosteric changes in MS2 CP that translate into efficient capsid assembly. These two ideas, RNA specificity and allosteric changes of the CP´s due to PS's, have not been deeply explored for CCMV assembly *in vitro*, hence they are extremely promising topics for further CCMV virion assemblies studies.

Two very appealing aspects to focus on in the future would be, the effects of PS's (if they exist at all for CCMV) on the relative packaging efficiencies for equal-length RNAs and the kinetics of assembly and capsid nucleation as a function of RNA sequence and length. Our group has started collaborating with Prof. Peter G. Stockley at The University of Leeds to try to determine by a SELEX experiments if CCMV and the closely related brome mosaic virus (BMV) exhibit high-affinity sequences. The idea of doing SELEX experiments in these two viruses is motivated by the unexpected results of the competition experiment between CCMV and BMV RNAs 1 mentioned above. We have suggested that it is possible that either the presence of a packaging site having particularly high affinity for CCMV CP or a particularly large number of PS-like sequences on BMV RNA 1 could be the responsible for its high relative packaging efficiency when compared to CCMV RNA 1.

The second set of future experiments suggested by results presented in this doctoral dissertation is the determination of the kinetics and assembly pathway of short RNAs around CCMV CP. By using single-molecule fluorescence correlation spectroscopy (sm-FCS) and a small fraction of labeled short RNAs, the rate of capsid assembly could be followed as a function of RNA length, sequence and concentration. This study could be further expanded to a more biologically relevant scenario by looking at the copackaging of CCMV RNA 3 and 4. This could lead to the first mechanistic study of RNA copackaging for a single-stranded RNA virus.

Finally, the results presented in this thesis open the door to start studying viral assembly pathways as a function of RNA length, sequence and concentration by taking concerted advantage of state-of-the art techniques such as SELEX, cryo-electron microscopy and single-molecule FCS. They will also serve as a basis for theoretical work carried out in our group and in collaboration with Prof. Avinoam Ben-Shaul at the Hebrew University of Jerusalem.

While there is still a great deal to understand about the assembly of single-stranded RNA viruses. In particular, future experiments have at testing whether the *in vitro* results have any biological significance. The ultimate molecular dissection of single-stranded RNA viruses has to be based on combining the simplicity of *in vitro* experiments with the biological relevance of the *in vivo* systems. While the results presented in this thesis create a clear scenario for the role of RNA in CCMV assembly *in vitro,* it is not obvious that PS, the length and number of packaged RNAs affect assembly *in vivo* in the same way. There are at least three *in vitro* results that could be tested *in vivo*: the relative packaging efficiency of non-viral RNAs with the same length as that of CCMV RNA1, the role of the putative PS and copackaging of RNA3 and 4. Protoplasts have been used as a good model to understand the life cycle of plant viruses, and hence could be

used to test the role of the length of the competitor RNA and the presence of PS. These two problems can be approached by carrying out competition experiments of non-viral RNAs (with the same length as CCMV RNA 1) with and without the putative PS (determined by the previously mentioned SELEX experiments) in the presence of all of CCMV genomic RNAs. The last *in vivo* project would be aimed at dissecting the effects of RNA length on CCMV RNAs 3 and 4 copackaging during an infection. The main idea would be to co-infect plants with wild-type CCMV RNAs 1 and 2 and different CCMV RNA3 molecules with lengths ranging from 2.1 to 2.8 kb (the same length as RNA2) to test if RNA4 co-packaging could be completely inhibited.

Appendix I

Labeling of cowpea chlorotic mottle virus capsid protein

with Alexa Fluor 647

Protein labeling with small fluorophores is a wide-spread technique that allows selective tagging, for a variety of *in vivo* and *in vitro* purposes. One of the most common methods involves reacting a solvent-exposed lysine with a fluorescent probe that is bound to a reactive N-Hydroxilsuccinimide (NHS) ester. In neutral to slightly alkaline conditions (pH 7.2 to 9) the NHS group reacts with primary amines ($-NH_2$) to yield stable amide bonds see (Fig. 1). Buffers such as TRIS ($(HOCH_2)_3CNH_2$) or glycine (NH_2CH_2COOH) have to be avoided since they contain primary amines, therefore establishing a competition between the cross-link reaction with the desired protein.
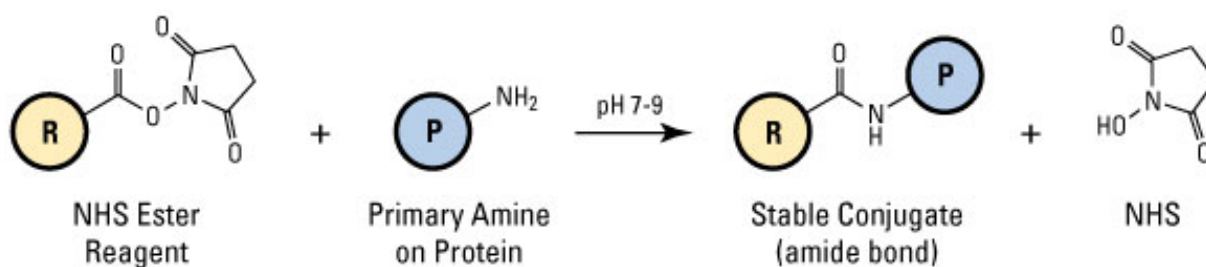
FIG 1. NHS-ester reaction scheme for conjugation to a primary amine. R represents the molecule (e.g., fluorescent dye) that will be cross-linked to the molecule of interest (i.e., a protein); (P) is the molecule of interest containing at least one solvent exposed primary amine (i.e., lysine) (Figure modified from http://www.piercenet.com/browse.cfm?fldID=F330F14F-EBCC-97DB-7F6E-9664D3ACE886)

By using this cross-link chemistry Gillitzer *et al* labeled CCMV virions by using the succinimidyl ester of 5-(and-6)-carboxylfluorescein (FAM)[1]. They pointed out that there are seven exposed lysines (K54, K84, K87, K65, K106, K131) per capsid protein that can be cross-

linked. By having an excess of 100 (300) FAM dyes molecules per protein they were able to add 180 (380) dyes per capsid.[1] However, such label density is too high for FCS; the ideal density for single-molecule experiments should be at least an order of magnitude lower.

