



What goes right and wrong during virus self-assembly?

Citation

Williams, LaNell Alexandria. 2023. What goes right and wrong during virus self-assembly?.
Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37375736>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences




DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Physics
have examined a dissertation entitled


What goes right and wrong during virus self assembly?

presented by LaNell Alexandria Williams

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature  _____

Typed name: Professor Vinodh Manoharan, Chair

Signature  _____

Typed name: Professor Daniel Needleman

Signature  _____

Typed name: Professor Michael Brenner

Date: May 5, 2023

What goes right and wrong during virus self assembly?

Understanding how bacteriophage MS₂ coat protein assembles
around RNA

A DISSERTATION PRESENTED
BY
LANELL ALEXANDRIA WILLIAMS
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2023

©2023 – LANELL ALEXANDRIA WILLIAMS
ALL RIGHTS RESERVED.

What goes right and wrong during virus self assembly? Understanding how bacteriophage MS₂ coat protein assembles around RNA

ABSTRACT

Self-assembly is a vital part of the RNA virus life cycle. The assembly of viral coat proteins around viral RNA occurs both *in vivo* and *in vitro*, suggesting that viral capsid assembly may be driven by minimization of free energy. To better understand this process, we study the assembly of bacteriophage MS₂ virus-like particles (VLPs) as a function of coat-protein concentration, RNA concentration, ionic strength, temperature, and viscosity. We use dynamic light scattering, transmission electron microscopy, and gel electrophoresis to determine the size distribution, morphology, and composition of the assembled structures. We show how the morphology of MS₂ VLPs varies with these conditions, and we relate these results to a model of assembly involving nucleation and growth.

Contents

| | |
|--|-----------|
| Title Page | i |
| Copyright | ii |
| Abstract | iii |
| Table of Contents | iv |
| List of Figures | vi |
| Dedication | xi |
| Acknowledgements | xii |
| 1 INTRODUCTION | 1 |
| 1.1 Virus Self-Assembly and Structure | 1 |
| 1.2 Overview | 3 |
| 2 EFFECT OF CAPSID PROTEIN CONCENTRATION ON MS₂ COAT PROTEIN ASSEMBLY AROUND RNA | 4 |
| 2.1 Introduction | 5 |
| 2.2 Results | 7 |
| 2.3 Discussion | 14 |
| 2.4 Conclusions | 18 |
| 2.5 Materials and Methods | 20 |
| 3 THE ROLE OF ELECTROSTATICS IN THE FORMATION OF MS₂ COAT PROTEIN SCAFFOLDS | 29 |
| 3.1 Introduction | 30 |
| 3.2 Results | 31 |
| 3.3 Discussion | 39 |
| 3.4 Conclusion | 40 |
| 3.5 Materials and Methods | 42 |
| 4 A PHASE DIAGRAM OF MS₂ VIRUS SELF-ASSEMBLY | 45 |
| 4.1 Introduction | 46 |
| 4.2 Results and Discussion | 47 |
| 4.3 Conclusions | 59 |

| | | |
|-----|--|----|
| 4.4 | Materials and Methods | 60 |
| 5 | CONCLUSIONS AND FUTURE WORK: HOW TO BUILD A VIRUS AND POTENTIAL APPLICATIONS | 62 |
| 5.1 | A phase diagram of virus self-assembly and beyond | 63 |
| 5.2 | MS2 coat-protein assembly around scaffolds | 65 |
| | REFERENCES | 71 |

Listing of figures

| | | |
|-----|--|----|
| 2.1 | Overview of experiments and results. We mix MS2 coat protein with MS2 RNA to make a solution with 50 nM RNA concentration and varying coat-protein concentration. The transmission electron microscopy images, gel electrophoresis measurements (full image is shown in Fig. 2.2), and dynamic light scattering results demonstrate a transition from well-formed MS2 VLPs to monster capsids to RNA-protein condensates with increasing coat protein concentration. The main text and subsequent figures elaborate on all of these results. | 7 |
| 2.2 | Images of agarose gels used to characterize the assembly products. We first stain with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat protein (bottom image). Lanes 1 (leftmost lane) and 20 (rightmost lane) show a DNA ladder. Lanes 2–4 show three controls: MS2 RNA (lane 2), wild-type MS2 capsids (lane 3), and MS2 RNA treated with RNase (lane 4). The other lanes show the results of gel electrophoresis on samples prepared with coat-protein dimer concentrations ranging from 2.5 μM to 30 μM . The region highlighted in purple shows that the amount of wild-type-sized products increases as dimer concentration increases from 2.5 to 7.5 μM and then decreases sharply at 8.7 μM | 9 |
| 2.3 | Plots of size distributions of wild-type MS2 virus particles and VLPs assembled <i>in vitro</i> at 50 nM concentration of free RNA and varying coat-protein dimer concentrations. The distributions are inferred from dynamic light scattering measurements. The first column shows the size distribution on a number basis, the second column shows the size distribution on a volume basis, and the third column shows the measured autocorrelation functions. Light gray peaks in the distributions show the results from eight individual experiments. Dark gray peaks show the results inferred from the average autocorrelaton function. The autocorrelation functions for each individual measurement are shown in light grey in the plots at right, and the average is shown in dark gray. | 11 |

