Effect of secondary structure on the size, configurational statistics, and packaging of long-RNA by viral capsid protein

Permalink
https://escholarship.org/uc/item/69z1x0kt

Author
Beren, Christian

Publication Date
2017

Peer reviewed|Thesis/dissertation
Effect of secondary structure on the size, configurational statistics, and packaging of long-RNA by viral capsid protein

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Chemistry

by

Christian Evan Beren

2017
ABSTRACT OF DISSERTATION

Effect of secondary structure on the size, configurational statistics, and packaging of long-RNA by viral capsid protein

by

Christian Evan Beren
Doctor of Philosophy in Chemistry
University of California, Los Angeles, 2017
Professor William M. Gelbart, Chair

Many viruses use long (thousands of nucleotides) single-stranded (ss)RNAs as their genomic material. Such viruses can be as simple as a single RNA molecule encapsidated inside a shell (capsid) composed of many copies of a single capsid protein (CP). Cowpea chlorotic mottle virus (CCMV) and bromo mosaic virus (BMV), the two sibling viruses studied throughout this work, are model ssRNA viruses that are capable of spontaneous self-assembly in vitro. This work aims to elucidate how RNA content affects both the structure of the particles formed and their physical properties. In particular, particles containing one or another of a variety of RNA molecules are compared using physical and biological tools. The results of this work further our understanding of the self-assembly of ssRNA viruses, and the physical forces that result in their structure and dynamical properties.

Chapter one is an introductory chapter, introducing the reader to viruses, ssRNAs and their secondary/tertiary structures, CCMV and BMV, and the in vitro self-assembly of these viruses. Chapters two and three discuss the production of viral-length polyU, an RNA molecule that lacks secondary structure due to its inability to base-pair or base-stack, and the packaging of this unique, structureless RNA with CCMV CP. Chapter four discusses the
cryo-electron microscopy (cryoEM) asymmetric reconstruction of BMV virions, which allows for unprecedented visualization of the RNA genome inside a multipartite virus—a virus with a genome comprised of more than one RNA packaged into several indistinguishable particles. Chapter five summarizes the results of this work, and provides some future experiments related to the work discussed herein.
The dissertation of Christian Evan Beren is approved.

Guillame Chanfreau

Giovanni Zocchi

William M. Gelbart, Committee Chair

University of California, Los Angeles

2017
Dedications

Thank you to my friends and family.
# Contents

## 1 Introduction

1.1 Secondary structure of long ssRNAs .................................................. 2

1.2 The bromoviruses: cowpea chlorotic mottle virus and brome mosaic virus . 6
   1.2.1 Self-assembly work ................................................................. 10

1.3 Thesis outline ....................................................................................... 13

## 2 Synthesizing and characterizing long polyU RNA molecules ............ 16

2.1 Synthesis and fractionation of long polyU RNAs .................................. 16
   2.1.1 Materials and methods .............................................................. 20

2.2 The size of polyU in solution ............................................................... 21

## 3 PolyU packaging by CCMV CP ............................................................ 23

3.1 Abstract ............................................................................................... 23

3.2 Introduction ......................................................................................... 24

3.3 Materials and methods ........................................................................ 27
   3.3.1 Buffers ....................................................................................... 27

3.3.2 PolyU synthesis and fractionation .................................................. 27

3.3.3 In vitro transcription of normal-composition RNA molecules .......... 28

3.3.4 Purification of CCMV CP ............................................................... 28

3.3.5 In vitro packaging of RNA by CCMV CP ..................................... 29

3.3.6 Electrophoretic mobility shift assay .............................................. 29
3.3.7 Transmission electron microscopy .............................................. 29
3.3.8 Competition experiments .......................................................... 30
3.3.9 RNase A digestion of virus-like particles ................................. 30

3.4 Results .......................................................................................... 30
3.4.1 PolyU is completely packaged at the magic ratio ...................... 30
3.4.2 PolyU RNA is packaged into 22-nm capsids ............................. 33
3.4.3 Multiplets form for longer lengths of polyU .............................. 35
3.4.4 PolyU outcompetes viral RNA for capsid protein .................... 37

3.5 Discussion ..................................................................................... 42
3.6 Acknowledgments .......................................................................... 46
3.7 Supplementary materials ............................................................... 47

4 CryoEM of BMV virions ................................................................. 49
4.1 Abstract ......................................................................................... 49
4.2 Introduction .................................................................................... 50
4.3 Results .......................................................................................... 53
4.3.1 In vivo synthesis of pure BMV3+4 ............................................ 53
4.3.2 Symmetric and asymmetric reconstruction .............................. 55
4.3.3 Capsid subtraction and asymmetric reconstruction ................. 57
4.4 Discussion ..................................................................................... 62
4.5 Supporting materials ..................................................................... 66
4.6 Materials and methods ................................................................. 68
4.6.1 Buffers ....................................................................................... 68
4.6.2 Strategy for in vivo assembly of pure virions ........................... 68
4.6.3 Sample preparation for cryo-electron microscopy ..................... 69
4.6.4 Cryo-electron microscopy and movie preprocessing ............... 69
4.6.5 Symmetric and asymmetric reconstruction ............................ 70
4.6.6 Structure refinement ................................................................. 70
5 Summary and perspective on future work

5.1 Summary ................................................................. 73
5.2 PolyU synthesis and packaging ........................................ 75
5.3 CryoEM of pure BMV virions ........................................ 78
  5.3.1 Future work on cryoEM analysis of BMV virions .......... 81

A Measuring the end-to-end distance of long ssRNAs

A.1 Introduction ............................................................. 82
A.2 Materials and methods ............................................... 84
A.3 Results ....................................................................... 87
A.4 Future work ............................................................... 90
A.5 Acknowledgments ......................................................... 90

Bibliography .................................................................... 90
Acknowledgments

I thank Professors William M. Gelbart and Charles M. Knobler for their advice throughout my thesis work, and for their unending thoughtfulness, positivity, and excitement towards science. I also thank Professors Hong Zhou and Jack Johnson for their support throughout the work discussed in Chapter 4.


Chapter 4 is a version of Beren, C., Cui, Y., Chakravarty A., Yang, X., Rao, A.L.N., Hong Zhou, Z., Knobler, C.M., and Gelbart W.M. Visualizing ssRNA inside brome mosaic virus capsids using high resolution cryo-electron microscopy. *In preparation.*
Vita

B. Reed Henderson High School, West Chester, PA 2011
Undergraduate Researcher with Professor Dmitri Vezenov, Lehigh University, PA, 2008-2011
Undergraduate Researcher with Professor Joachim Spatz, Max Planck Institute for Intelligent Systems, New Materials and Biosystems Division, Stuttgart, Germany, 2010
B.S. Chemical Engineering, Lehigh University, 2011
UCLA Hanson-Dow Teaching Assistant Award, 2013
Applications of Nanoscience Summer Institute Fellowship, 2015
Collegium of University Teaching Fellowship, 2016
Norma Stoddart Prize for Academic Excellence and Outstanding Citizenship, 2017
Publications


2. Beren, C.; Dreesens, L.L.; Liu, K.N.; Knobler, C.M.; Gelbart W.M. PolyU synthesis and fractionation. *In preparation, Bio-protocol*


Chapter 1

Introduction to long RNA molecules
and to in vitro packaging of long
RNA by viral capsid protein

The simplest viruses are made up of a genome comprised of long single-stranded (ss) messenger RNA and a single gene product, the capsid protein (CP), which forms a single-molecule-thick shell, encapsulating the RNA genome inside and conferring protection as the virus moves from cell-to-cell. These ssRNA viruses are capable of spontaneous self-assembly\cite{Fraenkel-Conrat1955}, and CCMV and BMV\cite{Bancroft1967}, the viruses studied in this work, are capable of spontaneous self-assembly in vitro. Additionally, both viruses have multipartite genomes made up of three ssRNA molecules plus a fourth subgenomic RNA (RNA4)–3200 nucleotide (nt) RNA1, 2800 nt RNA2, and one copy each of 2100 nt RNA3 and 700 nt RNA4–contained in three separate, indistinguishable, particles.

This work aims to understand how RNA sequence, and thus RNA secondary structure, affects the structure of virus particles. This is of interest as the CPs of these two virions have applications for drug delivery and are capable of protecting therapeutic RNAs for delivery to cells. Work that will not be discussed here further, but is ongoing in our lab, aims to use
CCMV and/or BMV CP to protect and deliver therapeutic, self-amplifying RNAs.

1.1 Secondary structure of long ssRNAs

Many biological systems, particularly many viruses, involve large ssRNA molecules (greater than 1000 nucleotides, nts) in a key role. While much is known about short RNAs, the physical properties of these larger RNAs are still poorly understood. Recently both theoretical modeling and experimental methods have been used to understand generic behavior of these large molecules[Yoffe 2008, Fang 2011a, Gopal 2012, Garmann 2015]. Most relevant here, is the fact that long ssRNAs represent a highly compact means of storing genetic information.

Figure 1.1: Left) Cryo-electron micrograph of the 2774 nt CCMV RNA 2 (blue circle) next to a CCMV virion (red circle). Right) Cryo-electron micrograph of 2117 nt ssRNA produced by in vitro transcription from the 2141 bp dsDNA template. Note that both the ssRNA and dsDNA contain the same amount of genetic information, but the dsDNA template is a larger, more rigid physical object[Gopal 2012].
An example, shown in Fig 1.1, illustrates how much smaller-in 3D space-a ssRNA molecule is as compared to its dsDNA template[Gopal 2012]. Both of these molecules encode the exact same genetic information, but they are clearly very distinct physical objects. Amazingly, Fig 1.1 shows that 2774-nt long CCMV RNA2 is roughly the same size as the capsid into which it is packaged[Gopal 2012]. This compact genomic material allows these viruses to spontaneously self-assemble, while their viral counterparts that use dsDNA have evolved complex motors to pack the DNA utilizing ATP-hydrolysis.

\[ \textbf{(5') CACAAACCACUGAACCCCCGGAACCGCUUCGUACCGGAUCAGAACCGCGU} \]
\[ \textbf{GAAGGAUCAACCGGUAGUUGCUCUGAUUCUCACCCAGUGCGAGG (3')} \]

**Figure 1.2:** Depicts how the linear primary sequence of ssRNA forms secondary structure through self-complementarity. The ssRNA behaves as a compact, branched polymer due to the formation of secondary structure.

The considerable amount (greater than 60%)[Yoffe 2008] of self-complementarity that arises in ssRNA molecules for any reasonable composition (e.g. roughly equal proportions of each nucleotide) allows them to encode genetic information efficiently in 3-D space. This self-base-pairing gives rise to RNA secondary structures, which cause ssRNA to behave as a branched polymer. Fig 1.2 illustrates how, through base-pairing, the linear sequence of
a ssRNA becomes a branched structure. This compaction due to the formation of branch-points becomes more striking when we look at the longer RNA secondary structure in Fig 1.3.

**Figure 1.3:** Left) Minimum free energy (as determined by M-FOLD) secondary structure of the 2774 nt CCMV RNA 2, and (Right) its mapping onto a branched tree-graph. The yellow line depicts the Maximum Ladder Distance (MLD), which is the longest path that can be followed across the RNA secondary structure.

This branched structure has a much smaller overall three-dimensional size, in this case represented by the yellow highlighted portion of the structure, called the Maximum Ladder Distance (MLD), which represents the longest path across the RNA and is proportional to the the 3D size of the RNA. Theoretical work has shown that viral RNAs have, on average, smaller MLDs than equal-length non-viral RNAs, implying that an evolutionary pressure exists for viral genomes to be compact\[Yoffe 2008\]. It has also been illustrated that the compaction of these viral genomes occurs because of the presence of higher-order vertices-branch points with more branches extending out of them-being positioned at the interior of secondary structure\[Borodavka 2016\] (see Fig 1.4).
Figure 1.4: The M-fold predicted minimum free energy secondary structure of the Qβ RNA genome (Top) and a random ssRNA (Bottom) of the same nucleotide length. Note that the viral RNA is smaller in 3D size due to the location of higher-order vertices at the center of the molecule. The Maximum Ladder Distance (MLD) of each RNA is highlighted in yellow [Yoffe 2008].
All of this work has illustrated that both the length of the primary sequence and the formation of RNA secondary structure are responsible for the size of a given RNA. A particular example of how secondary structure affects the physical properties of RNA was shown by Garmann et al.[Garmann 2015] In their work, two RNAs of equal nucleotide length were imaged using cryoEM. They found that these RNAs have different radii of gyration due to their respective secondary/tertiary structures—one RNA forms a highly branched, compact secondary structure with a smaller radius while the other is less branched and, therefore, more extended with a larger radius.

We are interested in how the formation of secondary structure affects the viral lifecycle. In particular, we are interested in studying how RNA secondary structure affects the assembly and disassembly of a virus, and much of this work is done through in vitro studies using either BMV or CCMV CP and physical tools for analysis (i.e. electron microscopy, fluorescence microscopy, uv-visible absorbance and gel electrophoresis).

1.2 The bromoviruses: cowpea chlorotic mottle virus and brome mosaic virus

Both viruses studied throughout this work are from the genus bromoviridae, and in fact, they are sibling viruses, i.e., the viruses are nearly identical in both physical and biological behavior. Both are plant viruses and utilize damage of the host-cell wall for cell-entry, and, as is true for all ssRNA viruses, they are directly translated and replicated in the host-cell cytoplasm to produce viral progeny. Their genomes are organized in identical fashion[Dasgupta 1982, Ahlquist 1984, Allison 1989, Dzianott 1991], with the only difference being the actual sequences encoded by their genetic material (see Fig 1.5), and their capsid proteins are nearly identical, utilizing a flexible, positively-charged N-terminus as an RNA-binding domain, a β-barrel core as the major structural motif, and a C-terminus that interacts with neighboring capsid proteins to hold the capsid together[Speir 1995, Lucas 2002].
Figure 1.5: Comparison of the BMV and CCMV genomes. While the sequences of the two viruses are different, their genome structure is nearly identical [Dasgupta 1982, Ahlquist 1984, Allison 1989, Dzianott 1991].

Physically, both wild-type virions are 28nm-diameter, T=3 Caspar-Klug [Caspar 1962] icosahedral capsids and are therefore comprised of 180 (60*T) copies of their respective capsid protein (CP) (see Fig 1.6) [Speir 1995, Lucas 2002]. Both viruses have genomes comprised of three molecules plus a fourth subgenomic RNA that are packaged into three distinct particles, a particle containing RNA1 (3200 nt), a particle containing RNA2 (2900 nt) and a particle containing RNAs 3 (2100 nt) and 4 (800 nt), and therefore, each particle has roughly 3000 nt of ssRNA inside [Dasgupta 1982, Ahlquist 1984, Allison 1989, Dzianott 1991].

Most importantly for this work, both viruses have been shown to be capable of spontaneous self-assembly in vitro to produce virions indistinguishable from wild-type virus [Bancroft 1967, Fox 1998]. It has been shown that a wide variety of cargo can be packaged inside these particles, ranging from synthetic nanoparticles [Douglas 1998, Aniagyei 2009], to heterologous RNAs of varying length and sequence [Cadena-Nava 2012], to negatively charged homopolymers like polystyrene sulfonate [Hu 2008, Cadena-Nava 2011] and polyU [Beren 2017].

Previous in vitro self-assembly work has shown that CCMV CP is capable of forming both wild-type, 28nm-diameter T=3 capsids and 22nm-diameter, T=2 capsids depending on the length of the RNA packaged (see Fig 1.7) [Cadena-Nava 2012]. T=3 capsids are composed of 90 CP-dimers (CP₂) organized into 12 pentamers and 10*(T-1)=20 hexamers, while T=2 capsids are composed of 60 (CP₂) subunits organized into 12 pentamers. Additionally, both
Figure 1.6: 3.1 Å resolution structure of the T=3 brome mosaic virus capsid determined by cryo-electron microscopy reconstruction. The 28-nm diameter particle has been colored radially.
Figure 1.7: Illustrates the difference between T=2 (left) and T=3 (right) capsids, both in cartoon form (top) and in negative-stain electron micrographs (bottom). The scale bar in the micrographs is 50 nm.
capsid structures have been shown to protect their encapsidated RNA from RNase digestion.