The reaction between a succinimidyl ester and a lysine depends on pH; it is optimum at pH 8.3. Given that the required label density for FCS has to be considerably lower that what is reported by the vendor (Invitrogen) and Gillitzer et al., we performed a series of reactions at neutral pH's (7.5 and 7.2) and at lower AF647:CCMV mass ratios than used fro FAM:CCMV. It is well known that CCMV swells at neutral pH and at low ion strength in the absence of divalent metal ions[3], and that a pH $\geq$ 7.5, high ionic strength ($\sim$ 1.0 M) and EDTA the capsids disassemble, we chose neutral rather than alkaline pH's.

We carried out a series of cross-linking reactions at different AF647:CCMV mass ratios and pHs. To increase the recovery yield after purification we also used different methods for buffer exchange and virus purification. The density of labeling (DOL) was determined by measuring the absorbances at 260 and 650 nm (see Chapter 4); a representative spectrum is shown in Fig. 2. Table 1 summarizes the different experimental conditions, purification methods and an estimate of the recovery yield.

Table 1. NHS-ester CCMV reaction conditions.						
AF647:CCMV (wt/wt)	pH	MgCl$_2$ (mM)	Buffer exchange method	Purification method	DOL (dyes/capsid)	Virion recovery yield (%)
0.2 – 0.1	7.5	0	Amicon filter	Amicon filter	118	$\sim$ 25
0.05	7.5	0	Amicon filter	Amicon filter	77	$\sim$ 25
0.005	7.2	5	Dialysis	Amicon filter	12	$\sim$ 50
0.005	7.2	5	Dialysis	Sucrose cushion	19	$\sim$ 100

As shown in Table 1 the lowest DOL was achieved by using an AF647:CCMV (virion) mass ratio (wt/wt) of 0.005 at pH 7.2, with Amicon filter purification. This degree of labeling also allowed CP recovery identical to that of WT CCMV; at higher DOL the recovery of purified

121

CP was an order of magnitude lower than for the WT.

Even though the DOL could be decreased by lowering the pH from 7.5 to 7.2 and by adding $MgCl_2$, the recovery yield of the purified CCMV-AF647 virions was still low, i.e. less than a quarter of the initial amount. One possibility is that the virions were not stable at pH 7.2 and that the force applied to the sample, while using filtering devices (five to seven washes with a 100 KDa Amicon filter; 10,000 g for 7 minutes at 4 ºC). could damage the virions. To avoid capsid disruption we therefore lowered the pH by overnight dialysis. Finally, we noticed that after cross-linking the filtering membranes had the same color as the sample, indicating adsorption of the modified virions onto them. By using a 10% sucrose cushion, exactly as for the WT purification[2], we achieved recovery yield of almost 100%.
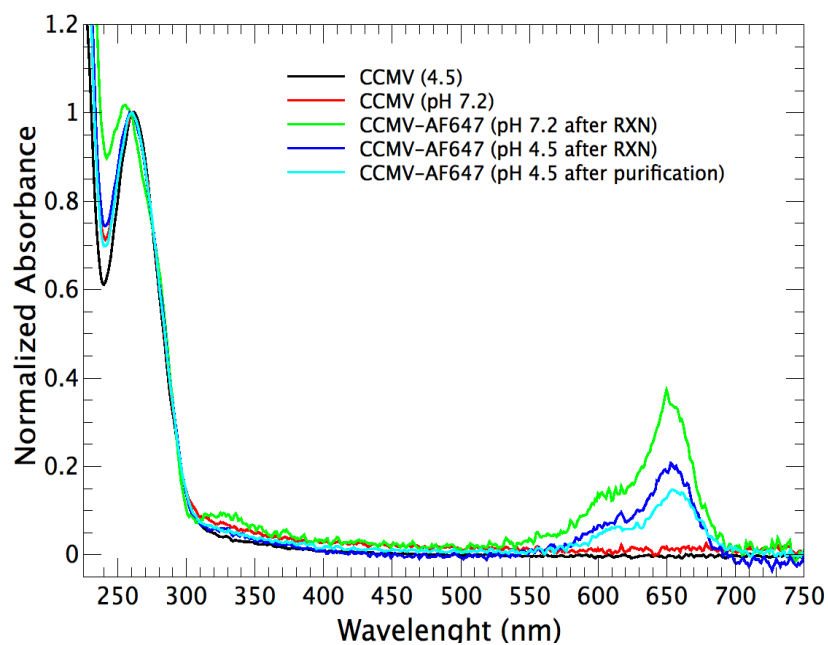
FIG 2. UV-Vis spectra of CCMV and AF647-labeled CCMV at different stages of the labeling and purification procedure. AF647:CCMV = 0.005, 0.1 M HEPES, 5 mM $MgCl_2$ and pH 7.2

As shown in Fig. 1 if the net charge of the AF647 succinimidyl ester is different from +1 (the original lysine charge), the formation of a covalent bond between a lysine and a NHS-ester

122

reagent should change the outer surface charge of the virion. The vendor has not released the chemical structure of this dye; however from the left panel in figure 3 we can deduce that the AF647 has a negative net charge. More explicitly, we observed that the electrophoretic mobility of labeled virions depended on the DOL (see Figs. 3 and 4): at a DOL of 77 dyes per capsid the electrophoretic mobility of the labeled virion is 13% higher than that of the WT (see Fig. 3); at 20 dyes (or less) per capsid, on the other hand there was no observable change with respect to the unlabeled virions (see Fig. 4). Since CCMV has 1260 solvent-exposed lysines, the net charge of AF647 is negative and a 6% lysine labeling increased the electrophoretic mobility by 13% then we can expect (in a first-order approximation) that AF647 has a net charge of -1.