1.2.1 Self-assembly work

The variety of cargo that can be packaged by CCMV is remarkable, but the virus requires that the cargo be negatively charged. This is because the driving force for assembly is an electrostatic attraction between positively charged residues on the disordered CP N-terminus—called the Arginine Rich Motifs (ARMs)[Bayer 2005]—and the cargo that will be packaged. Further still, the assembly process utilizes a two-step assembly protocol[Cadena-Nava 2012]: the first step involves dialyzing a mixture of the RNA of interest and CP in neutral pH, high-ionic-strength buffer, which screens particle interactions between the CP and RNA, into neutral pH, low-ionic-strength buffer turning on the electrostatic attraction between the RNA and the CP N-termini; the second step involves dialysis into low pH, low-ionic-strength buffer, which reduces charge on acidic residues on the CP, thereby reducing repulsions, allowing attractive ”hydrophobic interactions” between CPs to be dominant (see Fig 1.8). It has been shown that after the first step of assembly disordered CP-RNA aggregates have assembled, with all of the CP bound to the RNA, but that it is not until the pH is lowered that icosahedral, RNase resistant capsids are formed[Garmann 2014].

Previous work investigating the effect of RNA length—measured in number of nucleotides—on the self-assembly of CCMV have shown that the virus is capable of packaging molecules as short as tens of nucleotides up to as long as 12,000 nt[Cadena-Nava 2012]. For RNA molecules shorter than 2000 nt the assembly products are smaller-than-wild-type (22 nm), T=2-sized particles. Any RNA longer than 2500 nt produces the wild-type (28 nm) T=3-sized capsid, but molecules over 5000 nt are packaged into multiplet capsids, several capsids assembled around a single RNA molecule. In fact, it has been shown that 6000 nt long RNAs, twice the viral genome length, package into doublet capsids, while 9000 nt long RNAs package into triplet capsids and 12,000 nt-long-RNAs package into quadruplet capsids (see Fig 1.9). These assembly experiments highlight the diversity of cargo and capsid morphology that
**Figure 1.8:** Depicts the virus morphology at various stages in the two-step assembly protocol. The top row is a cartoon depiction, with the middle row being negative-stain electron micrographs and the bottom row are cryo-electron micrographs. The first column and second columns are RNA and RNA mixed with CP, respectively, at neutral pH. The third column shows capsids, which have formed after lowering the pH, while the fourth column illustrates that capsids are stable upon bringing the pH back up to neutral pH after the pH drop. The final column illustrates the resulting aggregation if RNA and CP are mixed directly at low pH, skipping the first step of the assembly [Garmann 2014].
CCMV CP will allow.

**Figure 1.9:** Left) Singlet capsid formed from 3000 nt RNA. Middle) Doublet capsid formed around 6000 nt RNA. Right) Triplet capsid formed around 9000 nt long RNA [Cadena-Nava 2012, Beren 2017].

Experiments were carried out to understand how CP selectively packages its cognate RNA *in vivo*, as particles purified from the plant host contain nearly exclusively (greater than 98%) viral RNAs. These experiments pitted two RNAs against one another in an *in vitro* competitive self-assembly reaction (see Fig 1.10), to determine which RNA is preferentially packaged by CCMV CP [Comas-Garcia 2012]. The results of these experiments showed that more compact, viral-length RNAs were the most efficiently packaged, with the exception of a short 2000 nt RNA, which formed T=2-sized capsids that was able to compete reasonably well considering its short length. Surprisingly, BMV RNA1 was found to outcompete CCMV RNA1 for CCMV CP despite the latter being the CP’s cognate RNA.

Van der Schoot et al. [van der Schoot 2013] calculated the free energy associated with packaging both linear and branched polymers, and they find that the efficiency of encapsidation of a given RNA increased dramatically with increasing compactness. Other theoretical work by Singaram et al. [Singaram 2015] showed that molecules with more compact, branched secondary structures preferentially bound CP at neutral pH. Then, upon lowering the pH, the RNA that had been saturated by protein at neutral pH would be encapsidated, winning
Figure 1.10: Illustrates the scheme used during a competition experiment. First, two RNAs are mixed in equal mass amounts followed by the addition of capsid protein. This mixture is allowed to equilibrate at neutral pH overnight, at which point the pH is lowered and virions are formed. Typically the sample is then treated with RNase and the reaction mix is probed to determine how much of each RNA has been packaged.

the competition.

1.3 Thesis outline

Two separate but related projects are discussed herein; the first utilizes a homopolymeric RNA–polyU–that lacks the traditional RNA secondary/tertiary structure to investigate how this lack of secondary structure affects the in vitro self-assembly of CCMV capsids, while the second uses cryo-electron microscopy (cryoEM) to investigate the effect of the packaged RNA on the structural properties of BMV virions.

The second and third chapters discuss how polyU RNAs, which are devoid of secondary structure due to their inability to self-base-pair and negligible stacking energy, are packaged by CCMV CP. Specifically, chapter two focuses on how fluorescently-labeled, viral-length polyU RNAs are produced, and characterizes the size of viral-length polyU RNAs compared to similar length normal-composition RNAs using Dynamic Light Scattering. Chapter three
discusses work on \textit{in vitro} packaging of viral-length polyUs, and competition reactions be-
tween viral RNA and viral-length polyU. The results of this work show that secondary
structure is necessary for the formation of well-formed virus-like particles (VLPs). In addi-
tion, this work brings to the fore many aspects of viral self-assembly that are still poorly
understood.

The fourth chapter focuses on cryoEM studies of BMV virions containing different RNAs.
We are in the unique position of having access, through our collaborators Dr. Hong Zhou at
UCLA and Dr. Annamalai Rao at UC Riverside, to both a world-class electron microscopy
facility and pure-RNA-containing BMV virions (i.e. virions containing only one or another
of the viral genes) produced \textit{in planta}. We have acquired cryo-electron microscopy data sets,
using the FEI Titan Krios microscope, for \textit{in planta}-produced BMV particles containing
BMV genomic RNAs 3 and 4 (BMV3+4), as well as for BMV particles containing only
BMV genomic RNA 3(BMV3).

We have determined both a near-atomic-resolution symmetric reconstruction of the
BMV3+4 particles, allowing for uniquely high resolution of the capsid protein shell, and a
near-atomic-resolution asymmetric reconstruction of these particles as well. Previous studies
have involved icosahedral-averaging of the data set, necessarily obscuring any structures that
do not have this symmetry (i.e. the viral RNA genome inside). We have acquired unprece-
dented data sets for these BMV virions, typically involving around 100,000 particles, and our
asymmetric structure is the first-ever of a multipartite, ssRNA genome inside a symmetric
viral capsid.

Using these singular data sets we visualize protein-RNA contacts that are crucial for
the packaging of viral genomes into well-formed, icosahedral capsids. Future work aims to
clarify differences in capsid structure and dynamics between the three distinct wild-type (wt)
particles of BMV, which may be relevant to the viral life-cycle insofar as different particles
interact differently with host-cell machinery based on their RNA cargo. Our collaborators
Peter Doerschuk (Cornell) and Jack Johnson (Scripps) have started using these cryoEM data
sets to compare the dynamical properties of the BMV3+4 and BMV3 particles.
Chapter 2

Synthesizing and characterizing long polyU RNA molecules

Long single-stranded (ss)RNA molecules are compacted by secondary structure, formed through self-base-pairing, resulting in ssRNA viruses having genomes roughly the same size as the viral capsid [Yoffe 2008, Gopal 2012]. This compactification of the genetic information allows ssRNA viruses to assemble spontaneously [Bancroft 1967, Hiebert 1969], and previous work has shown that more compact RNAs, namely viral RNAs, are more efficiently packaged by CCMV CP [Comas-Garcia 2012, Singaram 2015]. We propose to study viral-length polyU RNA, which cannot base-pair and does not form secondary structure [Martin 1962, Richards 1963], to compare its physical properties to those of normal-composition RNAs—composed of comparable numbers of each of A, U, G and C nucleobases. Furthermore, we plan to package polyU with CCMV CP to investigate how secondary structure influences viral assembly.

2.1 Synthesis and fractionation of long polyU RNAs

The synthesis protocol used herein was developed from existing protocols used by Vanzi et al. [Vanzi 2003] and van den Hout et al. [van den Hout 2011]. Specifically, polyU RNAs are
synthesized from UDP monomers using the enzyme polynucleotide phosphorylase (PNPase, Sigma Aldrich), which indiscriminately adds nucleotide diphosphates (NDPs) to the 3’ end of oligonucleotides, in this case a 20-nt-long polyU seed oligo with a 5’ cy5 fluorescent label (IDT DNA). The oligo can be tagged at the 5’ end with fluorophores, linkers, or modified bases, enabling the synthesis of polyU with customizable markers. Because of its lack of base-pairing, polyU cannot be visualized in electrophoresis gels by the usual intercalating nucleic acid stains. For this reason we chose to utilize a fluorophore at the 5’ end of the RNA.

PNPase has the ability to synthesize polynucleotides without a template, and a synthesis of polyU using Alexa-fluor 488 fluorescent fUTP was carried out as well, resulting in viral-length RNAs which have a low-level of fluorescent UTP incorporation (Data Not Shown). However, this avenue of synthesis was not explored further in this work as it makes labeling the resulting RNA more challenging.

The synthesis reactions produce a polydisperse mixture of fluorescent polyU RNAs ranging in length from 200-10000 nts, with shorter lengths being represented in significantly higher copy number as determined by fluorescence detection in denaturing agarose gel electrophoresis. This polydisperse RNA mixture is run on a denaturing agarose gel alongside a denatured ssRNA ladder and fractionated by electroeluting portions of the gel using the ladder as a reference.

We used this process to fractionate polyU into the following length fractions of polyU: 500-1500, 1500-2500, 2500-3500, 3500-5000, 5000-7000 and 7000-9000 nt. The amount of polyU purified was quantified using uv-vis absorbance measurements, and the absorbance ratio at 260nm/280nm is used as a measure of RNA purity (A260/A280 greater than 2.0). These fractionated polyU samples were then examined on a denaturing agarose gel to demonstrate that single contiguous RNA bands are obtained (see Fig 2.2).
Figure 2.1: Electrophoretic gel analysis of unfractionated polyU (red) after synthesis using PNPase next to a double-stranded (ds) DNA ladder (green), illustrating the high amount of polydispersity in the polyU generated from the synthesis reaction.
Figure 2.2: Denaturing electrophoretic gel analysis of fractionated polyU. From right to left: ssRNA ladder, 5000-7000, 4000-5000, 2500-4000, 1500-2500 and 500-1500 nt polyU RNAs. The decrease in band intensity for higher molecular weight polyU molecules is due to the fact that an equal mass of polyU was loaded per lane, resulting in many fewer RNAs in the higher molecular weight fractions, and subsequently, a decreased fluorescence signal.
2.1.1 Materials and methods

PolyU synthesis

The synthesis protocol used herein was developed from existing protocols used by Vanzi et al. [Vanzi 2003] and van den Hout et al. [van den Hout 2011] Polynucleotide phosphorylase from Synechocystis sp. (PNPase, Sigma) was used to polymerize fluorescent polyU from UDP monomers, which was initiated by a short “seed” oligo (500 pmol of a 5’ tagged cy5 polyU oligo 20-mer, Integrated DNA Technologies). The reaction was initiated by mixing 500 pmol of seed oligo with 20 mM UDP (Sigma), 1 ul of RNase Inhibitor (New England Biolabs) and 300 ng of PNPase in 100 mM Tris-HCl pH 9 buffer solution containing 1.0 mM EDTA. After a 2 hour incubation at 37°C, the UDP concentration was increased to a final concentration of 40 mM, and the solution was incubated for a total time of 4 hours at 37°C. The product was thoroughly washed using TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and filtered through a 100-kDa Amicon filter (EMD Millipore, Inc).

PolyU purification

Fluorescent polyU was run on a 1% agarose/TAE (40 mM Tris pH 7.4, 20 mM acetic acid, 1 mM EDTA) RNA electrophoresis gel at room temperature and 90 V for two hours alongside a ssRNA ladder. Before loading on the gel, the ladder was denatured with formamide (70% of sample volume per lane) and heated to 70°C for 10 min to eliminate all secondary structure. After disconnecting the power source the gel was soaked in TAE, allowing the ladder to reform secondary structures and to be stained with ethidium bromide. Fractions containing polyU of 500-1500, 1500-2500, 2500-3500, 3500-5000, 5000-7000, and 7000-9000 nt were cut from the gel and electroeluted with TAE buffer for one hour at 100 V. After electroelution, the RNA was washed using TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) through a 100-kDa Amicon filter. Concentrations of poly(U) were determined by UV-Vis spectroscopy using the absorbance at 260nm. The purity of the electroeluted sample was subsequently
verified by RNA gel electrophoresis.

2.2 The size of polyU in solution

The hydrodynamic radii of three polyU RNA fractions (500-1500, 1000-2500 and 2500-3500 nt) and three normal-composition RNAs (500, 2000 and 3200 nt) were determined by Dynamic Light Scattering (DLS) (see Fig. 2.3). The right-hand plot of radius against length shows that the polyU radii exhibit $N^{1/2}$ power-law scaling and that the normal-composition RNAs exhibit $N^{1/3}$ power-law scaling, the scaling predicted for normal-composition RNAs by previous work using folding algorithms\cite{Fang2011a}. In addition, and more importantly for this work, the radius of polyU RNAs are approximately 25 percent larger than normal-compositions of similar length.

We attempted to use cryo-electron microscopy, which utilizes scattering off of the phosphate backbone in double-stranded portions of the RNA to resolve RNA molecules in solution, to investigate the size of polyU. However, due to the lack of base-pairing in polyU, these RNAs were not able to be visualized with the microscope.
Figure 2.3: Left) Raw DLS data of polyU and BMV RNA 1 (3200 nt long normal-composition RNA). The fraction of molecules of a given hydrodynamic radius (nm) is plotted for each sample. The distributions are broad, but the average size is consistent. Right) Plot showing how the hydrodynamic radii of polyU and normal-composition RNAs varies with nt length. The magenta line that has been drawn through the data represents $N^{1/2}$ power-law scaling (polyU), while the black line represents the $N^{1/3}$ scaling expected for RNAs with secondary structure. Note that though the raw data exhibit a large standard deviation, the standard deviation of the mean (error bars) is relatively small, indicating that the average size is reproducible across measurements.
Chapter 3

PolyU packaging by CCMV CP

3.1 Abstract

Previous work has shown that purified capsid protein (CP) of cowpea chlorotic mottle virus (CCMV) is capable of packaging both purified single-stranded (ss)RNA molecules of “normal” composition (comparable numbers of A, U, G, and C nucleobases) of varying length and sequence, and anionic synthetic polymers such as polystyrene sulfonate. We find that CCMV CP is also capable of packaging polyU RNAs, which – unlike normal-composition RNAs – do not form secondary structures and which act as essentially structureless linear polymers. Following our canonical two-step assembly protocol, polyU RNAs ranging in length from 1000 to 9000 nucleotides (nt) are completely packaged. Surprisingly, negative-stain electron microscopy shows that all lengths of polyU are packaged into 22-nm diameter particles despite the fact that CCMV CP prefers to form 28-nm diameter (T=3) particles when packaging normal-composition RNAs. PolyU RNAs greater than 5000 nt in length are packaged into multiplet capsids, in which a single RNA molecule is shared between two or more 22-nm diameter capsids, in analogy with the multiplets of 28-nm diameter particles formed with normal-composition RNAs longer than 5000 nt. Experiments in which viral RNA competes for viral CP with polyUs of equal length show that polyU, despite its lack of
secondary structure, is packaged more efficiently than viral RNA. These findings illustrate that the secondary structure of the RNA molecule – and its absence – plays an essential role in determining capsid structure during the self-assembly of CCMV virus-like particles.