* Before purification
† After purification
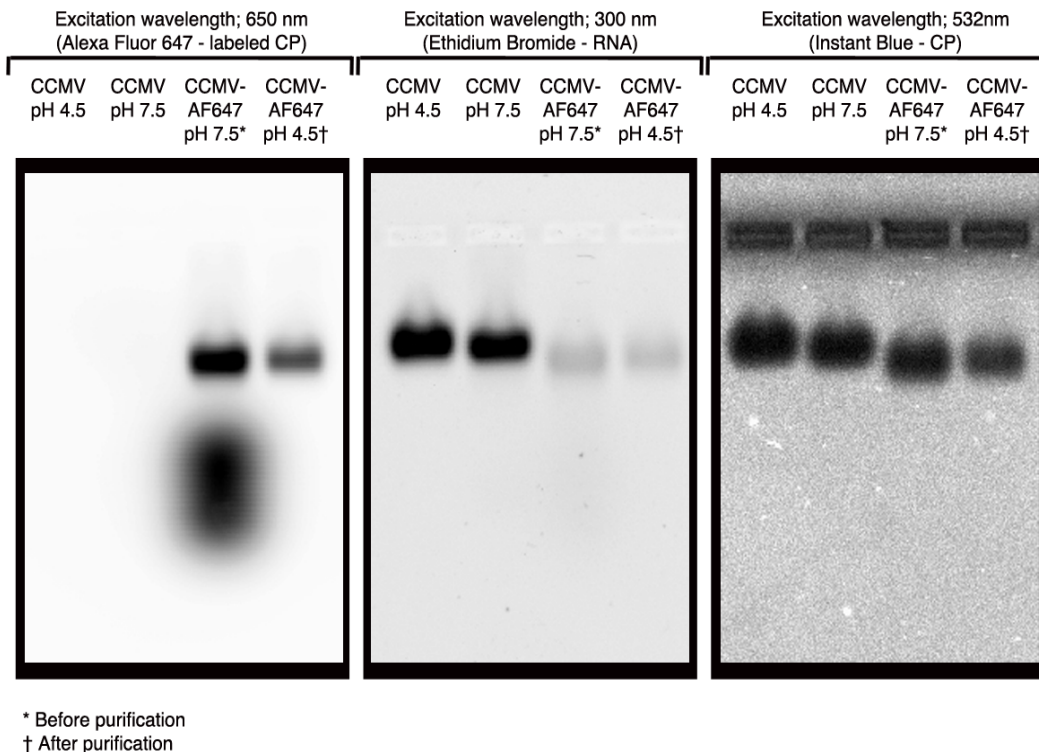
FIG. 3 EMA of labeled and unlabeled CCMV virions at different stages of the labeling protocol. The reaction was performed at a AF647:CCMV mass ratio = 0.05 and pH 7.5 (DOL = 77 dyes/virion). The electrophoretic mobility of the labeled virions was 13% greater than that of the unlabeled samples.
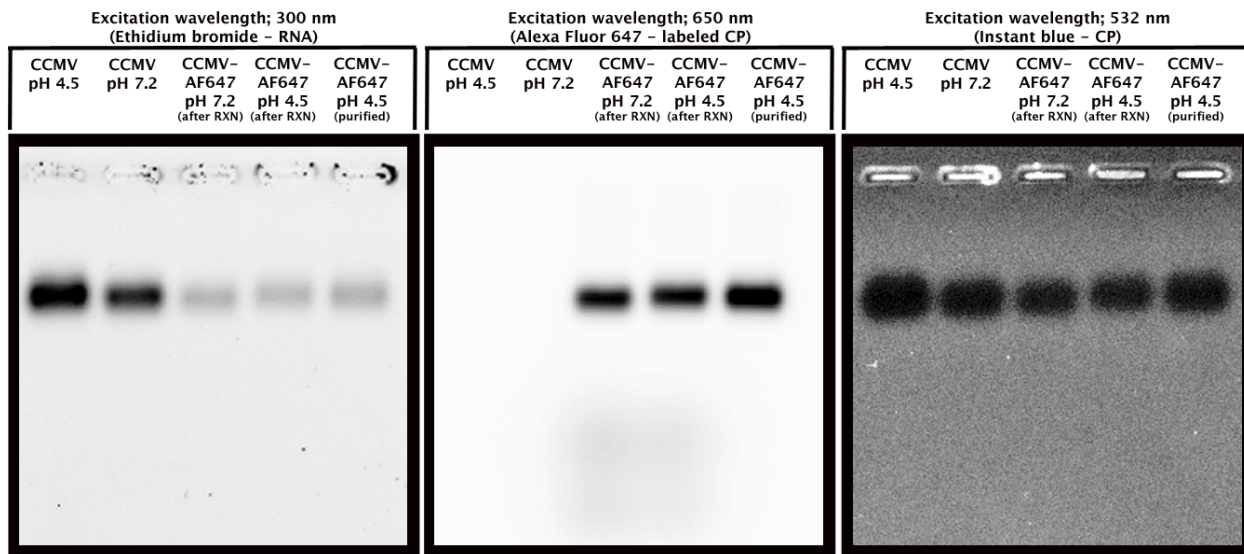
FIG. 4 EMA of labeled and unlabeled CCMV virions at different stages of the labeling protocol. The reaction was performed at a AF647:CCMV mass ratio = 0.005 and pH 7.2 (DOL = 19 dyes/virion). The electrophoretic mobility of the labeled virions was indistinguishable from that of the unlabeled samples.