3.2 Introduction

Many viruses use long (thousands of nucleotides, nts) ssRNAs as their genomic material. Such viruses can be as simple as a single RNA molecule encapsidated inside a shell (capsid) composed of many copies of a single capsid protein (CP). The capsid serves to protect the genome from degradation as it moves in and between host cells. Several ssRNA viruses are capable of spontaneous in vitro self-assembly[Bancroft 1967, Fraenkel-Conrat 1955], driven by a combination of attractions between the capsid protein and the RNA molecule (CP-RNA) and interactions between neighboring CP subunits (CP-CP)[Garmann 2014].

CCMV, the virus studied throughout this work, is a model ssRNA virus whose virions are composed of 180 identical CPs arranged into 28-nm-diameter icosahedrally symmetric capsids with a T=3 Caspar-Klug triangulation number[Caspar 1962], and the virions contain either one copy of 3200-nt RNA1, or one copy of 2800-nt RNA2, or one copy each of 2100-nt RNA3 and 700-nt RNA4 – hence about 3000 nt per capsid. They can be reconstituted in vitro from purified components producing particles that are indistinguishable from the native virus[Fox 1998]. More importantly still, for our purposes, is the fact that CCMV CP is able to package non-viral ssRNA molecules of varying length and sequence [Hiebert 1969, Cadena-Nava 2012, Hu 2008]. It is important to note that these in vitro assembly reactions rely on a two-step protocol in which the components are dialyzed into low pH (4.5) buffer, which may not be physiologically relevant. Further, a superstoichiometric amount of CP is used in these experiments, as this has been found to be necessary to package all of the RNA in vitro.

Recent experimental and theoretical work has suggested that long ssRNAs exhibit generic
properties that enhance their ability to be packaged by viral CP. Specifically, any long RNA molecule with comparable numbers of the bases A, U, G and C – i.e., ”normal” composition – results in secondary structures in which there is extensive Watson-Crick base pairing, typically involving more than 50 percent of the nts [Fang 2011b]. Because these secondary structures involve a large number of single-stranded loops from which three or more duplexes emanate, they cause these RNAs to behave effectively as branched polymers in solution [Fang 2011a]. This branching results in RNA molecules being significantly more compact than structureless linear polymers of similar chain length. The secondary structures associated with viral sequences of RNA involve a larger number of three-fold and higher-order ”branch points” than do non-viral normal-composition RNAs, and are as a consequence more compact [Garmann 2015, Yoffe 2008]. Further, ssRNA viral genomes have been shown to have radii of gyration comparable to the radii of the virions into which they are packaged [Gopal 2012].

For RNA molecules of comparable length and similar extent and nature of branching, particular local components of secondary structures, e.g., stem loops with specific sequences in their duplex and single-stranded portions, have been implicated in the preferential packaging, in vivo, of viral RNA over cellular RNA. ”Packaging signals” of this kind have been identified in the RNA genomes of several different viruses, and have been established as the origin of nucleation of viral capsids in several in vitro bulk-solution and single-molecule measurements [Basnak 2010, Turner 1986, Rolfsson 2016, Stockley 2016, Patel 2015].

Competitive self-assembly studies have recently probed the key role of sequence and length in determining the packaging of an RNA [Comas-Garcia 2012]. In these experiments, equal masses of two different RNA molecules are allowed to compete for an amount of CCMV CP sufficient to completely package only one of them. It was found in such competitive assemblies of branched RNAs of different lengths and sequences that the 3000 nt CCMV RNA1 (C1) outcompeted shorter and longer RNAs. But, surprisingly, the heterologous Brome Mosaic Virus RNA1 (B1), essentially identical in length to C1, but with different
sequence and hence different secondary structure, outcompeted C1 for its own CP.

The role of secondary structure on packaging has been examined in reaction scheme descriptions of the assembly process[Zlotnick 2013], theory[Perlmutter 2013], and computer simulation[Erdemci-Tandogan 2016]. Erdemci-Tandogan, et al.[Erdemci-Tandogan 2016] employed mean-field theory to investigate how the secondary structure of an RNA affected the electrostatic interactions with CP and the stability of the assembled virions. They found a deeper minimum for the packaging free energy of a branched polymer, compared to that of a linear one with the same number of monomers, and that the optimal packaged length was greater. Perlmutter, et al.[Perlmutter 2013] found similarly that the formation of RNA secondary structure allows longer RNA molecules to be packaged by viral CP. Additionally, Monte Carlo simulations of competition experiments have shown that more compact, branched, RNA molecules outcompete less branched RNAs for binding of CP[Singaram 2015].

To provide a direct experimental test of the effect of branching on the self-assembly of virions we have undertaken in vitro packaging studies of polyU, an RNA molecule with no secondary structure. In addition to having no base pairing, the stacking interaction between uridines is very weak, so polyU is an essentially structureless linear polyelectrolyte except at low temperature, where it exhibits some helicity[Richards 1963]. We examine how these unstructured RNAs are packaged in vitro by CCMV CP, and how they compete with viral RNA for CP, to determine if the presence of secondary structure/branching allows the viral genome to be selectively packaged by its cognate capsid protein. Throughout our discussion we will refer to polyU molecules alternately/interchangeably as ”linear” and structureless polymers, and normal-composition RNAs as ”branched” and structured polymers.

We find that polyU RNAs are packaged by CCMV CP into capsids with a characteristic size of 22 nm, smaller than the 28-nm diameter of the wild-type (wt) virus, and that their size is independent of the length of the polyU packaged. For lengths of polyU significantly longer than the 3000-nt wt genome, polyU is packaged into multiplets, in which a single RNA
is shared between two or more capsids. The packaged polyU is protected against RNase, demonstrating that the particles are closed. Electron microscope images of the particles show that although they have sizes typical of “T=2” capsids, a majority of the particles are not icosahedrally symmetric and are significantly polydisperse. A series of experiments in which polyU competes with normal-composition RNAs of equal length demonstrates that, surprisingly, and in contrast to theoretical expectations, the unstructured RNA is preferentially packaged.

3.3 Materials and methods

3.3.1 Buffers

The following buffers were employed: Synthesis Buffer-100 mM Tris-HCl pH 9, 1 mM EDTA and 5mM MgCl₂; Buffer B (protein buffer)-20 mM Tris-HCl pH 7.2, 1 M NaCl, 1 mM EDTA, 1 mM DTT, and 1 mM PMSF; RNA Assembly Buffer (RAB)-50 mM Tris–HCl pH 7.2, 50 mM NaCl, 10 mM KCl, and 5 mM MgCl₂; Virus Storage Buffer (VSB)-50 mM sodium acetate pH 4.5 and 8 mM magnesium acetate; TE Buffer-10 mM Tris-HCl pH 7.5 and 1mM EDTA; Disassembly Buffer (DAB)-50 mM Tris-HCl pH 7.5, 0.5 M CaCl₂, 1 mM EDTA, 1 mM DTT, 0.5 mM PMSF; Virus Running Buffer (VRB)-100 mM sodium acetate pH 5.5 and 1 mM EDTA; and TAE buffer-40 mM Tris-HCl pH 8, 20 mM acetic acid and 1 mM EDTA.

3.3.2 PolyU synthesis and fractionation

PolyU RNAs were prepared in Synthesis Buffer using the enzyme polynucleotide phosphorylase (Sigma Aldrich, isolated from Synechocystis sp) from a 20-nt-long polyU seed oligo with a 5’ cy5 fluorescent label[Vanzi 2003]. Briefly, 500 pmol of cy5-labeled seed oligo was mixed with 40 mM UDP and 1 unit of PNPase, followed by a 4-hr incubation at 37°C. These polymerization reactions produce a polydisperse mixture of fluorescent polyU RNAs ranging in length from 200-10000 nts, with shorter lengths being represented in significantly higher
copy number, as determined by fluorescence detection in denaturing 1% TAE agarose gel electrophoresis.

This polydisperse RNA mixture was run on a denaturing agarose gel alongside a ssRNA ladder and fractionated into discrete-length segments using the ssRNA ladder as reference. This fractionation process resulted in the following length fractions of polyU: 500-1500, 1500-2500, 2500-3500, 3500-5000, 5000-7000 and 7000-9000 nts (see Fig. S1). The amount of polyU purified was quantified by uv-vis absorbance measurements, with the absorbance ratio at 260nm/280nm used as a measure of RNA purity ($A_{260}/A_{280}$ greater than 2.0). These fractionated polyU samples were run on a denaturing agarose gel to demonstrate that single RNA bands were obtained.

### 3.3.3 In vitro transcription of normal-composition RNA molecules

Normal-composition RNAs were made by in vitro transcription of linearized DNA plasmids with T7 polymerase, rATP, rUTP, rCTP, and rGTP. This work utilized the following plasmids, produced and described in detail by Cadena-Nava et al.[Cadena-Nava 2012]: pT7B1 for BMV RNA1, pT7B1-0.5kbp for the 500 nt RNA, and pT7B1-2kbp for the 2000 nt RNA. Transcription reactions were carried out for 4 hr at 37°C, followed by digestion using DNase I (New England Biolabs) for 50 min at 37°C [Cadena-Nava 2012]. RNA was then purified from transcription mixes by washing through a 100-kDa Amicon filter with TE buffer. The concentration of RNA was determined by uv-vis, and the quality of the RNA produced was verified by native 1-percent agarose gel electrophoresis in TAE buffer.

### 3.3.4 Purification of CCMV CP

CCMV was purified from infected California cowpea plant (Vigna unguiculata cv Black eye pea) following the procedure described by Bancroft et al.[Bancroft 1967]. Virus was dialyzed into disassembly buffer, and CP was isolated as described by Annamalai and Rao[Padmanaban 2005]. An Applied Biosystems Voyager-DE-STR MALDI-TOF was used
to assess CP quality, in particular that the N-terminus was not cleaved during purification.

### 3.3.5 In vitro packaging of RNA by CCMV CP

RNA and CP were mixed in buffer B at a 6:1 mass ratio of CP:RNA, unless otherwise stated, to a final RNA concentration of 30 ng/µl, then dialyzed overnight into RAB at 4°C. The assembly mix was then dialyzed against VSB for 6 hr at 4°C, at which point the VLPs were collected for analysis.

### 3.3.6 Electrophoretic mobility shift assay

The products of assembly reactions were run on 1 percent native agarose gels in VRB. Generally, 5 µg of virus was loaded per lane, and run alongside the same mass of wt CCMV, as well as a dsDNA ladder control. Virions containing normal-composition RNAs were visualized using ethidium bromide, while polyU VLPs were visualized fluorescently using the cy5 fluorophore at their 5’ end. Virions were also stained using the protein stain Coomassie Blue. Gel bands were visualized on a Pharos FX Plus Molecular Imager, and the quantity of RNA in a particular lane was determined using ImageJ analysis.

### 3.3.7 Transmission electron microscopy

Negative-stain electron microscopy grids were prepared by depositing 6 µl of assembly reaction on glow-discharged copper grids (400 mesh), which were previously coated with Parlodion and carbon. After a 1-min deposition, grids were blotted and stained with 6 µl of 2 percent uranyl acetate for 1 min, then blotted dry. Grids were stored in a desiccator, and imaged on a FEI Tecnai TF20 electron microscope (Electron Imaging Center for Nanomachines, California Nanosystems Institute) at 50kX magnification. Particles were sized manually from images taken in negative stain TEM. At least 200 particles were analyzed per sample.
3.3.8 Competition experiments

The competition experiments follow the same basic protocol as the in vitro packaging experiments reported earlier [Comas-Garcia 2012], with a few changes. Equal masses of two RNAs were incubated in buffer B at a total RNA concentration of 30 ng/µl and a final volume of 200 µl, with only a 3:1 mass ratio of CP introduced into the system. In other words, there was the same total amount of RNA but half as much capsid protein as in a normal assembly reaction. This RNA-protein mixture was then dialyzed into RAB overnight at 4°C, followed by another 6-hr dialysis into VSB. Because of the lower (than 6:1) CP:RNA mass ratio involved, significantly less than all of the RNA was packaged.

3.3.9 RNase A digestion of virus-like particles

Virions were subjected to RNase digestion to confirm that they were fully-closed shells. Briefly, VLPs are incubated with RNase A for 50 min at 4°C at a ratio of 2.5 ng RNase A (PureLink RNase A, Thermo Fisher Scientific) per 1 µg VLP. This mild RNase treatment has been shown to have no effect on the wt virus, but is capable of digesting at least 10 µg of unpackaged RNA. After RNase treatment, the mixture was incubated with 40 Units of RNase inhibitor (Protector RNase Inhibitor, Sigma-Aldrich) for 15 min at 4°C, then washed through a 100kDa Amicon filter at least 3 times with TE buffer to remove the RNase and the RNase inhibitor.

3.4 Results

3.4.1 PolyU is completely packaged at the magic ratio

Purified polyU fractions (see Fig. S1) were subjected to the two-step assembly protocol that has been shown to package normal-composition RNAs ranging in length from 140 to 12,000 nts into CCMV VLPs. In this procedure we use a superstoichiometric 6:1 mass
ratio of CP to RNA, dubbed the "Magic Ratio," to provide the minimum amount of CP required to package completely the RNA\cite{Cadena-Nava2012}. It corresponds to an equality between the negative charges on the RNA phosphate backbone and the positive charges in the arginine-rich N-terminus of the CCMV CP.

The RNA and CP are first incubated in neutral-pH high-ionic-strength buffer (buffer B), in which the CP and RNA have little interaction because of the electrostatic screening. This mixture is then dialyzed into neutral-pH low-ionic-strength buffer (RAB), "turning on" RNA:CP interactions involving attraction between the RNA backbone and the CP N-terminus\cite{Bayer2005,Ni2012}. At this point in the assembly protocol it has been shown for branched RNAs that all of the CP is bound to the RNA in solution. It has also been shown that although all of the protein is bound to the RNA, regular capsid structures do not form under these conditions\cite{Garmann2014}. Virus-like particles are formed after a final dialysis into low-pH low-ionic-strength buffer (VSB), which significantly enhances the lateral attractions between CP subunits.

The assembly products were put on a native agarose gel to determine the extent of assembly; the DNA ladders and wild-type CCMV were visualized in red with ethidium bromide staining, while the cy5-labeled-polyU is visualized by its fluorescence, in green. Fig. 3.1 shows the resulting assembly products for B1 RNA and a variety of polyU lengths packaged at half the Magic Ratio (where explicitly stated) and at the Magic Ratio (when not explicitly specified). It is clear to see, for example, the effects of RNase treatment of assembly products shown in lanes 4 and 5 (unfractionated polyU), and in lanes 6 and 7 (B1 RNA and 2500-3500nt polyU), that self-assembly at half the Magic Ratio yields a large fraction of the RNA not being packaged into RNase-resistant capsids, but rather into a broad range of faster-moving RNA/protein complexes that are susceptible to nuclease digestion. In contrast to this, see lane 8 (1500-2500nt polyU), self-assembly at the Magic Ratio results in complete packaging of the RNA into fully-closed capsids that are RNase resistant.
Figure 3.1: Electrophoretic gel analysis of assembly products. Top: Lane 1, double-stranded (ds)DNA ladder; lane 2, wt CCMV; lane 3, 1500-2500 nt polyU VLPs; lane 4, 2500-3500 nt polyU VLPs; and lane 5, 3500-5000 nt polyU VLPs. The bands were visualized using two different fluorescent signals: cy5 attached to the polyU RNA, in green; and ethidium bromide in structured RNAs, in red. Bottom: Lane 1, double-stranded DNA ladder; lane 2, wt CCMV; lane 3, wt CCMV + RNase; lane 4, complexes of unfractionated polyU and CP at half the Magic Ratio (hMR); lane 5, unfractionated polyU complexes at half the Magic Ratio + RNase; lane 6, 3000 nt polyU and B1 RNA-protein-complexes at half the Magic Ratio; lane 7, 3000 nt polyU and B1 RNA-protein-complexes at half the Magic Ratio + RNase; and lane 8, 1500-2500 nt polyU VLPs at the Magic Ratio (MR) + RNase. Data from the ethidium bromide channel (left) and cy5 fluorophore channel (right) placed next to each other to create this lower gel image.
3.4.2 PolyU RNA is packaged into 22-nm capsids

**Figure 3.2:** Negative-stain EM images of polyU packaged by CCMV capsid protein and treated with RNase. Top row, from left to right: 1500-2500, 2500-3500 and 3500-5000 nt; bottom row, from left to right: 5000-7000, 7000-9000 nt and wt CCMV. All lengths of polyU assembled (ranging between 1500-9000 nt) have resulted in the formation of 22-nm diameter particles, despite the fact that wild-type virions are 28-nm in diameter. Scale bar is 50 nm.

The assembly products were visualized before and after RNase A digestion using negative-stain electron microscopy (Fig. 3.2). Analysis of the micrographs allows for size determination of the particles, measured by taking the geometric mean of their diameters along the horizontal and vertical axes of the image. Surprisingly, as shown in Fig. 3.3, all polyU VLPs have a characteristic size of 22-nm, smaller than the 28-nm wt particles of CCMV. Similarly small particles have previously been found for VLPs containing branched RNAs shorter...
than 2300 nt[Cadena-Nava 2012], whereas longer RNAs are packaged into 28-nm diameter capsids.

![Figure 3.3: Histogram of particle sizes formed from assemblies using various lengths of polyU as compared to the size of particles from wt CCMV virions purified from plants. All polyU VLPs were treated with RNase before imaging, although the plot for VLPs that have not been subject to RNase treatment is essentially identical.](image)

To quantify their asymmetry, the axial ratio of the VLPs as a function of length of polyU packaged was determined from negative-stain electron micrographs. Fig. 3.4 shows the axial ratios of VLPs containing various lengths of polyU in comparison to particles containing normal-composition RNAs, and at the right distributions of axial ratios for those same samples. The average axial ratios of the polyU VLPs are larger than the ratio determined for both wt virus and VLPs containing the viral RNA 1 (3000 and 3200 nt data points shown in red). Interestingly, VLPs containing 500 nt normal-composition RNAs, which were previously reported to be T=2 in size[Comas-Garcia 2014], have average axial ratios similar to those for VLPs formed around shorter polyU fractions. It is also clear that, for
all samples, only certain ranges of axial ratios are populated, indicating that the assembly reactions result in specific structures. Specifically, axial ratios between 1-1.1 (these are considered to be spherical given the fact that in negative-stain even wt CCMV exhibits some asymmetry) and 1.15-1.25 have majority populations, while axial ratios around 1.3, 1.4, and 1.5 show minority populations. VLPs containing 3000 nt polyU imaged using cryo-electron microscopy, in which the particles are suspended in vitreous buffer, show that the T=2-sized particles are indeed asymmetric in solution, and that this asymmetry does not occur due to adsorption onto the negative-stain EM grid. Additionally, the fact that discrete axial ratios are seen for all samples suggests that the particles are asymmetric (see DISCUSSION), and that this asymmetry is not a result of particles being distorted by adhesion to a substrate.

3.4.3 Multiplets form for longer lengths of polyU

While polyU is packaged into 22-nm sized particles regardless of its length, longer polyUs begin to be packaged into "multiplet" capsids in which RNA is shared between two or more capsids. Fig. 3.5 depicts the various types of capsids formed for polyU: a singlet has at least one RNA molecule in a single capsid (for RNAs shorter than 1000 nt, several RNAs are copackaged into the same capsid), a doublet contains RNA shared between two capsids, while triplets have RNA shared between 3 capsids. Similar multiplets have been found previously for branched RNAs that are significantly longer than the genome of CCMV (i.e. more than 4500 nt)[Cadena-Nava 2012].

As Fig. 3.6 shows (see Left), both normal-composition and polyU RNAs shorter than 4000 nt are packaged exclusively into singlet particles. RNAs with lengths between 4000 and 7000 nt are packaged predominantly into doublets; triplets begin to form at lengths above 7000 nt. Despite multiplet formation, all the capsids containing polyU are still the smaller 22-nm diameter capsids while the capsids containing normal-composition RNAs are preferentially 28-nm in diameter. It is seen further (see Right) that, after RNase digestion, the RNA shared between the capsids has been digested, resulting in only singlet VLPs. The right-
Figure 3.4: Left) Distributions of axial ratios for VLPs containing polyUs of various length, for normal-composition RNA VLPs (500 nt NC and B1), and for wt CCMV. For clarity, average distributions for short-polyU VLPs (polyU’s shorter than 5000 nt) and long-polyU VLPs (polyU’s longer than 5000 nt) are plotted, as the individual samples within each set exhibited similar distributions. Additionally, the axial ratios of long-polyU VLPs were inspected in cryo-electron microscopy to rule out the affect of drying the samples on EM grids. Right) Axial ratios of polyU VLPs, as a function of length, measured from negative-stain electron micrographs. The black data points represent the average axial ratio of polyU VLPs of a given length. Linear regression (black line) shows the increase in the axial ratio as a function of polyU length. The red data points represent the average axial ratio of several normal-composition RNA VLPs[Comas-Garcia 2014]. The normal-composition RNA data point at 3000 nt corresponds to the wt virus. Error bars indicate the standard deviation of the mean.
Figure 3.5: Representative negative-stain electron micrographs of singlet, doublet and triplet capsids formed by \textit{in vitro} self-assembly of 7000-9000 nt polyU. For this length of polyU, predominantly doublets are formed.

hand plot also shows that the apparent fraction of doublets seen at the left for short polyU lengths is due simply to the density of particles on the EM grid. The normal-composition RNA plots refer to data from significantly higher dilution, and therefore doublets are not seen before RNase digestion for RNAs shorter than 4000 nt, as particles are far less likely to be near one another under these conditions. PolyU VLPs imaged at higher dilutions show that multiplets disappear for shorter polyU assemblies, but that for longer polyU assemblies (greater than 5000 nt) the number of multiplets stays constant upon dilution.

3.4.4 PolyU outcompetes viral RNA for capsid protein

Competition experiments – self-assembly reactions in which equal masses of two different RNAs compete for CP that is sufficient to package only one of them – were carried out between viral RNA and polyU of the same length (2500-3500 nt). As these competition experiments are performed \textit{in vitro} they necessarily lack the host-cell machinery that could influence RNA packaging. Previous competitions between structured RNAs have shown that RNA1 of brome mosaic virus (BMV), a sibling virus of CCMV, outcompetes other RNAs (including RNA1 of CCMV) for viral CP in an \textit{in vitro} assembly reaction[Comas-Garcia 2012].
**Figure 3.6:** Left) Fractions of singlet (black), doublet (blue) and triplet (red) capsids as a function of increasing RNA length for polyU (squares) and for normal-composition RNAs (triangles). Note that the onsets of doublets (around 4000 nt) and triplets (around 7000 nt) occur at similar lengths for both polyU and NC-RNA assemblies. Right) Effect of RNase on the fractions of singlet (black), doublet (blue) and triplet (red) capsids formed from polyUs of increasing length. RNase digestion converts multiplets into singlet capsids, demonstrating that the polyU RNA is threaded between the capsids. The 20 percent baseline of doublets is very probably due to particle crowding upon drying on the EM grid, as demonstrated by imaging solutions at higher dilution, resulting in decrease in the number of doublets for VLPs containing short polyUs while doublets survive dilution for VLPs containing longer polyUs.
**Figure 3.7:** Negative-stain electron micrograph of the assembly products of the Both First competition reaction between 3000 nt polyU (red circles) and B1 RNA (blue circles) after RNase treatment. The smaller 22-nm diameter VLPs contain polyU, while the larger 28-nm diameter particles contain B1. Scale bar is 50 nm.
Figure 3.8: The size distribution of the VLPs formed from competitive self-assembly reactions after RNase treatment. The peak centered around 22 nm corresponds to polyU VLPs, while the peak centered at 28 nm represents B1 VLPs. The separate colors represent distinct self-assembly reactions, in which the order of mixing of the assembly components was altered. In the Both First reaction (black, left), the two molecules were mixed and CP was then added, while in the B1 First (blue, right) and polyU First (red, right) assemblies the respective RNA was first mixed with CP and allowed to equilibrate before the other RNA was introduced. The figure shows that polyU wins the competition, and that the order of mixing is important.
In addition, these experiments showed that the order in which components are mixed at neutral pH does not affect the outcome of the experiment. In other words, if RNA1 out-competes RNA2, it will do so even if RNA2 is first equilibrated with CP at neutral pH before the addition of RNA1. This result suggests that the battle for CP is fought at neutral pH, and that whichever RNA binds CP more readily at neutral pH will be packaged upon acidification.

Three distinct competitions were carried out between linear and branched RNAs of equal length: the viral RNA and polyU were incubated at neutral pH followed by the addition of CP (called Both First), viral RNA was first equilibrated with CP followed by the addition of polyU (B1 First), and polyU was first equilibrated with CP followed by addition of B1 (polyU First). It is important to note that all of these steps were taken before the pH was lowered; acidification results in complete capsid formation, which is an irreversible process.

The relative amounts of each RNA packaged were analyzed after RNase treatment using two methods. Negative-stain electron micrographs (Fig. 3.7) were taken of the mixture and the numbers of 22- and 28-nm diameter particles were compared. The relative numbers of each type of RNA packaged could then be determined under the assumption that polyU was packaged in the smaller capsids and B1 in the larger. The amounts of each RNA packaged were also determined by gel electrophoresis of the assembly mixtures alongside calibration curves of viral RNA VLPs and fluorescently-labeled polyU VLPs. The molecules can be distinguished because the polyU is labeled with cy5 and the viral RNA is stained by ethidium bromide. The intensity attributed to each RNA molecule was then compared to its respective calibration curve, allowing a quantitative method of determining the amount of each type of VLP formed (data not shown).

The results of the EM analysis are shown in Fig. 3.8. Surprisingly, polyU outcompetes the viral RNA in head-to-head competition for CP. Additionally, unlike the case of the competition between B1 and other branched RNAs, the order in which the polyU and B1 RNAs are incubated with CP has a significant effect on the outcome of the competition.
Neither of these results was expected.

Table 3.1 shows the percentages of 22- and 28-nm diameter capsids produced for each competition reaction, as analyzed by negative-stain EM or by fluorescence gel electrophoresis (in parentheses). The table shows good agreement between the two forms of analysis, and that polyU outcompetes B1 RNA for CP. Additionally, the data show that the order in which the RNAs sees the CP alters the results of the competition experiment: if polyU is equilibrated with CP first, it wins the competition outright; if B1 is equilibrated with CP first, B1 is preferentially packaged; and if both RNAs interact with CP simultaneously, polyU is preferentially packaged. This suggests that polyU does not bind CP reversibly, as the order of mixing changes the outcome of the competition. In contrast, for normal-composition RNAs the result obtained for a competition between two RNAs was independent of the order of mixing.

<table>
<thead>
<tr>
<th>Particle Diameter (nm)</th>
<th>Both First</th>
<th>polyU First</th>
<th>B1 First</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>78 (76)</td>
<td>99 (91)</td>
<td>35 (30)</td>
</tr>
<tr>
<td>28</td>
<td>22 (24)</td>
<td>1 (9)</td>
<td>65 (70)</td>
</tr>
</tbody>
</table>

Table 3.1: EM and agarose-gel fluorescence intensity analyses (in parenthesis) showing the percentages of polyU and B1 VLPs formed during the competition reactions.

3.5 Discussion

As mentioned in the INTRODUCTION, it is significant that while RNA molecules of the same nucleotide length and composition have comparable amounts of intramolecular self-complementarity (Watson-Crick base-pairing) non-viral sequences result in significantly less compact 3D structures. This is because even though secondary structure formation leads in all cases to an effective branching of RNA, associated with the presence of many single-stranded loops from which three or more duplexes emanate, viral sequences involve more "branch points" of this kind as well as a concentration of them in the center of the overall
secondary structure rather than on its periphery (where they do not contribute much to the "gathering-in" – compaction – of the molecule). In addition, recent theoretical work has shown that this branching and compactness of RNA results in a more efficient binding of capsid protein [Singaram 2015].

Accordingly, in choosing to work with polyU – for which neither self-complementarity nor base stacking plays a role – we expected that it would be both less compact than viral RNA of the same length and less efficiently packaged. Indeed DLS measurements (See Fig. S2) indicate that in assembly buffer at 25°C in the absence of CP, polyU molecules are about 25% larger in average diameter than branched RNAs of the same length. And yet we find that polyU outcompetes viral RNA for binding capsid protein and for being packaged into RNase-resistant virus-like particles. Further, independent of its length, even when it is three times longer than viral RNA, the VLPs formed are always of a size (22 nm) typical of T=2, rather than the wt 28-nm T=3 nucleocapsids formed in the case of viral (or normal-composition) RNAs.

This work therefore raises many more questions than it answers. Basically, there are two sets of issues. One involves the binding of RNA by capsid protein with subsequent formation of a VLP, and its dependence on RNA secondary structure (or the lack thereof). The other involves the spontaneous selection of a preferred size of capsid, and its dependence on RNA secondary structure and on overall length/size. These two issues are of course interconnected, in particular because the preferred capsid size depends not just on the preferred curvature of protein-protein interactions, but also on the sequence-dependent local secondary structure of the RNA molecule to which the capsid protein binds. One way to disentangle these effects is to keep the secondary structure fixed and change only the overall length/size of the RNA, and this can be done most simply by working with synthetic linear anionic polymers (e.g., PSS) or with RNA that has no secondary structure (e.g., polyU).

In the case of PSS, earlier work with in vitro self-assembly with CCMV capsid protein has shown that the spontaneously-formed VLPs change from T=1-size to wild-type T=3
nucleocapsids as the size of PSS is increased from hundreds of kDa to several MDa [Hu 2008, Cadena-Nava 2011]. This result was understood in terms of their being three curvatures that are intrinsically preferred by interacting capsid protein, corresponding to T=1, T=2 and T=3 sizes, with the final result being determined by the size of the polyanionic “template” on which the protein binds.

A similar result was found for normal-composition RNA: when the length of RNA was increased from hundreds to thousands of nts, the dominant VLP product switched from T=2 to T=3 sizes [Cadena-Nava 2012]. But, as found in the present work, in the case of polyU the only VLP formed is of T=2 size, even when the polyanion length is increased by a factor of ten, up to 10,000nt. In both (PSS and normal-composition RNA) cases, the polyanion is shared by two or more T=3 capsids when it becomes sufficiently long; with polyU the increasingly long polyanion is shared by two or more T=2 capsids.

But the increase in size of the polyU could also be accommodated by a switch from a T=2 size to T=3. Why does this not happen? It is here that the absence of secondary structure in polyU likely plays a role. Structural determinations of T=3 capsids show that the protein residues are essentially coplanar, with the 180° dihedral angle between the residues in the hexamer the result of double-stranded RNA that runs beneath it, a mechanism clearly not operative for the necessarily single-stranded polyU [Fox 1998, Speir 1995].