After determining the reaction conditions to achieve low DOL, we investigated whether the self-assembly process of VLPs with AF488-500-nt RNA was inhibited or disrupted by the presence of a label on the CP. Figure 5 shows the EMA of a series of assembly reactions with labeled CP and RNA, which demonstrate that the degree of labeling did not affect the amount of packaged RNA. However there were some changes in the behavior of the VLP. When the DOL was 4.1% of the CP, the electrophoretic mobility of the VLPs was only slightly higher than that of the unlabeled ones, but the fluorescence from AF488-labeled 500-nt RNA was greatly affected by the presence of labeled CP. As the fraction of AF647 increased, the degree of both ethidium bromide and AF488 quenching increased (see Figure 6). The latter suggests that AF488 is acting like a FRET-pair donor and AF647 a FRET-pair acceptor. Although this FRET interaction has no effect on the assembly process it might affect the quality of AF488 fluorescence when using this VLPs for sm-FCS.
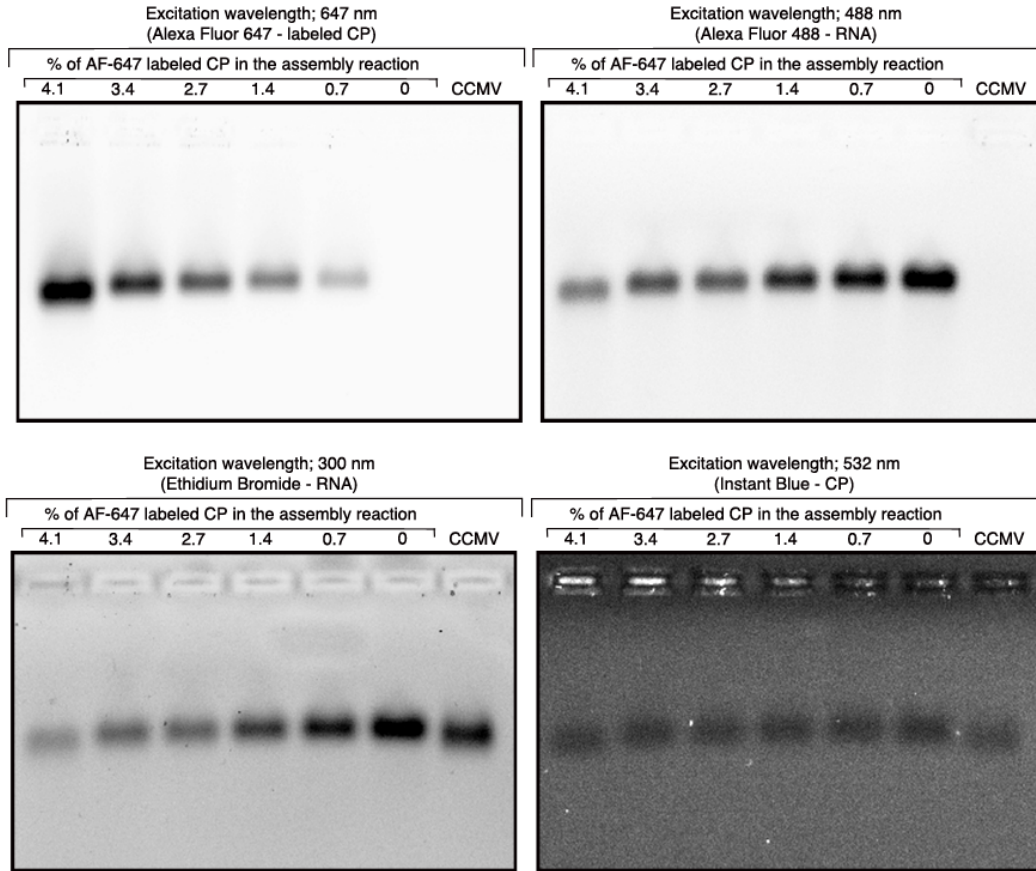
FIG 5. EMA of assemblies of AF488-labeled 500-nt RNA with various amounts of AF647 labeled CP. The electrophoretic mobility of the VLPs was affected only at a percentage of labeled CP of 4%. The quenching of ethidium bromide and AF488 increases as the fraction of labeled CP in the reaction mixture increased.
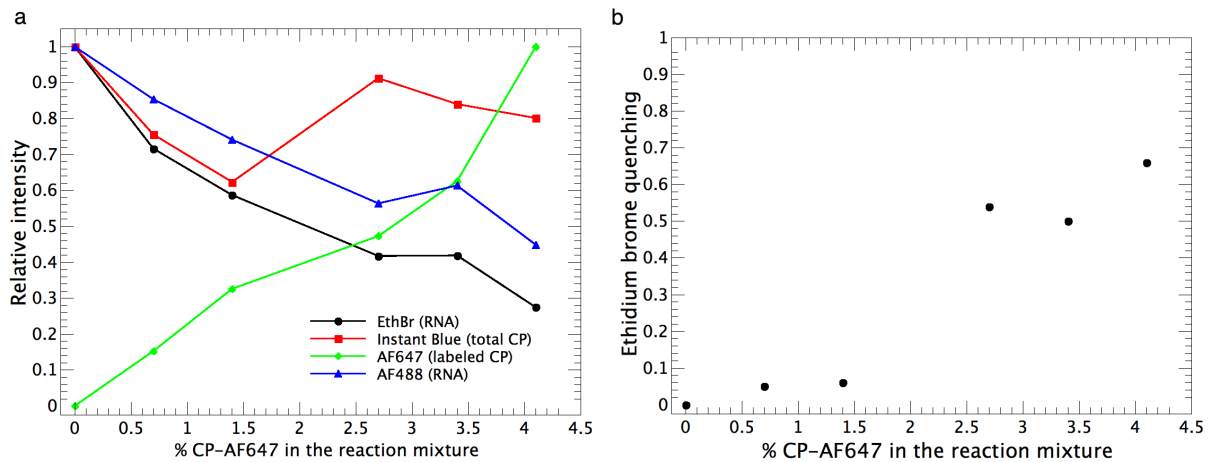

FIG 6. a) As the fraction of labeled CP increases (green diamonds) the relative intensity of both RNA dyes (ethidium bromide [black circles] and AF488 [blue triangles]) decreases while the intensity of Instant Blue (dye that measures total CP concentration) remains roughly constant. b) Quenching of the ethidium bromide increases as a function of % of labeled CP; Quenching = $1 - (E/I)_x/(E/I)_R$ where E and I are the signal from ethidium bromide and Instant Blue, respectively, and R and x are the samples with and without AF647-labeled CP, respectively.

125

Labeling protocol

Based on the findings described in this appendix we developed the following protocol that produces virions with a DOL of about 19 dyes/virion (1.5% of the solvent-exposed lysines) and almost 100% of recovery yield. This protocol requires CCMV, AF647 (10 mg/mL in DMSO), RXN buffer (0.1 M HEPES buffer, 5 mM $MgCl_2$ pH 7.2) and VSB (50 mM sodium acetate, 8 mM magnesium acetate pH 4.5).

Sample preparation

1) Concentrate 2 mg of CCMV to 5 mg/mL.