In this context it is important to mention the example of MS2 capsids for which it has been shown that interaction of the capsid protein with specific local components of RNA secondary structure – the ”packaging sequences/signals” – are essential for the formation of appropriate size (T=3) shells for efficient packaging of the viral genome [Basnak 2010]. In fact, Krol et al. [Krol 1999] showed that, in the presence of replication-incompetent genomic RNAs, BMV capsid protein forms T=2 particles in vivo. However, these particles were polydisperse and only the most spherical particles were chosen for the structure determination. In reviewing several previous works which involved the formation of T=2-sized particles made from CCMV or BMV capsid proteins, it appears that all of these assemblies
have resulted in heterogeneous particles\cite{Krol 1999, Cadena-Nava 2011, Comas-Garcia 2014, Hu 2008, Tresset 2014, Tang 2006}. A mutated CCMV CP lacking most of its cationic N-terminal domain has been shown to produce, \textit{in vitro}, heterogenous, T=2-sized particles, in addition to producing T=1 and T=3 particles. Additionally, Tresset et al.\cite{Tresset 2017} found that polydisperse, T=2-sized, VLPs containing PSS were less thermally stable than wt CCMV.

Small non-isometric closed capsids have been described in two theoretical papers treating polymorphism of empty capsids formed from capsid protein alone\cite{Nguyen 2009, Wagner 2015}. They both find a family of closed-shell structures based on icosahedral caps of pentamers separated by belts of hexamers. Similar structures have been inferred for the \textit{in vivo} virions of Alfalfa Mosaic Virus (AMV)\cite{Cusack 1983} and Flock House Virus (FHV)\cite{Dong 1998}. In both these cases the virus particles contain RNAs shorter than the wild-type genome and have characteristic sizes that lie between T=1 and T=3 structures. The aspect ratios observed in electron micrographs for FHV are similar to those expected for one or two belts of hexamers. More explicitly, Dong, et al.\cite{Dong 1998} have made estimates of the maximum number of nt that could be packaged within different polymorphs. They assume that one nt occupies 0.655 nm$^3$ and calculate that a T=1 FHV particle could accommodate 600 nt and that bacilliform particles with one belt and two belts of hexamers could accommodate 1740 and 2700 nt, respectively. From sucrose gradient sedimentation studies, however, Martin and Ames inferred a value of 0.57 cm$^3$/g for the partial specific volume of poly U\cite{Martin 1962}; this corresponds to a nucleotide volume of 0.31 nm$^3$, about half that of the volume used in the computation in the Dong paper, and suggests that single-belt and two-belt capsids could accommodate about 3,500 and 5,400 nt of polyU. These values are consistent with the onset of multiplet structures, the point at which the capacity of a single capsid is exceeded. Additionally, axial ratios of polyU VLPs are discretized, suggesting "hexamer-belted" structures similar to those shown for AMV and FHV.

This polymorphism suggests that T=2 sized particles have a high probability of form-
ing kinetically trapped, non-icosahedral, structures. In fact, it has been suggested that T=2 particles form following a kinetically trapped pathway en route to T=3 [Wang 2015, Zlotnick 2000]. The idea that kinetics play a role in the assembly of T=2 sized polyU VLPs is further supported by the results of the competition experiments between polyU and BMV RNA1, wherein the order of mixing the RNAs resulted in different amounts of each RNA encapsidated. This is unique to polyU RNA, in contradiction to competitions between normal-composition RNAs where the order of mixing is irrelevant [Comas-Garcia 2012]. We have not investigated whether CP prefers to bind polyU or B1 at neutral pH, but expect that polyU is preferred, as previous work has shown that the RNA that binds CP more readily at neutral pH is the one that is subsequently packaged upon acidification [Comas-Garcia 2012, Garmann 2014].

This work highlights the importance of RNA secondary structure in capsid formation, as it affects both the ability of an RNA to bind capsid protein as well as the size and structure of the capsid formed. While polyU both outcompetes viral RNA for capsid protein and is packaged into RNase-resistant nucleocapsids, the capsids formed appear to involve kinetically trapped assembly pathways, resulting in structural heterogeneity. It is also clear that models of capsid assembly need to be refined to include the effect of secondary structure on capsid assembly, specifically addressing the impact of particular secondary structures on resultant capsid morphology.

### 3.6 Acknowledgments

We thank Dr. Feng Guo for providing the T7 RNA polymerase, Dr. A.L.N. Rao for providing the cDNA used to generate BMV RNA1, and Dr. Hong Zhou, Dr. Ivo Atanasov and Dr. Yanxiang Cui for helpful discussions related to electron microscopy. Electron microscope images were acquired at the California NanoSystems Institute (CNSI) Electron Imaging Center for NanoMachines. This research was supported by grant CHE1051507 from the
NSF. KNL was supported by the Amgen Scholars Program over the course of this work. None of the authors has a conflict of interest.

### 3.7 Supplementary materials

![Electrophoretic gel analysis of fractionated polyU.](image)

**Figure 3.9:** Electrophoretic gel analysis of fractionated polyU. From left to right: double-stranded (ds)DNA ladder, 5000-7000, 4000-5000, 2500-4000, 1500-2500 and 500-1500 nt polyU RNAs. The decrease in band intensity for higher molecular weight polyU molecules is due to the fact that an equal mass of polyU was loaded per lane, resulting in many fewer RNAs in the higher molecular weight fractions and, subsequently, a decreased fluorescence signal.
Figure 3.10: Left) Raw DLS data of polyU RNAs of varying length and BMV RNA1. The fraction of molecules of a given hydrodynamic diameter (nm) is plotted for each sample. The distributions are broad, yet the average size is consistent. Right) Plot showing how the hydrodynamic radius of polyU and normal-composition (NC) RNAs varies with nt length. Theoretical points show $N^{1/2}$ scaling. Note that although the raw data exhibits a large standard deviation, the standard deviation of the mean (error bars) is relatively small, indicating that the average size is reproducible across measurements.
Chapter 4

Visualizing ssRNA inside viral capsids using high resolution cryo-electron microscopy

4.1 Abstract

We report the first asymmetric reconstruction of the single-stranded (ss) RNA content of a multipartite virus, i.e., one whose genome consists of two or more molecules that are contained in two or more otherwise-identical protein shells (“capsids”). The best-studied example of such a virus is Brome Mosaic Virus (BMV), whose genome is made up of 3 RNA molecules, plus a fourth subgenomic RNA that is packaged, distributed in three separate spherical particles – one containing RNA1 (3200nt), a second containing RNA2 (2800nt), and a third containing one each of RNA3 (2100nt) plus the subgenomic RNA4 (700nt). Because the wild-type virus involves a roughly uniform mix of these particles it is difficult to fractionate them beyond imperfect separation into RNA1-containing particles (3200nt) and a mix of RNA2- and RNA3+4-containing particles (2800nt). In our work we exploit a sample of “pure” BMV particles containing only RNA3+4, produced \textit{in planta} by agrobacterium
transformation. We find that the interior of the particle is nearly empty with most of the RNA genome situated at the capsid shell. But this density is flexible and disordered, suggesting that the RNA inside is associated with an ensemble of secondary/tertiary structures that interact with the capsid.

4.2 Introduction

BMV is a spherical, positive-sense, single-stranded (ss)RNA virus with a multipartite genome composed of 3 molecules, plus a fourth subgenomic RNA that is packaged, split into three capsids. Specifically, wild-type (wt) BMV is a mix of virions containing either one copy of 3.2 kilo-nucleotide (knt) RNA1, or one copy of 2.9 knt RNA2, or one copy each of 2.1 knt RNA3 and 0.9 knt subgenomic-RNA4; accordingly, each virion contains about 3 knt per particle[Lane 1974]. The three distinct wt virus particles are indistinguishable when imaged with a transmission electron microscope (TEM)[Fox 1998] and separation techniques (e.g. sedimentation) have not been successful in producing pure single-RNA-containing fractions[Hoover 2016].

The BMV capsid is composed of 180 identical capsid proteins (CPs) arranged into a 140-Å-radius icosahedrally symmetric T=3 Caspar-Klug capsid composed of 12 pentamers and 20 hexamers[Caspar 1962, Lucas 2002]. The structure of the BMV capsid has been solved to near-atomic resolution using both x-ray crystallography[Lucas 2002] and cryo-electron microscopy (cryoEM)[Wang 2014]. These high-resolution structures have benefited from the icosahedral symmetry present in the capsid, but have necessarily obscured any features lacking icosahedral symmetry, particularly the internal genetic material.

A common RNA binding motif used by ssRNA viruses is the Arginine-Rich-Motif (ARM), a relatively unstructured, positively-charged amino acid sequence found at the N-terminus of the CP that has been shown to extend into the interior of the virion to bind to the RNA. BMV CP has been shown to interact nonspecifically with the RNA genome through
electrostatic attractions between its positively-charged N-terminus—specifically, there are 8 positive residues at the CP N-terminus—and the negatively charged phosphate backbone of the RNA[8]. In fact, CCMV, a sibling virus to BMV, has been shown to package a wide variety of molecules including polystyrene sulfonate[9], polyU[10] and heterologous RNA with normal-composition (i.e. roughly equal numbers of the A, U, G, C nucleobases) independent of length or sequence [11].

While BMV has not been shown to require specific RNA packaging signals for assembly [12], recent experiments [13, 14, 15] indicate that the secondary structure of the encapsidated RNA plays a key role in BMV assembly, highlighting the importance of learning more about the interactions between the genome and capsid proteins. One such experiment has shown that a 187 nt sequence in the nonstructural movement protein gene in BMV RNA2 is required for its specific packaging in vivo[16], while a similar experiment has shown that the TLS (the 3’ end sequence of each of the BMV RNAs) is required for specific packaging in vivo[17, 18].

Recent work by Hoover et al.[19] investigated virions that were partially separated on a cesium chloride gradient into two fractions, one rich in virions containing 3200 nt-RNA1 (“BMV1”) and another rich in virions containing either 2800 nt RNA2 or 2100 nt-RNA3 plus 700nt-RNA4 (“BMV2.3+4”): they found that the RNA organization at the capsid inner surface differs between the two fractions. RNA1 virions have the RNA arranged in a dodecahedral cage, sitting under the hexamers and forming rings around the pentamers. In contrast, BMV2.3+4 virions have a thicker ring of RNA density that is also more amorphous in structure and generally situated under the 2-fold axes of the capsid. This gives the appearance of RNA density being organized into rings around both the hexamer and pentamer subunits. In both virion fractions they also find strong density associated with the N-termini of the CP, but are unable to distinguish whether that density represents RNA or
These earlier investigations have provided important indications of the RNA structure in multipartite viral capsids, but they are limited by working with samples containing multiple RNAs and by the imposition of icosahedral symmetry, which prevents the detection of RNA density that does not conform to the capsid’s icosahedral symmetry. We can, for the first time, examine samples prepared \textit{in planta} consisting of virions with a single RNA species—either RNA1, RNA2 or RNA3+4. Additionally, recent developments in cryoEM, particularly the development of direct electron detecting cameras, greatly improve the overall quality of the images taken, insofar as fewer electrons can be detected allowing for more data to be acquired before the sample is damaged by the incident electron beam. Coupled with significant improvements in computing power, improved images allow high-resolution structures to be solved for samples without assuming symmetry during the image analysis, such as the recently solved structures of the monopartite bacteriophages MS2 and Q\textbeta[20] [Zhang 2015, Koning 2016, Gorzelnik 2016, Dai 2017].

This work investigates the structure of RNA inside a multipartite virus using BMV3+4 virions and asymmetric reconstruction. Utilizing a pure–unique RNA content–BMV particle (BMV3+4) produced \textit{in planta} and asymmetric reconstruction, we avoid both averaging over different RNA contents and averaging out asymmetric portions of the virion, in principle, allowing for visualization of the organization of the structured portions of the RNA. We chose to work with BMV3+4—instead of BMV1 or BMV2—samples because BMV3+4 can be produced in higher quantities \textit{in planta}.

We find in BMV that the RNA is only organized at the capsid surface and that there is not a single dominant conformation for the RNA inside the virus, but rather an ensemble of RNA structures that interact strongly with the capsid protein. In contrast, in MS2 a single RNA secondary/tertiary structure is resolved to very high-resolution[Dai 2017]. Our results indicate a fundamental difference between the ssRNA organization in the multipartite, symmetric BMV viral capsid and the ssRNA organization in the monopartite, asymmetric
capsids of the bacteriophages MS2 and Qβ.

We have also analyzed the data set by subtracting out the capsid protein signal from the raw data, in an attempt to extract more of the asymmetric RNA genome[Zhang 2015, Liu 2015]. These methods allow for more of the RNA genome to be visualized at low-resolution, but a high-resolution structure of the RNA genome was not able to be determined. These findings demonstrate that while some viruses rely on specific protein-RNA interactions to drive virion assembly, other viruses, such as BMV, utilize nonspecific electrostatic interactions between CP and RNA to mediate virion assembly.

4.3 Results

4.3.1 In vivo synthesis of pure BMV3+4

Genome packaging in BMV is functionally coupled to replication, i.e., only the replication-derived progeny RNA is packaged into virions[Annamalai 2005, Annamalai 2006b, Seo 2012]. More explicitly, the interaction between CP and replicase protein p2a has been shown to dictate packaging specificity. With these essential requirements in mind, an agrobacterial transformation strategy has been developed to separately assemble virions packaging either RNA1 (BMV1 virions) or RNA2 (BMV2 virions) or RNA3 and 4 (BMV3+4 virions)[Annamalai 2006a].

These particles are then purified from plants following protocols outlined by Padmanaban and Rao[Padmanaban 2005], and their purity is determined by both gel electrophoresis and negative-stain electron microscopy. Additionally, the purity of their RNA cargo is determined by extracting the RNA from virions, and detecting the RNA species by northern blot (Fig 4.5, SUPPORTING MATERIALS).
Figure 4.1: Top Left) Exterior of the 3.1Å-resolution symmetric reconstruction of BMV3+4, highlighting the highly symmetric capsid, in close agreement with previous structures determined for wild-type BMV[Lucas 2002, Wang 2014]. Top Right) Exterior of the 3.9Å-resolution asymmetric reconstruction of BMV3+4, highlighting the similarity between the symmetric and asymmetric structures. The particles are colored radially, with the red density being closest to the center of the particle and the blue density being furthest away. The red and orange density is attributable to the RNA, while the blue and green density is attributable to the capsid protein. The yellow density is primarily attributed to the RNA, but probably contains a mixture of RNA and protein. Bottom Left) Fourier shell correlation (FSC) of the 3.1Å-resolution symmetric reconstruction shown above. Bottom Right) Fourier shell correlation (FSC) of the 3.9Å-resolution asymmetric reconstruction shown above. Note that the structures reach atomic-resolution.
4.3.2 Symmetric and asymmetric reconstruction

We obtained 3150 images, containing 79,140 particles, and have obtained high-resolution symmetric and asymmetric reconstructions of the data set. The top left image of Fig 4.6 (SUPPORTING MATERIALS) shows BMV3+4 particles obtained on the FEI Titan Krios with a K2 camera; while at the top right, the fourier transform of the same image is shown, illustrating the quality of the data, as the 3.8Å resolution ring associated with water is easily visible.

The symmetric reconstruction is at 3.1Å resolution (Fig 4.1), while the asymmetric reconstruction is at 3.9Å resolution. The particles have been colored radially from the center, with the red and orange density, which is internal to the capsid, primarily attributable to a mixture of the unstructured capsid protein N-termini and the ssRNA genome and the green, cyan and blue density attributable to the structured portion of the capsid protein (see Fig 4.3).

The asymmetric reconstruction is nearly identical to the symmetric reconstruction, with both structures lacking features at the interior of the capsid. This similarity is attributed to the dominance of the CP signal in the reconstruction, which obscures the signal of the genome. Both capsid structures agree well with those determined previously using cryoEM symmetric reconstruction[Wang 2014] and x-ray crystallography[Lucas 2002]. Our structure has an average radius of 137 Å, in good agreement with the structure solved previously by cryoEM. The structure determined by crystallography is significantly smaller, with an average radius of 113 Å, possibly due to the low pH buffer conditions used during the crystallization or due to effects on the capsid structure during crystallization.