2) Transfer the sample to a dialysis tube and dialyze it against RXN buffer for 24 hours at room temperature.

3) Add 0.01 mg of AF647 to the sample.

4) Incubate the reaction for 2 hour at room temperature with continuous shaking.

5) Transfer the sample to a dialysis tube and dialyze it against VSB for 24 hours at 4 ºC.

6) Collect the sample and measure the volume.

7) Transfer it to a TLA110 tube and bring the volume up to 1.5 mL.

8) Add 0.25 mL a 10% Sucrose in VSB at the bottom of the tube.

9) Centrifuge the tubes at 100,000 rpm for 2 hours and 4 ºC.

10) Immediately discard the supernatant.

11) Add 0.5 mL of VSB to each tube.

12) Seal the tube with parafilm and cover it with aluminum foil.

13) Keep overnight at 4 ºC.

14) Collect the sample, transfer to an Eppendorf tube, vortex and microfuge.

15) Measure the UV-Vis absorbance

Determining the degree of labeling

1) The absorbance of the virions can be affected by the absorbance of AF647, hence the concentration of the virion has to be corrected; $Abs_{260}* = Abs_{260} - Abs_{650}*0.03$

2) Calculate the degree of labeling (DOL); $DOL = \frac{Abs_{650} \times Mw}{[Protein] \times \varepsilon_{AF647}}$

3) Calculate the concentration of CCMV; $[CP] = \frac{Abs_{260nm}^*}{5.8 \; a.u \cdot mL \cdot mg^{-1}}$

where M_w of the protein within a capsid $= 3.65 \times 10^6$ g/mol, $[Protein] = Abs_{260}^* \times \frac{0.785}{5.8} = [mg \; of \; CCMV \; CP/mL]$ and $\varepsilon_{AF647} = 239000$ cm^{-1}M^{-1}

References

1. **Gillitzer, E., Willits, D., Young, M.J. and Douglas, T.** 2002. Chemical modification of a viral cage for multivalent presentation. Chem. Commun. **20**:2390-2391

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

3. **Speir, J.A., Munshi, S., Wang, C., Baker, T. and Johnson, J.E.** 1995. Structure of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. Structure. **3**(1):63-78

Appendix II

Factors that affect single-molecule fluorescence correlation spectroscopy

The experiments in Chapter 4 involved extensive sm-FCS measurements. This appendix focuses on set-up optimization, artifacts due to laser intensity and high and low sample concentrations, and data treatment. As explained in Chapter 4, sm-FCS measures the correlation of fluctuations in the fluorescence intensity of a fluorescently-labeled molecule at extremely low concentrations (~ 1 to 10 molecules in a femtoliter confocal volume). Figure 1 describes the main conceptual idea behind the method.
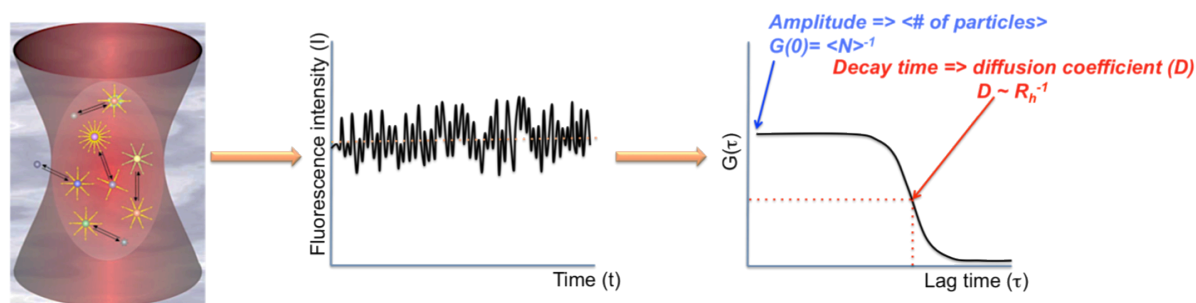
FIG 1. Schematic representation of sm-FCS. First, an extremely small confocal volume (~ fL) is achieved by using a confocal microscope and highly monochromatic source of light (a laser). The fluctuations of the fluorescence intensity, I(t), of a highly diluted labeled sample of molecules that diffuse through the confocal volume is recorded as a function of time. By using a built-in correlation algorithm the correlator card generates an autocorrelation curve, $G(\tau) \propto <I(t)I(t+\tau)>$. Depending on the nature of the experimental sample the autocorrelation curve can be fitted to an autocorrelation function that at least describes the size of the diffusion molecules.

Uncorrelated background and laser power

Accurate FCS measurements depend on working at the maximum laser power that produces the lowest amount of photobleaching of the fluorescent probe. Figure 2 shows that for pure AF488 (fixed concentration) at laser powers ≤ 50 µW there is a linear relationship between the average

number of counts and the laser intensity. At higher laser powers the intensity does not increase linearly, meaning that some fraction of the fluorophores has been photobleached.
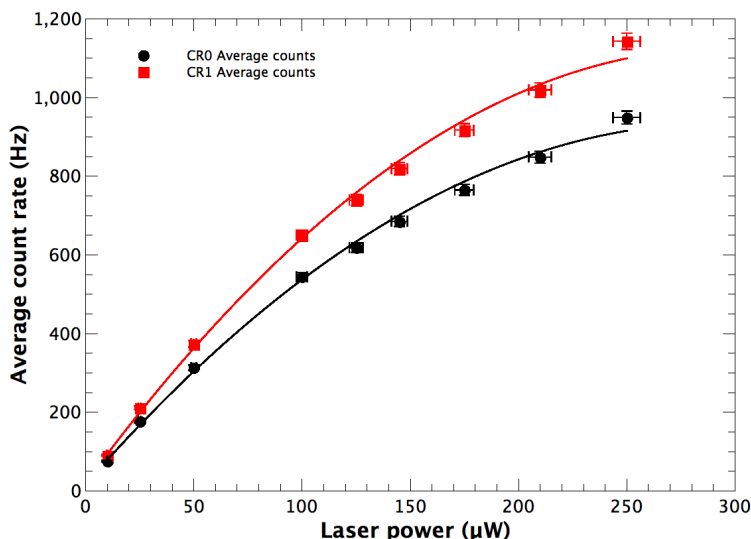

FIG 2. Saturation curve of the dye AlexaFluor 488 measured at a 50 nM concentration.