An atomic model of the asymmetric unit of the capsid (Fig 4.2) was generated from the asymmetric structure shown at the top right of Fig 4.1, and this model represents the first atomic model of the BMV CP generated using asymmetric reconstruction. The model illustrates in atomic detail the organization of the three (A,B and C) conformations of the CP in the T=3 icosahedral capsid; the asymmetric unit is very similar to previous structures.
Figure 4.2: Top Left) 2.9Å atomic model of the BMV3+4 virion, with the asymmetric unit of the capsid protein highlighted in color. Bottom Left) Atomic model of the asymmetric unit of the capsid protein, with the A (red), B (blue), and C(green) conformers indicated. The A and B conformers are nearly identical and make up the hexamers of the capsid, while the C conformer is distinct and makes up the pentamers. Right) Zoom in on the N-terminus of the A subunit (Top), C-terminus of the A subunit (Middle) and C-terminus of the C subunit (Bottom). The atomic model is shown, with a mesh overlay representing the electron density determined via reconstruction. This figure illustrates the high quality of this data set, as individual amino acids are easily resolved.
of BMV that were determined by symmetric reconstruction[Lucas 2002, Wang 2014]. The atomic model notably excludes any density internal to the capsid, and an overlay of this model onto our symmetric structure with internal density illustrates that the red, internal density is probably a mixture of the disordered CP N-termini and RNA.

Fig 4.3 shows a hint of the ssRNA genome, with some internal density surviving icosahedral averaging during the structure determination, interacting with the N-termini of the hexameric subunits of the capsid. The RNA density resolved in Fig 4.3 agrees well with previous work suggesting that the RNA in BMV3+4 is found predominantly at the hexamers and forms rings around the pentamers[Speir 1995, Hoover 2016]. In fact, zooming in on the structure hints that the N-termini of the CP–the yellow-colored portions of the capsid protein that extend into the capsid interior at the hexamers–are binding to the RNA–the red and orange colored density–and ordering it between the two hexameric subunits. This is illustrated most clearly in the image at the bottom-right of Fig 4.3, in which the density threshold has been lowered to allow more of the internal signal to be shown.

Unfortunately this density cannot be ascribed to a particular structure (e.g., a dsRNA hairpin-loop), but is rather most probably a mixture of capsid protein N-termini and RNA that is consistently present at this location. In this context, the HIV-1 A-rich hairpin loop[Puglisi 1998] was used as a model 23 nt dsRNA, and an attempt to fit it into our structure made clear that the density in our structure is significantly flatter than dsRNA. It is interesting, however, that we see strong density, attributed to the RNA genome, at these locations on the capsid and, the fact that this density lacks a particular structure implies that a specific RNA-protein motif is not present, but rather that nonspecific interactions between CP and RNA are consistently found at this position inside the capsid.

### 4.3.3 Capsid subtraction and asymmetric reconstruction

While the symmetric structure (Fig 4.3) shows fragments of the internal RNA genome that survive icosahedral averaging, asymmetric reconstruction can allow for visualization of nearly
Figure 4.3: Interior views of the 3.1Å resolution symmetric reconstruction of BMV3+4 highlighting the RNA density that survives icosahedral averaging. The RNA density is seen interacting with the N-termini of the hexameric proteins, which agrees with previous structures determined with symmetric cryoEM reconstruction\cite{Wang2014, Hoover2016} and x-ray crystallography\cite{Lucas2002}. The red and orange density is attributable to a mixture of the RNA and the unstructured residues at the N-terminus of the CP. The yellow density is attributed to the structured portion of the capsid protein N-terminus, which extends into the interior of the particle. The rest of the capsid protein has been colored gray.
the entire RNA genome\cite{Koning2016, Gorzelnik2016, Dai2017}. However, the asymmetric structure in Fig 4.1 shows that simply using the raw data is not sufficient for visualizing the asymmetric genome in BMV because the symmetrical capsid dominates the signal of the asymmetrical RNA, resulting in a complete loss of the RNA signal.

Therefore, we subtract the symmetric capsid signal from the raw data resulting in a new data set that contains both the RNA density and any internal capsid protein density (i.e. the disordered N-terminus) that was not resolved in the symmetric reconstruction\cite{Zhang2015, Liu2015}. This subtracted data set is then used to determine a new RNA-dominant structure, in which the orientation of each particle has been determined by the asymmetric RNA. However, BMV has a symmetric T=3 capsid and therefore, there are 60 distinct orientations for each particle, which makes aligning the RNA-structures difficult.

To work around this, we have taken 60 copies of each RNA particle, and sorted them into 60 3D classes using RELION\cite{Scheres2012}. We then chose representative classes with fewer than 1/60 of the total number of particles, as classes with more than this number have been overpopulated by copies. Finally, we removed any duplicate particles that have been sorted into the representative classes to ensure that each particle is only utilized once in the sorting. Once the data set was classified, the orientations of these RNA-dominant structures were used to build a full capsid model from the original, raw data (Fig 4.4). For a more detailed explanation of this procedure, and to see more representative 3D structures, see the MATERIALS AND METHODS section and the SUPPORTING MATERIALS.

Using this approach, we found many similar but distinct 3D structures; Fig 4.4 shows four representative 3D structures, all of which have an identical capsid structure with minor variations in the RNA density. The structures shown in Fig 4.4 represent roughly 5000 particles (6\% of the data set). Notably, the capsid is resolved to high resolution, yet the RNA genome shows little density and has been low-pass filtered to 15Å resolution for visualization.

While this procedure is technically complicated, these models allow us to visualize how the RNA genome is interacting with the capsid protein shell. The fact that this procedure
generates many classes based on the internal RNA organization is already interesting, as it illustrates that many RNA secondary structures are present in BMV3+4 virions. Furthermore, the low resolution to which these structures are determined mean that each of these structures itself represents an average of many RNA secondary structures.

We are able to resolve particular RNA structural motifs, such as a long dsRNA duplex that extends across the interior of the dense RNA shell in Fig 4.4. In fact, this procedure generates many distinct classes, each corresponding to a unique set of RNA secondary structures. Yet there are similarities in the structure of the RNA across many classes. For example, we see that there is a strong association of the RNA with the N-termini of the CP, with much of the RNA density being organized into an asymmetric ring at the capsid surface, and many of the classes exhibit the dsRNA duplex discussed above. Additionally, each of the classes shown in Fig 4.4 has several other classes, that are not shown, generated by 3D classification in RELION with very similar RNA secondary structures.

Comparison of the asymmetric reconstruction (bottom right, Fig 4.1) and the asymmetric reconstructions determined using capsid-subtraction (Fig 4.4) shows the advantage of subtracting the capsid signal before doing the asymmetric reconstruction, namely that the signal of the asymmetric RNA genome survives the structure determination.

While the resolution of the RNA is very low, we are seeing for the first time the organization of the ssRNA genome inside a multipartite, symmetric virus particle with pure RNA content. We find that the RNA is organized strongly by the CP at the capsid surface, and our results indicate that there is not a single dominant conformation for the RNA but rather an ensemble of structures for which the RNA interacts strongly with the capsid protein.
Top Left) High-resolution exterior of the capsid determined by asymmetric reconstruction using the capsid-subtracted data set. The capsid is resolved at 4 Å resolution, illustrating that the asymmetric reconstruction is capable of resolving high-resolution information. It has been colored radially, with the red density being internal RNA, the yellow and green densities contain a mixture of RNA and protein, and the blue and gray density is the capsid protein.
Figure 4.4 *(previous page)*: Top Right) Interior of the capsid, left, determined by asymmetric reconstruction using the capsid-subtracted data set. This structure has been low-pass filtered to 15Å for better resolution of the internal RNA. It has been colored as described below. All others) Interior views of representative asymmetric reconstructions of BMV3+4 illustrating how the RNA interacts with the capsid protein. The RNA density is seen interacting with the N-termini of the hexameric proteins, which agrees both with our symmetric reconstruction and with previous structures determined with symmetric cryo-EM reconstruction[1] and x-ray crystallography[2]. The particles are colored radially, with the red density being internal RNA, the orange density contains a mixture of RNA and disordered CP N-termini, the yellow density is a mixture of the disordered N-termini and the ordered CP, and the remaining capsid protein has been colored gray.

4.4 Discussion

We show that many RNA conformations exist within BMV virions, all of which exhibit a shell of RNA interacting with the capsid, yet each conformation has distinct RNA secondary/tertiary structures at the interior. In fact, we find that if we allow the data set to be sorted into 60 3D structures then 60, roughly equally populated, structures will be generated. If that number is increased to 100, then 100 roughly equally populated structures are built. This suggests that the genome of BMV is organized into many different conformations, and that the structures generated from the 3D classification are average structures of similar RNA conformations.

Our symmetric reconstruction is the highest-resolution (3.1 Å) structure of BMV to date, and is unique in that it was solved at neutral pH using a pure BMV virion (BMV3+4). We determined, using asymmetric reconstruction, an atomic model for the asymmetric unit of
the capsid, and find that it agrees well with previous atomic models derived from symmetric reconstruction. Interestingly, our symmetric and asymmetric reconstructions are nearly identical, showing that the capsid shell is truly highly-symmetric even when the particle symmetry is left unconstrained during reconstruction.

Recent work investigating important capsid protein-RNA interactions in BMV has shown that these interactions depend on both the RNA content in the virion and on the host in which the virus is generated [Ni 2014, Vaughan 2014, Hoover 2016]. These studies use Clip-Seq analysis in which the binding footprints of CP for RNA, and vice-versa, are investigated by cross-linking capsids with different RNA cargo. It is found that over 60% of the CP of BMV2.34 particles purified from N. Benthamiana, the sample studied that is most similar to the BMV3+4 studied in this work, interacts with the RNA genome. In particular, they find that residues 1-64, 90-103, 131-142 and 166-189 interact strongly with the genome. Similarly, in BMV1 virions residues 1-103 and 131-142 were found to interact with the RNA genome. We find good agreement between our structural determinations of the locations of CP-RNA interaction and those found using Clip-Seq, and both studies illustrate that the CP of BMV is highly nonspecific, with much of the RNA and CP interacting with one-another.

We find that most of the RNA density is organized into a shell at the capsid surface, indicating that much of the RNA is interacting with the capsid protein. Additionally, we see that many RNA conformations exist across different particles, indicating that many different portions of the RNA genome are interacting with the capsid protein in a variety of ways. These results suggest that BMV packaging is highly nonspecific, a result that agrees well with in vitro packaging experiments confirming that BMV CP will package any viral-length RNA regardless of sequence.

Our results indicate that the multipartite, plant virus BMV differs dramatically from the monopartite, bacteriophage MS2 in the way that their genomes are packaged into virions. Structurally these viruses are very similar, with both viruses having 30 nm diameter, T=3, icosahedral capsids with around 3000 nt of ssRNA packaged inside; yet these viruses have
been shown to have very different self-assembly pathways. In MS2 it has been shown that a single RNA secondary structure is present in each MS2 particle and that the RNA density is conserved even at the center of the particle[Koning 2016, Dai 2017], while in BMV we do not resolve a high-resolution RNA secondary structure and we only see RNA density at the capsid surface.

While the MS2 capsid is structurally similar to BMV, in contrast to BMV, it has a monopartite genome consisting of only a single RNA molecule[Fiers 1976]. Additionally, the capsid of MS2 is not perfectly symmetrical as one of the capsid protein dimers that comprise the capsid has been replaced with the maturation protein (also called the A protein)[Dent 2013]. This protein is known to interact strongly with both the 5’ and 3’ ends of the RNA genome, imparting an asymmetry—and strong constraints—on the internal genomes configuration[Toropova 2011, Bakker 2012]. The unique asymmetry of these capsids also serves to reduce the possible orientations of a T=3 capsid from 60, the number of quasi-equivalent units in these structures, to 1.

Additionally, MS2 is known to have strong, highly-specific contacts between the CP and the RNA genome, called packaging signals[Rolfsson 2008, Toropova 2008, Rolfsson 2010], which dictate that the RNA be structured so as to make those specific interactions[Dykeman 2011, Dykeman 2013]. BMV CP, on the other hand, is known to interact with RNA through the electrostatic attractions between the positively-charged N-termini of the CP and the negatively-charged phosphate backbone of the RNA[Bayer 2005]. These interactions have been shown to be nonspecific in nature, and in fact, CCMV CP is capable of packaging any normal-composition, viral-length RNA in vitro independent of its sequence[Cadena-Nava 2012].

The last major difference between MS2 and BMV is in the way that the CPs of MS2 and BMV interact with their internal genomes. The contacts between MS2 and its RNA-genome are made directly at the internal surface of the capsid, again requiring the RNA to organize itself in accord with the structure of the capsid[Koning 2016, Dai 2017]. BMV CP, in contradis-
tinction, interacts with its genome through disordered N-termini that extend into the internal portion of the virion, allowing the RNA more freedom in its organization [Lucas 2002]. All of these differences act in concert, resulting in an RNA genome that is highly ordered particle-to-particle in MS2, and an RNA genome that is disordered in BMV.

A comparison of the asymmetric reconstructions of BMV and MS2 illustrates that ss-RNA viruses can exhibit disordered and ordered genomes, respectively. This result provides further evidence that these viruses assemble differently, with BMV utilizing nonspecific electrostatic interactions to drive assembly and MS2 using highly-specific CP-RNA contacts. We hypothesize that BMV utilizes nonspecific interactions for virion assembly to enable its polyamorous, multipartite lifestyle, in which it packages 4 RNAs into 3 identical particles. The ensemble of RNA secondary/tertiary structures allowed by the capsid protein enables the virus more flexibility in packaging its many genomic RNAs.

We plan to prepare in planta BMV capsids containing either RNA1 alone, or RNA2 alone, or cellular mRNA, allowing for a comparison between pure one- (RNA1 or RNA2) and two (RNA3+4) virions and between viral and nonviral RNA-containing virions. We can also prepare any of the pure virions by in vitro assembly allowing for a comparison of in vitro reconstituted virus-like particles (VLPs) and in vivo produced virions that have packaged the same RNA molecule.
4.5 Supporting materials

Figure 4.5: Northern blot showing the purity of the RNA in the BMV3+4 virions. The lanes contain either total RNA (T) or RNA purified from virions (V) from two different agroinfiltrations. At the left, the plants have been infiltrated with wt plasmids producing wt RNAs 1, 2, 3 and 4 and all the viral RNAs are replicated and packaged. At the right, the plants have been infiltrated with modified plasmids to produce the gene products of RNAs 1 and 2 and with a plasmid producing wt RNA 3s and 4. This infiltration produces pure BMV3+4 particles as RNAs 1 and 2 are not replicated in this system.
Figure 4.6: Left) Representative cryo-electron micrograph of BMV3+4. The particles are frozen in vitreous ice suspended across holes in a carbon support film. Right) Fourier transform of the micrograph pictured at the left. The outer ring is the diffraction ring associated with vitreous water, requiring a resolution of 3.8 Å.

Figure 4.7: Flow chart describing the image acquisition and data processing for the reconstructions discussed in this work. See the MATERIALS AND METHODS section for greater detail.
4.6 Materials and methods

4.6.1 Buffers

The following buffers were employed: RNA Assembly Buffer (RAB)-50 mM Tris-HCl pH 7.2, 50 mM NaCl, 10 mM KCl, and 5 mM MgCl$_2$; Virus Storage Buffer (VSB)-50 mM sodium acetate pH 4.5 and 8 mM magnesium acetate; and Virus Running Buffer (VRB)-100 mM sodium acetate pH 5.5 and 1 mM EDTA.

4.6.2 Strategy for in vivo assembly of pure virions

Genome packaging in BMV is functionally coupled to replication, i.e., only the replication-derived progeny RNA is packaged into virions[Annamalai 2005, Annamalai 2006b, Seo 2012]. More explicitly, the interaction between CP and replicase protein p2a has been shown to dictate packaging specificity. With these essential requirements in mind, an agrobacterial transformation strategy has been developed to separately assemble virions packaging either RNA1 (B1 virions) or RNA2 (B2 virions) or RNA3 and 4 (B3+4 virions)[Annamalai 2006a].

The first inoculum is formulated to yield pure B1 virions by infiltrating N. benthamiana leaves with a mixture of agrotransformant plasmids–pB1, p2a and pCP: transcription of pB1 results in the synthesis of a biologically active full-length genomic RNA1 whose translation gives functional replicase protein p1a; similarly, agrotransformant p2a results in an mRNA competent to give replicase p2a but not competent to be replicated because it lacks the requisite non-coding regions; finally, translation of the mRNA transcribed from agrotransformant pCP gives the CP subunits for directing virion assembly. As a result, a functional replicase complex is assembled from proteins p1a and p2a that ensures the replication of B1 RNA followed by its packaging into virions by the transiently expressed CP subunits. The second inoculum is identical to that of the first, except that pB1 and p2a are substituted respectively by p1a and pB2. Agroinfiltration of this inoculum results in the assembly of virions containing B2 RNA. Finally, the third inoculum is formulated to assemble virions
packaging B3 RNA and subgenomic B4 by mixing agrotransformants of p1a and p2a and pB3. Transiently expressed proteins p1a and p2a then direct replication of B3 followed by the synthesis of sgB4 for CP production.

These particles are then purified from plants following protocols outlined by Rao et al. [Padmanaban 2005], and their quality is determined by both gel electrophoresis and negative-stain electron microscopy. Additionally, the purity of their RNA cargo is determined by extracting the RNA from virions as described by Rao et al. [Rao 2006], and then detecting the RNA species by northern blot (see Fig S.1).

### 4.6.3 Sample preparation for cryo-electron microscopy

After purification from plants, virus was washed three times with RAB through a 100 kDa Amicon filter, and concentrated to a final concentration of 5 mg/ml. CryoEM grids were prepared by depositing 2.5 µl of virus sample on a glow-discharged Quantifoil holey-carbon grid (SPI Quantifoil R1.2/1.3). The grid was then blotted dry using a Vitrobot Mark IV cryo-sample plunger, leaving behind an approximately 100-nm-thick film of solution suspended across the holes in the grid. This solution was then rapidly frozen, producing a vitrified glassy solution within the holes, by plunging the grid into a 2:1 mixture of liquid propane:liquid ethane cooled to liquid nitrogen temperatures. This rapid freezing produces thin films of amorphous ice containing the sample in the holey carbon film. The grid is then loaded into the microscope for imaging.

### 4.6.4 Cryo-electron microscopy and movie preprocessing

3150 CryoEM movies were acquired using Leginon [Suloway 2005] on a FEI Titan Krios (Electron Imaging Center for NanoMachines, California NanoSystems Institute) operated at 300 kV. All images were collected with a K2 Summit direct electron detection camera with a gatan imaging filter (GIF), and a 20 eV energy filter to remove inelastically scattered electrons. Images were acquired in super-resolution counting mode at a nominal magnification
of 130,000x, corresponding to a real pixel size of 1.07Å (0.535Å for super-resolution pixel size), with an electron dose rate of 8 electrons per real pixel per s. An accumulated dose of 48 electrons per Å², collected over a 6 second exposure, was fractionated into a movie stack of 30 image frames.

The movies were aligned for drift correction–discarding the first two frames of each movie to remove noise–and binned 2x2 by Fourier copping using MotionCor2[Li 2013] and 79,140 particles were picked using Ethan[Kivioja 2000]. Using CTFFIND4[Rohou 2015], the defocus values were determined to be in the range of 0.5 to 3 µm in the images. The data set was then preprocessed with RELION[Scheres 2012] to generate a stack of CTF-corrected, extracted particles for processing in RELION. The images averaged from all 30 frames were used for initial model searching and structure refinement, while images averaged from the first 14 frames were used for calculation of the final density map to avoid radiation damage to the sample.

4.6.5 Symmetric and asymmetric reconstruction

3D classification was carried out using I2 symmetry and a Gaussian ball as a reference, resulting in the 3.5Å-resolution structure pictured in Fig 4.3. This structure was then refined using RELION Refine3D, resulting in the 3.1Å-resolution capsid protein structure pictured in Fig 4.1. 3D classification was then carried out using C1 symmetry (no symmetry), using the symmetric structure as a model. The 3.9Å-resolution asymmetric structure that resulted from this classification is shown in Fig 4.1.

4.6.6 Structure refinement

Data was divided into two random halves and refined separately using RELION Refine3D. The final density map was then generated by merging the two half-data sets, and the resolution was determined by the "Gold-Standard" Fourier shell correlation (FSC)=0.143 [Rosenthal 2003] (see Fig 4.1).
4.6.7 Building an atomic model

An atomic model of the capsid protein was built using Coot\citep{Emsley2010} and visualized with Chimera\citep{Pettersen2004}. The model was refined with Phenix, a real space refinement program\citep{Adams2010}.

4.6.8 Capsid subtraction and asymmetric reconstruction

A 2D projection of the capsid from the symmetric structure was generated using RELION and used to subtract most of the capsid signal from the raw images, leaving behind only the density associated with the internal genome\citep{Zhang2015, Liu2015}. Since each particle has 60 quasi-equivalent orientations, we took 60 copies of each particle, increasing the total data set to 4,748,400 particles.

We then used RELION to sort the data set into 60 3D classes, and discarded any classes that became populated by more than 1/60th of the total data set, as classes with more than this number have been overpopulated by copies. We were also careful to remove any duplicates of particles that were classified into the same 3D class, to avoid any particle being used more than once in the final data. We then used the classifications and orientations determined by the subtracted, RNA-only data set and applied them to the original data that includes the capsid protein signal. In this way, full virus structures were generated, but with their orientations determined by the internal genome.

4.6.9 Transmission electron microscopy

Negative-stain electron microscopy grids were prepared by depositing 6 $\mu$l of assembly reaction on glow-discharged copper grids (400 mesh), which were previously coated with Parlodion and carbon. After a 1-min deposition, grids were blotted and stained with 6 $\mu$l of 2\% uranyl acetate for 1 min, then blotted dry. Grids were stored in a desiccator, and imaged on a FEI Tecnai TF20 electron microscope (Electron Imaging Center for Nanomachines,
California Nanosystems Institute) at 50kX magnification.

4.7 Acknowledgments

We thank Dr. Ivo Atanasov for helpful discussions related to electron microscopy, Dr. Jack Johnson for discussions related to nucleocapsid structure, and Dr. Robijn Bruinsma for discussions related to RNA-genome packaging. Electron microscope images were acquired at the California NanoSystems Institute (CNSI) Electron Imaging Center for NanoMachines.
Chapter 5

Summary and perspective on future work

5.1 Summary

The object of the work described in this thesis was to investigate how RNA secondary structure affects the structure and assembly of ssRNA viruses. Our approach was two-fold.

i) Since the complex branched structure of ssRNA arises from self-complementarity, we could greatly simplify the structure by working with the homopolynucleotide polyU, which is essentially structureless, because it lacks both self-base-pairing and helical base-stacking. Our findings highlight the important role that RNA secondary structure plays in the assembly process of RNA viruses. We show that RNAs lacking secondary structure can serve as templates for viral assembly, but that the particles formed are polydisperse when compared to wild-type particles and, in the case of CCMV CP assemblies, not wild-type sized capsids[2017].

ii) Examine the RNA structure directly by carrying out cryoEM reconstructions. In order to do so it was necessary to perform, for the first time, an asymmetric reconstruction for BMV and to work with a pure sample, containing virions with the same RNA content—
in this case RNAs3+4. Asymmetric reconstruction allows for direct visualization of the asymmetric RNA genome inside of BMV3+4 virions, and the results indicate that many RNA secondary/tertiary structures exist within the particles. Furthermore, sequence-specific RNA-CP contacts are not resolved, but rather, an amorphous shell of RNA density is seen interacting with much of the capsid interior, suggesting that the interactions driving virion assembly are nonspecific.

It has been suggested by some that there is a single paradigm for the assembly of ss-RNA viruses, one in which the assembly process is controlled by local secondary structure motifs[Turner 1986, Tang 2006, Rolfsson 2008, Calhoun 2008, Rolfsson 2010, Rolfsson 2016], i.e. in the form of sequence-specific RNA stem loops. However, work using BMV and CCMV has shown that global secondary structure that reflects the extent of RNA branching enables particular RNAs to be preferentially packaged[Cadena-Nava 2012, Comas-Garcia 2012, van der Schoot 2013, Singaram 2015]. In fact, CCMV has been shown to package a wide variety of negatively charged objects, including polyanions[Hu 2008, Cadena-Nava 2011], RNAs of varying length and sequence[Cadena-Nava 2012], polyU of various lengths[Beren 2017], and nanoparticles[Douglas 1998, Aniagyei 2009].

The work discussed in this thesis shows that BMV and CCMV utilize nonspecific interactions—proposed to be electrostatic attractions between the disordered, positively-charged N-terminus of the CP (the Arginine Rich Motif, ARM) and the negatively-charged phosphate backbone of the RNA—to drive virion assembly. In fact, many ssRNA viruses utilize disordered ARMs for RNA-binding. Our results suggest that the bromoviruses (CCMV and BMV) and bacteriophages (MS2 and Qβ) represent two different, yet effective, methods of genome packaging.

Additionally, our findings have inspired the use of CCMV as a potential drug delivery platform. Our group is working to utilize CCMV to package and subsequently deliver therapeutic RNA molecules. The applications of this delivery system are diverse, because any therapeutic agent that can be encoded by ssRNA is capable of being packaged and delivered to cells.
5.2 PolyU synthesis and packaging

Previous work has shown that secondary structure of long RNA molecules impacts the efficiency of RNA packaging into viral capsids. Both experimental [Comas-Garcia 2012] and theoretical [Singaram 2015] work on virus assembly have suggested that RNAs with more compact RNA secondary/tertiary structures are preferentially packaged by CCMV CP. Therefore, we want to investigate how an viral-length RNA lacking secondary structure, specifically viral-length polyU, is packaged into virus-like particles. Unexpectedly, we found that the structureless polyU is packaged into smaller-than-wild-type, T=2-sized capsids, even for polyU RNAs that are as much as three times longer than the viral genome. Even more surprising is the result that polyU is more efficiently packaged than the viral genome in head-to-head competition experiments.

For this work, fluorescent, viral-length polyU is synthesized using the enzyme Poly nucleotide Phosphorylase [Vanzi 2003, van den Hout 2011]. The synthesis produces a polydisperse mixture of polyU RNAs between 500-10000 nt in length, which is then fractionated to produce relatively monodisperse samples of polyU (i.e. 2500-3500 nt). These fractions are then packaged by CCMV CP following the canonical two-step self-assembly protocol. We find that the "magic ratio" [Cadena-Nava 2012], a 1:1 charge-matching between the negatively-charged phosphate backbone of the RNA and the positive resides on the N-termini of the CP, is required for complete packaging of polyU RNAs. Below the "magic ratio" some of the RNA will not be packaged and is susceptible to digestion by RNase [Garmann 2014].

We find that all lengths of polyU, ranging from 500-9000 nt in length, are packaged into smaller-than-wild-type, T=2 capsids. These capsids are resistant to RNase treatment, but exhibit morphological defects and are not perfect icosahedra (see Fig 3.4). PolyU RNAs that are longer than the CCMV viral genome (3000 nt) are packaged into multiplet capsids, in which a single RNA is shared—probably threaded—between several capsids. Such multiplet capsids have been shown for normal-composition RNAs as well [Cadena-Nava 2012], and in fact, the onset of doublet and triplet capsids occurs at the same lengths—4500 nt and 7500 nt,
respectively–for polyU RNAs and normal-composition RNAs. Interestingly, for both polyU and normal-composition RNAs, treatment of multiplets with RNase cuts the RNA that is threaded between capsids resulting in a sample containing only singlet capsids–demonstrating that the threaded RNA is exposed to nucleases, while the internal RNA is still protected.

Three distinct competition experiments[Comas-Garcia 2012], in which equal masses of two different RNAs compete for CP that is sufficient to package only one of them, were carried out between viral-length polyU and BMV RNA1 (B1): polyU and B1 were incubated at neutral pH followed by the addition of CP (called Both First), polyU was incubated with CP first followed by the addition of B1 (called polyU first), and B1 was incubated with CP first followed by the addition of polyU (called B1 first). All three samples were first allowed to equilibrate at neutral pH, at which point the pH was lowered resulting in capsid formation.

The results of these experiments are two-fold: (i) polyU RNA outcompetes B1 in the both first competition experiment, although B1 had beaten every normal-composition RNA tested, and (ii) the order of mixing affects the relative amounts of each RNA packaged. If polyU is incubated with CP first polyU is the predominant RNA packaged (95% polyU, 5% B1), while in the ”Both First”-case polyU wins the competition to a lesser degree (77% polyU, 23% B1), and in the ”B1 First”-case B1 wins the competition but by the smallest margin yet (35% polyU, 65% B1).

It is surprising that polyU outcompetes viral RNA for CP, as it has a radius of gyration 25% larger than the viral RNA and lacks the secondary-structure-motifs canonically associated with protein binding. However, that the order of incubation affects the results of the assembly illustrate that their is a kinetic effect associated with the competition results. Curiously, a 2000-nt normal-composition RNA (in fact, a truncation of B1), which like polyU also forms T=2-sized capsids, also did unexpectedly well in a head-to-head competition with B1. However, a 1500-nt truncation of B1 was unable to compete despite the fact that it forms T=2-sized capsids[Comas-Garcia 2012]. These results suggest that the formation of
T=2-sized capsids may help polyU to compete during the competition reactions, possibly due to faster kinetics of assembly, but that the formation of T=2 particles is not the sole reason polyU is so effectively packaged.

**Future work on packaging**

We have begun packaging viral-length polyA RNAs using CCMV CP, and preliminary results suggest that polyAs are not well-packaged. Specifically, we find that polyAs shorter than 2000 nt in length are able to be packaged into T=2-sized capsids but that the resulting capsids are not resistant to RNase treatment. Curiously, polyA RNAs longer than 2000 nt are not packaged into virus-like particles. While these results are preliminary they suggest that polyA, which unlike polyU exhibits base-stacking resulting in helical ordering of the RNA[Rich 1961, Leng 1966, Seol 2007], is not effectively packaged by CCMV CP. Surprisingly, previous works have demonstrated that polyA is double-stranded below pH 5, which may impact the self-assembly process since the second-step of the assembly utilizes a drop in pH to 4.5[Petrovic 2005, Ke 2009].

Another interesting set of experiments that has been considered involve synthesizing viral-length polyU RNAs from seed oligos with sequences associated with "Packaging Signals". These packaging signals have been shown to have a high affinity for CP, and these experiments allow for direct comparison of each packaging signal within a viral-length polyU RNA.

Lastly, we are interested in studying how RNAs are shared between multiplet capsids. These experiments will help us to understand how multiplets form, specifically, whether the assembly occurs via two distinct, irreversible nucleation and growth events, or if the RNA is equally shared across multiplets suggesting that the RNA is allowed to equilibrate across particles after capsid formation. We plan to use RNase digestion followed by RNA extraction from virus-like particles as a means of assaying what lengths of RNA are packaged in each virion within multiplet structures. Additionally, gold nanoparticles can be used to label the ends of the RNA, allowing for visualization of where the ends are located within
multiplets using cryoEM. In fact, we have previously investigated the end-to-end distance for long ssRNA molecules labeled with 5-nm gold nanoparticles using electron microscopy, and find that the ends are close for RNAs of different length and sequence (Data not shown). We plan to use smaller—less than 2-nm in diameter—gold nanoparticles in the future, to avoid perturbing the structure of the RNA.