Based on Fig. 2 a series of experiments at a laser power of 100 µW were carried out. The first one was a dilution series with AF488 in water. From this experiment it was obvious that as the concentration decreases the average number of particles, <N>, has a non-monotonic behavior (see chapter 4 for a definition of <N>). This counterintuitive behavior is a consequence of an increasing contribution of the uncorrelated background signal, which is prominent at low sample concentrations. Correspondingly, there is a reduction of G(0) (amplitude of the autocorrelation function at $\tau = 0$) and increase in <N>. The effects of the uncorrelated background can be taken into account through a correction factor χ^2:[1-2]

$$\frac{1}{\chi^2} = \frac{1}{(1+\langle b\rangle/\langle f\rangle)^2} \quad \text{(Eq. 1)}$$

where <b> is the average background count rate measured in pure solvent, <f> is the virtual count rate of the actual sample without any background ($< f > = < F > - < b >$), which is calculated from the the measured count rate <F>. Figure 3 shows χ^2 increasing with decreasing

129

dye concentration. Even though the relationship between the effects of the uncorrelated background and the amplitude of the autocorrelation (Eq. 2) function were originally derived for dyes without triplet states (not the case for AF488), Fig. 4 shows that this is also applicable for AF448[1-2]. Such a relationship is described by the following equation:
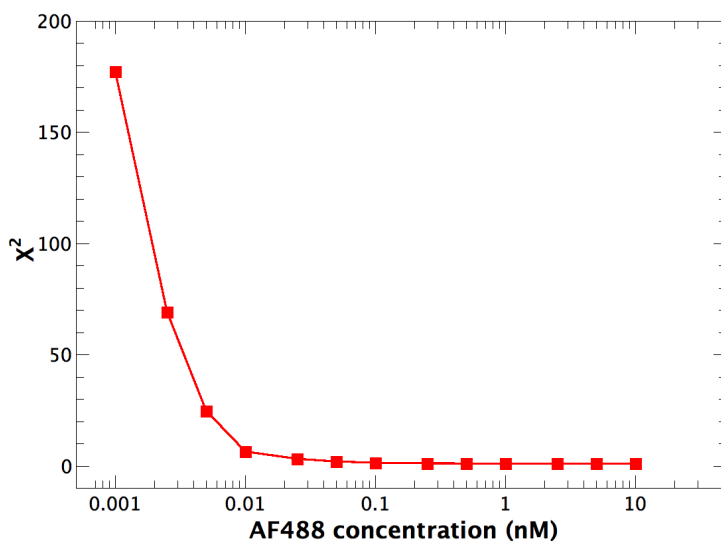
$$< N^* > = \frac{1}{\chi^2 G(0)} \quad \text{(Eq. 2)}$$

FIG 3. Correction factor that takes into account the uncorrelated background at low dye concentration.
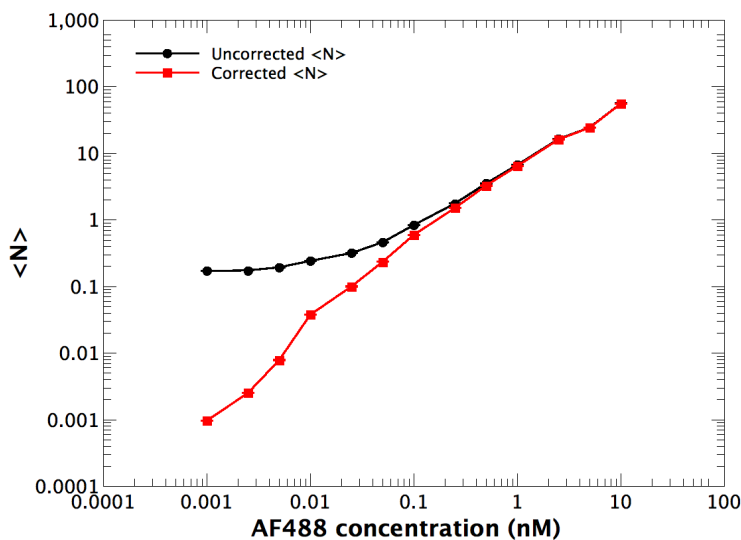
FIG 4. A dilution series for Alexa Fluor 488 shows that the uncorrected counts are not a linear function of AF488 concentration, but it is indeed linear if the uncorrelated background signal is taken into account.

130

Artifacts due to RNA concentration and fluorophore dynamics

Particle concentration in sm-FCS is crucial; at high concentrations the intensity fluctuations are too small and at low concentrations the signal-to-noise ratio increases; hence it was imperative to find the optimal particle concentration. The solid black line in Fig. 5 shows oscillatory fluctuations of $G(\tau)$ between 1 and 10 ms at high particle concentrations (130 nM 500-nt RNA [2.5 nM VLP]), which affect the quality of the fit but not the calculated R_h. More importantly there are other artifacts that decrease the quality of the measured data. For example, at extremely low concentrations (< 0.25 nM VLP [1 nM 500-nt RNA]) the part of the $G(\tau)$ related to the diffusion coefficient shifts by an order of magnitude towards shorter τ (see Fig. 5). Without further knowledge of the nature of the sample one could conclude that the R_h at 0.0125 nM is smaller than at 5 and 132 nM. These three samples are exactly the same, i.e. VLPs containing 500-nt long RNAs in VSB, but at different dilutions. Since at this pH VLPs are extremely stable (their melting point is 65 ºC) it is unlikely that they will dissociate at high dilutions.
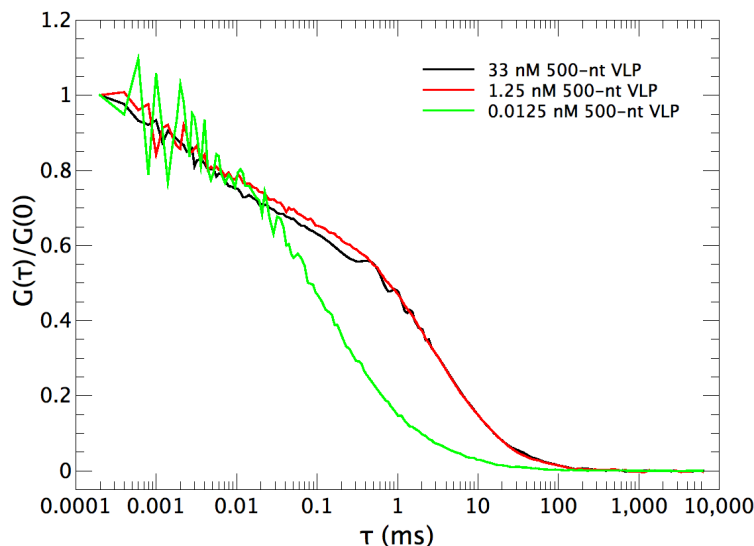