5.3 CryoEM of pure BMV virions

This work sets out to determine the structure of the genome of a multipartite ssRNA virus. We have used pure BMV3+4 virions—a sample containing only virions with one copy each of RNA3 and RNA4—produced in planta to avoid obscuring the genome through the presence of a mixture of particles. Additionally, this is the first high-resolution structure determined in a physiological buffer for BMV. Two different methods are used for the 3D structure determination: the first imposes icosahedral symmetry on the data set and will necessarily result in an icosahedral structure, obscuring any asymmetry in the particle; the second does not impose any symmetry and instead uses the asymmetry present in the particles to align the data into a 3D structure.

For both the symmetric and asymmetric reconstructions the capsid shell agrees well with structures determined previously using cryoEM symmetric reconstruction[Wang 2014] and x-ray crystallography[Lucas 2002], as does the RNA density seen at the hexamers of the symmetric structure[Hoover 2016]. We have determined the first atomic model of BMV CP using asymmetric reconstruction, and find good agreement with the atomic model determined by symmetric reconstruction. In fact, the symmetric and asymmetric structures are nearly identical due to the dominance of the capsid protein signal during the asymmetric structure determination. Therefore, the capsid protein signal was subtracted from the data set[Zhang 2015], leaving behind only the signal of the asymmetric RNA, and new structures were determined using only the RNA signal.
The orientations of the particles are determined using the RNA signal, and then the full particle—including the capsid protein—is calculated using the raw (unsubtracted) data with the orientations determined from the RNA-only data. We find many distinct secondary/tertiary structures for the RNA genome, suggesting that the RNA has many conformations inside the virion. Most of the RNA genome is resolved to 15Å resolution near the capsid protein surface, with much of the interior of the particle remaining empty. It is important to note that the RNA itself is limiting this resolution, as the capsid protein is resolved to near-atomic resolution.

Our results indicate a fundamental difference between the ssRNA organization in the multipartite, symmetric BMV viral capsid and the ssRNA organization in the monopartite, asymmetric capsids of the bacteriophages MS2 and Q-beta. For these bacteriophages, it has been shown that a dominant RNA conformation is found inside the assembled viral capsids, and RNA density is conserved even at the center of the particle[Koning 2016, Dai 2017, Gorzelnik 2016]. We find in BMV, on the other hand, that the RNA is only organized at the capsid shell, and our results indicate that there is not a single dominant conformation for the RNA inside the virus but rather an ensemble of RNA structures that interact strongly with the capsid protein shell. This is not surprising, as the two differ significantly in a variety of ways, all of which would lead to the conclusion that the encapsidated genomes of bacteriophages will be more ordered than the packaged genomes of bromoviruses.

The most obvious difference between the two virions is the asymmetry of the bacteriophage capsid, as compared to the highly-symmetric capsid of BMV. Moreover, the MS2 maturation protein, which takes the place of a capsid protein dimer in the assembled virion, is known to bind the ends of the RNA genome, imparting an asymmetry on the internal genome that is congruent with the capsid protein[Koning 2016, Dai 2017]. The unique asymmetry of these capsids also serves to reduce the possible orientations of a T=3 capsid from 60, the number of quasi-equivalent units in these structures, to 1.

Another major difference between the two virions is the number of RNA molecules en-
capsidated. MS2 has a single RNA molecule inside, while the BMV3+4 particle discussed here has two RNAs inside, which could result in less genome-ordering as compared to MS2. It is worth noting that we carried out a symmetric reconstruction on mutant virions—one positive charge in the N-terminus has been made neutral—containing only RNA3 (BMV3) and we find that the RNA is organized in rings similar to the RNA rings seen in BMV3+4 (Fig 4.3). This result suggests that the RNA organization is not dependent on the number of RNAs packaged.

Additionally, MS2 is known to have strong, highly-specific contacts [Toropova 2008, Rolfsson 2008, Rolfsson 2010, Rolfsson 2016] between the CP and the RNA genome, which dictate that the RNA be structured so as to make those specific interactions[Dykeman 2011, Dykeman 2013]. BMV CP, on the other hand, is known to interact with RNA through the electrostatic attractions between the positively-charge N-termini of the CP and the negatively-charged phosphate backbone of the RNA[Bayer 2005]. These interactions have been shown to be non-specific in nature, and in fact, BMV is capable of packaging any normal-composition, viral-length RNA in vitro independent of its sequence[Cadena-Nava 2011, Cadena-Nava 2011, Hu 2008].

The last major difference is in the way that the CPs of MS2 and BMV interact with their internal genomes. The contacts between MS2 and its RNA-genome are made directly at the internal surface of the capsid[Koning 2016, Dai 2017], again requiring the RNA to organize itself in regard to the structure of the capsid. BMV CP, in contradistinction, interacts with its genome through disordered N-termini which extend into the internal portion of the virion[Lucas 2002, Wang 2014, Hoover 2016], allowing the RNA more freedom in its organization. All of these differences act in concert, resulting in an RNA genome which is highly ordered particle-to-particle in MS2 and Qβ, and an RNA genome which is disordered in BMV. Consequently, highly-sequence specific CP-RNA interactions are considered essential for proper assembly of MS2 but not for BMV, and in vivo experiments suggest that BMV packaging is directly linked to replication[Annamalai 2005].
5.3.1 Future work on cryoEM analysis of BMV virions

In the same manner as described above for pure BMV3+4 virions, we propose to obtain data for \textit{in planta} produced BMV1 and BMV2 virions and to apply the tools of asymmetric reconstruction to investigate how both the capsid structure and the RNA structure differ depending on the encapsidated RNA. It will also be important to similarly investigate the pure BMV3 virion, which is not a wild-type particle but is of special interest because it contains only the 2100 nt of RNA3 and yet has the same T=3 capsid structure (as opposed to the T=2 nucleocapsids formed in vitro from the same RNA3) as the longer RNA2 (2800nt) and RNA1 (3200nt); this potentially affords us the biggest difference in RNA and capsid fluctuations and hence translation efficiencies. We can also prepare \textit{in planta} BMV capsids containing cellular mRNA, allowing for a comparison between viral and nonviral RNA-containing virions, and we can prepare any of the pure virions by \textit{in vitro} assembly allowing for a comparison of \textit{in vitro} reconstituted virus-like particles (VLPs) and \textit{in vivo} produced virions that have packaged the same RNA molecule. Recent work by Hoover et al.\cite{Hoover2016} has suggested that post-translational modifications to the virus can impart different physical properties, and they have shown that the virion cargo can affect the extent of post-translational modifications as well.

In addition to the high-resolution structure determination work illustrated above, we plan to compare the pure BMV virions using cryoEM variance analysis\cite{Tang2014}, which utilizes the heterogeneity in a cryoEM dataset to determine the relative structural variance amongst similar 3D structures. We plan to use this analysis to determine whether the pure virions have different structural heterogeneities that can be used to determine their dynamical properties, which is directly related to the propensity of a particle to allow RNA to fluctuate out of the assembled capsid.
Appendix A

Measuring the end-to-end distance of long ssRNAs

A.1 Introduction

Biological processes are often regulated by RNA circularization [Gallie 1991]. A large number of RNA viruses require circularization for initiation of genome replication, while circularization of messenger RNAs recruits cellular cofactor-proteins to initiate translation [Karetnikov 2006].

In addition, many of these RNA viruses have evolved complementary sequences near the ends of their RNA, suggesting evolution favors close proximity of the ends. Specifically, Yellow Fever Virus [Corver 2003], an 11,000-nt long RNA virus has evolved a 21 nucleotide base-pairing sequence at its termini, while HIV-1 [Ooms 2007], Influenza A [Hsu 1987] and Sindbis [Frey 1979] have also been shown to have sequences that promote base-pairing at their ends.

While many of these systems utilize base-pairing sequences to “lock in” circularization (many others do not), a general mechanism is still necessary to bring the ends close for this base-pairing to occur. Recent theoretical studies have suggested that the ends of long ssRNAs
are in close proximity independent of length and sequence, providing a simple means for
circularization to occur across many biological systems. In particular, Yoffe et al.[Yoffe 2011]
demonstrated by a variety of arguments based on RNA folding algorithms that the ends of
any large, single-stranded oligonucleotide (both biological and random sequences were tested)
are necessarily close (within several nanometers), independent of length or sequence.

This general statement about all large oligonucleotides can be attributed to the consider-
able amount (greater than 60%) of self-complementarity that arises in these molecules for any
reasonable composition (e.g. roughly equal proportions of each nucleotide). This significant
degree of self-complementary inherently leads to base-pairings between nucleotides that are
far from one another along the RNA, resulting in close proximity of the ends independent of
the length or sequence[Fang 2011b]. This research aims to test these results experimentally,
and preliminary results have already been obtained. A related experiment using fluorescence
resonance energy transfer (FRET) spectroscopy, with RNAs labeled with a fluorophore (half
of the FRET pair) at each end of the RNA, has recently been described[Leija-Martinez 2014].
While these results have measured the end-to-end distances of a variety of RNAs using
FRET, we would like to measure this distance directly by visualization with cryo-electron
microscopy (cryoEM).

We plan to investigate the end-to-end distances of RNAs before and after packaging them
with cowpea chlorotic mottle virus (CCMV) capsid protein (CP). This work is of particular
interest when applied to the investigation of RNAs that are too long to be packaged into a
single CCMV capsid, but instead are packaged into "multiplet" capsids, in which a single
RNA molecule is threaded between several capsids. We hope that the use of end-labeled
RNAs will allow us to resolve how the RNA is threaded between the multiple capsids.
A.2 Materials and methods

To examine experimentally if the ends of any RNA are necessarily close, single-stranded RNA’s ranging from 500-6000 nt with varying sequences are synthesized (by in vitro transcription) and tested. Many of these proposed RNA’s are derived from viral RNA genomes, and subsequently, the results of this work are of particular interest to RNA virus research.

These RNA’s are labeled at both the 3’ and 5’ ends with 1.8 nm gold nanoparticles [Ackerson 2005] that are visualized using cryoEM. Measurements of the proximity of the ends of the RNAs will be made for varying amounts of denaturant, to show that as secondary and tertiary interactions are suppressed, the distance between the ends of the RNA will approach those of a linear polymer chain (i.e. the distance will increase in a known way with increasing RNA length instead of staying the same as predicted for the native state).

RNAs are labeled through the method of dual splinted ligation. While this procedure is fairly common for short RNA’s (up to 100 nt), a novel procedure for the ligation of long RNA’s must be developed here. This novel ligation protocol is of particular interest to virus research, as viral RNA genomes are notoriously long, and dual labeled oligonucleotides can be followed in vivo using FRET measurements (although the method can be used for any large RNA molecule and many labeling techniques).

During splinted ligation, two 20 nt RNAs (thiol-terminated for thiol-gold chemistry) are stapled to the desired RNA by T4 DNA ligase (see Fig A.1. The sequences of these short RNA oligonucleotides have been chosen to be non-complementary, to prevent them from bringing the ends of the RNA close together. The oligos are also sufficiently short compared with the size of the RNAs tested (less than 5% of the total size) so as to not affect the measurement of the end-to-end distance of the RNA. Preliminary results show that this ligation methodology produces dual labeled long RNA molecules.

The thiol modified RNAs are conjugated to 1.8-nm gold nanoparticles via incubation. After attachment of the nanoparticles, the samples are flash-frozen to preserve their secondary/tertiary structures. In addition, flash-freezing produces a vitrified medium contain-
Figure A.1: Splinted ligation protocol. First, a thiolated RNA oligo is linked at both the 5’ and 3’ ends of the RNA using a DNA staple. The RNA oligos are then covalently attached to the 5’ and 3’ ends of the RNA using T4 DNA ligase, resulting in an RNA molecule with reactive thiol groups at both of its ends. Oligos that are complementary to the DNA staple are then added in excess, removing them from the RNA, and the RNA is then equilibrated by heating and cooling. Lastly, gold nanoparticles are added to the ends of the RNA by thiol-gold chemistry.
Figure A.2: Left) Cryo-electron micrograph of 1.8 nm gold nanoparticles[Ackerson 2005] Right) Cryo-electron micrograph of the 3200 nt BMV RNA1. This figure shows that both the RNA and the gold nanoparticles can be imaged in cryoEM. Scale bar is 20nm.
ing the RNA, allowing the RNA to be visualized without spoiling its structure through interactions with the substrate. Using cryoEM, it is possible to see both the RNA (when the microscope is slightly under-focused) and the nanoparticles (when the microscope is in focus) at subsequent times (see Fig A.2). In this way, real-space images of the entire RNA-nanoparticle conjugates can be constructed. These cryoEM images can then be analyzed for various amounts of denaturant, and the end-to-end distances of the RNA at different denaturant concentrations can be determined.

**A.3 Results**

The aim of this work is to experimentally test the validity of a theoretical prediction, namely that the ends of any RNA are in close proximity independent of length and sequence[1].

It is possible to discern between paired nanoparticles (two nanoparticles attached to the same RNA) and unpaired nanoparticles (nanoparticles bound to two separate RNAs) in cryoEM images by diluting the sample, i.e. controlled dilution of the sample results in a means of determining the difference between a paired set of nanoparticles (which would both be bound to the same RNA) and unpaired nanoparticles.

Preliminary results (see Fig A.3) using 5-nm gold nanoparticles suggest that the ends are close for both the viral-length BMV RNA1 (3234 nt) and a 500-nt-truncation of BMV RNA1. These results show that the splinted ligation protocol is capable of adding thiols to the ends of the RNA, and that 5-nm gold nanoparticles can bind to the thiol-modified ends of the RNA. These results are preliminary, as they utilize 5-nm gold nanoparticles, which may be too large for these measurements, and they were acquired using negative-stain electron microscopy, which involves drying samples on a substrate and could possibly alter the secondary/tertiary of the RNA.

The results of this work are shown in Fig A.4, in which the radial distribution function of gold nanoparticles is plotted as a function of radius (nm). The figure shows that the ends of
Figure A.3: Left) Negative-stain electron micrograph of 5 nm gold nanoparticles attached to the ends of a 500-nt truncation of BMV RNA1. Right) Negative-stain electron micrograph of 5 nm gold nanoparticles attached to the ends of 3200-nt BMV RNA1. Scale bar is 20 nm.

Both RNAs are close, around 7 nm and 10 nm, respectively, for the 500 nt and 3234 nt long RNAs. Leija-Martinez et al. found that the end-to-end distances of long RNA molecules ranged from [Leija-Martinez 2014]. In particular, they measured the end-to-end distances of several 574-nt RNAs and found them to range from 6- to 8-nm, while BMV RNA 1 (the same RNA we have looked at here) had an end-to-end distance between 9- and 10-nm.

However, since the diameters of the gold nanoparticles are roughly the same size as the end-to-end distances being measured, we would like to acquire data in cryoEM using smaller gold particles for better resolution, and to ensure that our measurements are not being perturbed by the size of the particles or an interaction with the negative-stain electron microscopy grid.
Figure A.4: Left) Negative-stain electron micrograph of 5 nm gold nanoparticles attached to the ends of a 500-nt-truncation of BMV RNA1. Right) Negative-stain electron micrograph of 5 nm gold nanoparticles attached to the ends of 3234-nt BMV RNA1. Scale bar is 20 nm.
A.4 Future work

We plan to label the ends of RNAs of varying length and sequence with small, 1.8-nm diameter gold nanoparticles, and image them with cryoEM to measure their end-to-end distance in solution. We will then investigate the end-to-end distances of RNAs before and after packaging them with CCMV CP. This work will utilize RNAs labeled with both 1.8-nm diameter gold nanoparticles and with a FRET pair. The gold nanoparticles will allow for direct visualization of the ends of the RNA, while the FRET analysis can give dynamic information as well. The FRET work will be done in collaboration with the group of Professor Jaime Ruiz-Garcia at the Universidad Autonoma de San Luis Potosi, as they have already published measurements of the end-to-end distance of RNAs in solution using FRET[Leija-Martinez 2014].

A.5 Acknowledgments

Much of this work was done by Mauricio Comas-Garcia, a former PhD student in the group, and Chanrith Siv, an outstanding former undergraduate. I would like to thank them very much for their work and for helping me to start working on this project.
Bibliography