FIG 5. Normalized autocorrelation curves of 500-nt VLPs in VSB at different dilutions.

The reason for the apparent decrease in size that happens at picomolar concentrations is

not clear; it could be an issue related to the signal-to-noise ratio or a specific problem with the choice of dye. To determine if it was due to a decrease in the signal-to-noise ratio, we carried out a series of dilution experiments with VLPs labeled with either RNA (AF488) or CP (AF647). In order to fit the correlation curves and extract the value of the hydrodynamic radii the confocal volume has to be known. We therefore calculated it at different concentrations by using VLPs in VSB as standards. Fig. 6 shows the variation of the calculated confocal radii of the experimental setup for two different dyes at different concentrations. By using these calculated confocal radii we calculated the R_h for a series of diluted samples when exciting either the fluorophore in the exterior of the VLPs (AF647) or inside (AF488). From Fig. 7a is clear that the size of the measured complexes depends on the dilution factor only when exciting the dye incorporated into the RNA (AF488); when looking at the signal from AF647 the size of the CP-RNA complexes is independent of the dilution factor.
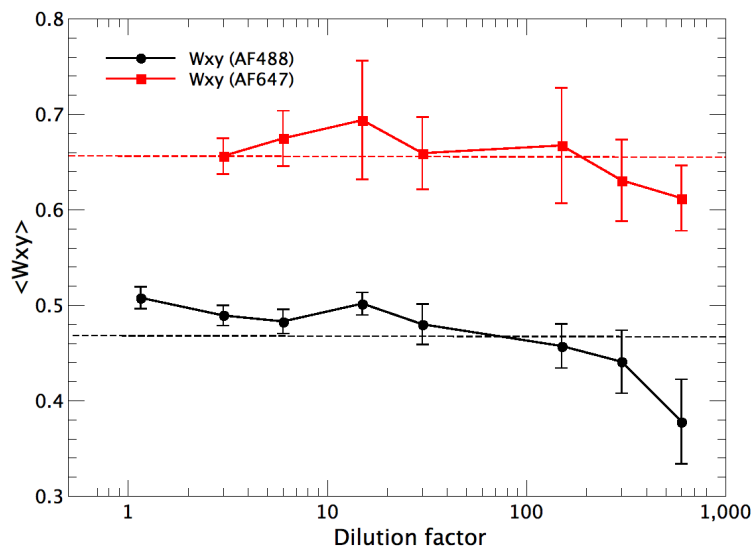

FIG 6. Radius of the confocal volume (Wxy) as a function of the dilution factor for two lasers (488 and 635) and two different dyes (AF488 and AF647). From this radius the confocal volume can be calculated by assuming a 2D Gaussian laser profile.

Figs. 7a and b are in disagreement; when looking at the signal from AF488 the size of the

132

CP-RNA decreases with increasing dilution factor while it is constant when exciting AF647. By combining these results with that of Fig. 5 we can conclude that at high dilutions there are artifacts related to the photophysics of AF488.

Fig. 7b shows that the size of CCMV CP at low pH decreases with increasing dilutions, which is in agreement with previous data for the self-assembly of CP at low pH and medium ionic strength. In 1969 J.B. Bancroft showed that pure CCMV CP forms multishells[3] at the pH and ionic strength of VSB. This self-assembly process is equivalent to that of micelle formation for which there is a critical concentration, which for the multishells is between 5 and 1 μM.
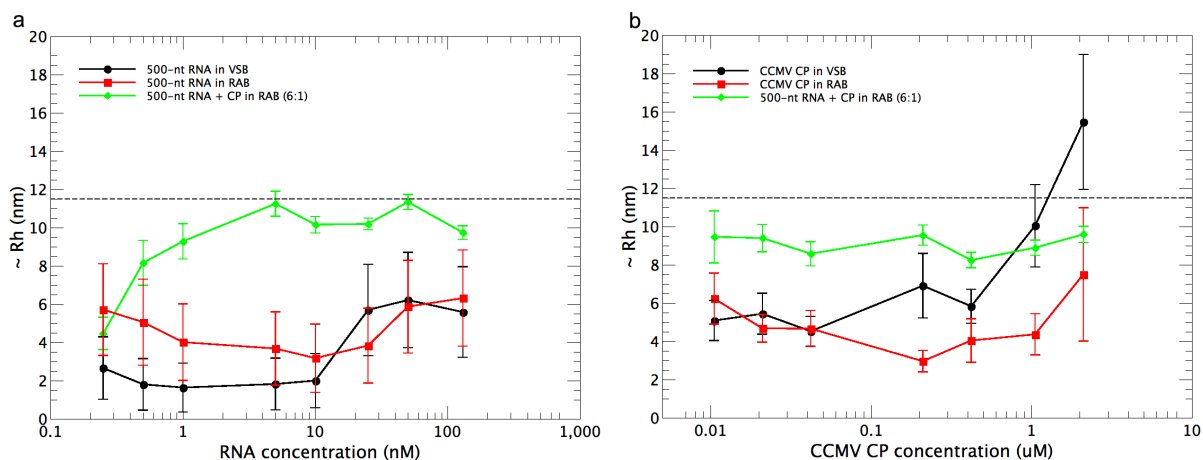

FIG 7. a) The R_h of CP-(AF488-RNA) complexes at neutral pH (green diamonds) decreases with dilution from 100 to 0.25 nM (while keeping the CP:RNA ratio constant at 6:1), suggesting that the number of RNAs in the complexes decreases. The red squares and black circles are RNA controls in VBS and RAB, respectively. b) The R_h of (AF647-CP)-RNA shows little change with dilution, indicating that the complexes do not dissociate at high dilutions. The large decrease in R_h for the CP in VSB (black circles) is consistent with the dissociation of multishells that form at low pH and medium ionic strength. The dotted line indicates the R_h of the 500-nt VLPs that was used to calibrate the confocal volume.

Fitting models

As described in Chapter 4 the shape of the confocal volume depends on the experimental configuration and it can be described by either a 2- or 3-dimensional Gaussian volume; the setup used for the sm-FCS measurement has a confocal volume better described by a 2-dimensional

volume (see Chapter 4). A detailed description of the mathematical model used to fit the data is explained in Chapter 4. This model includes two terms that describe the photophysics of the fluorophore: triplet state and photon bunching. To determine if these terms faithfully describe this process the acquired data were fitted by including them one by one, see Fig. 8.

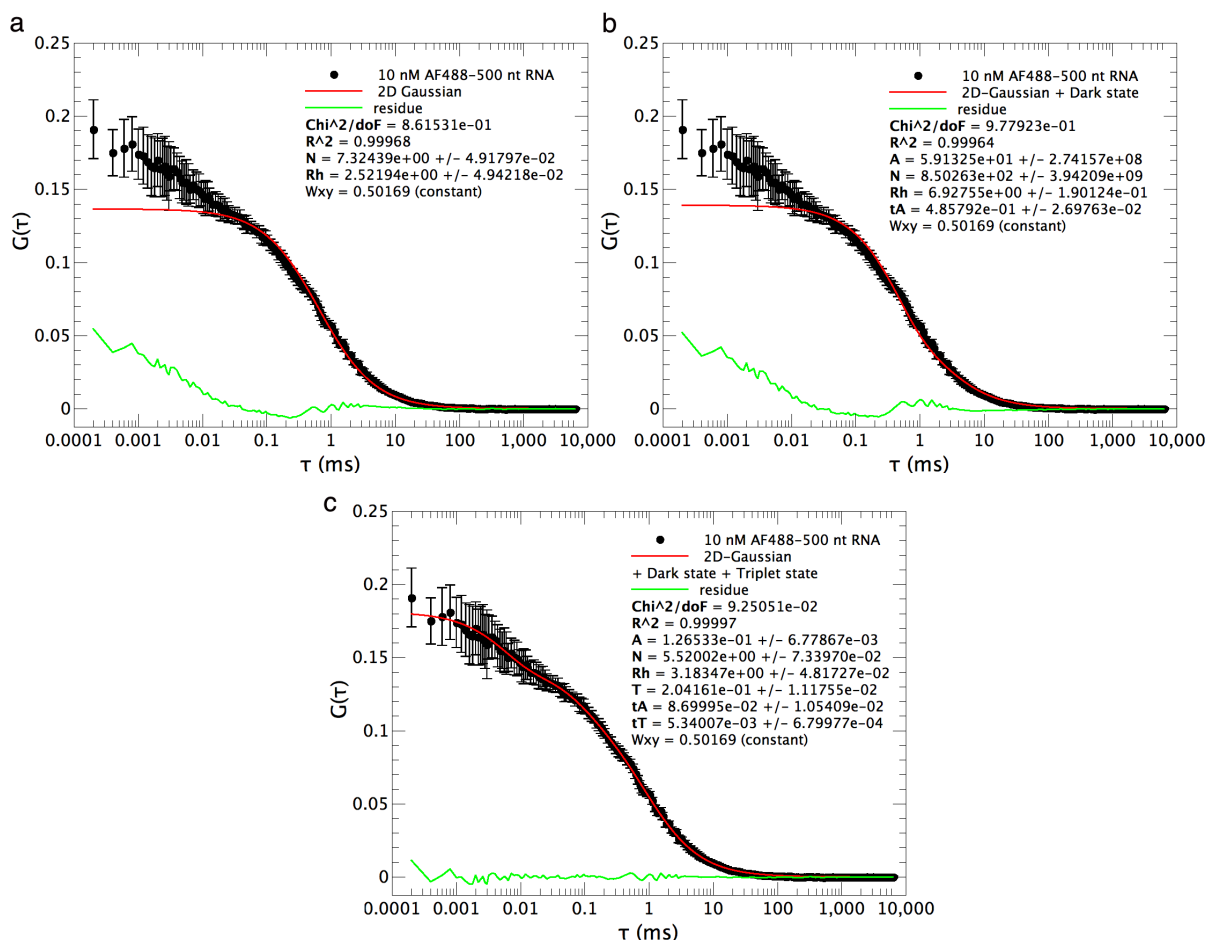When the fitting model has no other component other than diffusion (Fig. 8a) or when it

FIG 8. FCS data for a 500-nt RNA labeled with AF488 fitted by using different models (red solid curve); the green line is the residuals for each model and the error bars are the standard deviations of 40 measurements. Fitting models are shown that include: a) only the diffusion component; b) diffusion component and photon bunching; c) and diffusion component, photo bunching and triplet state.

only includes photon bunching (Fig. 8b) it cannot describe the experimental data at $\tau < 0.1$ ms.

Only by including all three components (Fig. 8c) is the fitting model able to describe the data. It

is worthwhile pointing out that even though these fast processes (photon bunching and triplet

134

states) occur at $\tau < \tau_D$ ($\tau_D \sim D^{-1} \sim R_h$) they significantly affect the calculated R_h.

References

1. Whal, M., Koberling, F., Patting, M. and Erdman, R. (2004) Curr. Pharm. Biotech. 5:299-308

2. Rigler, R., Mets, Ü., Widengren, J. and Kask, P. (1993) Eur. Biophys. J. 22:169-175

3. Bancroft, J.B., Bracker, C.E. and Wagner, G.W. (1969) Structures derived from cowpea chlorotic mottle and brome mosaic virus protein. Virol. 38:324-335