| | | |
|-----|--|----|
| 2.4 | TEM images from negatively stained samples of wild-type MS2 particles and products of assembly at varying coat-protein dimer concentrations. At concentrations less than 10 μM , most particles have the shape and size of wild-type capsids. At higher concentrations, we observe clusters of partially formed capsids that increase in size with concentration. The dotted line in the inset of the 15 μM image shows the outline of one such partial capsid. All scale bars are 100 nm. | 12 |
| 2.5 | TEM images of the same samples as shown in Fig. 2.4, but at different magnifications for each sample. The low magnification images of the samples at 12.5 μM coat-protein dimer concentration and higher show that the assembled structures are several micrometers in size. | 13 |
| 2.6 | Representative snapshots from Monte Carlo simulations of capsomer-and-polymer systems with (A) a 12:1 ratio of capsomer to polymer and (B) a 50:1 ratio. The volumes of the simulation cells and reduced temperature of the simulations are identical (see Section 2.5). Each capsomer is modelled as a hard disk with five sticky patches on its rim to mediate capsomer-capsomer interactions, and a large sticky patch on its face to mediate capsomer-polymer interactions. At the low capsomer:polymer ratio, discrete particles form containing 12 capsomers surrounding a polymer (a select few are highlighted in panel A). At the high capsomer:polymer ratio, extended clusters form, mediated by capsomer-capsomer interactions, that contain multiple polymer chains and which resemble the RNA-protein condensates found in the experiments. In the images at right, the capsomers are made transparent so that the polymer conformations are more visible. | 15 |
| 2.7 | Overview of the experiments and results, showing how coat protein concentration affects the structures of MS2 coat protein assembly around RNA. We mix 50 nM MS2 RNA with varying amounts of MS2 coat protein. We find three kinds of assembly products. The first (top row), which forms at 5 μM coat protein, consists of well-formed capsids that in some cases appear to stick together. The second (middle row), which forms at 10 μM protein concentration, consists of monster particles: multiple partial capsids that decorate the RNA. The third (bottom row), which forms at 15 μM protein concentration, consists of condensates of RNA and coat protein. Diagrams at left and middle illustrate potential pathways for formation of these products. Images at right are from transmission electron microscopy (TEM) of negatively stained MS2 VLPs. | 17 |
| 3.1 | TEM images of negatively stained samples of wild-type MS2 particles mixed with varying coat-protein dimer concentrations. At concentrations less than 5 μM , we begin to see the formation of an additional coat-protein layer and, in some cases, clusters of particles. The number of layers and clusters increases with increasing concentration. | 32 |

| | | |
|-----|--|----|
| 3.2 | TEM image of a negatively stained sample of MS2 virus-like particles (VLP) mixed with 15 μM coat-protein dimers. Again, we see the formation of an additional coat protein layer that forms around the VLP. | 33 |
| 3.3 | Images of agarose gels used to characterize samples. We first stain the samples with ethidium bromide to detect the RNA (top image) and then with Coomassie Blue R-250 to detect protein (bottom image). Lanes 1 (leftmost lane) and 12 (rightmost lane) show a DNA ladder. Lane 2 shows a control sample of wild-type MS2 capsids. Lanes 4–6 and 8–10 show samples prepared with coat-protein dimer concentrations from 1.25 to 20 μM (lane 7 represented a concentration of 7.5 μM but there was a mistake in the preparation of the sample). An additional sharp band appears that broadens with increasing coat-protein concentration. | 34 |
| 3.4 | Plots of size distributions of wild-type MS2 virus (top row) and of 50 nM MS2 wild-type virus mixed with varying concentrations of coat-protein dimers (bottom five rows). The distributions are inferred from dynamic light scattering measurements. The first column shows the size distribution on an intensity basis, the second column shows the size distribution on a volume basis, and the third column shows the measured autocorrelation functions. The light grey peaks in the distribution show the results from eight individual experiments. Dark gray peaks show the results inferred from the average autocorrelation function. The autocorrelation function for each individual measurements is shown in light gray. The diameter d corresponding to each peak is noted on each distribution. | 36 |
| 3.5 | Analysis of infectivity of wild-type MS2 and wild-type virus with added coat protein. A. Confocal micrographs (raw images and histogram adjusted images to show individual pixels) of wild-type MS2 labeled with AF 647 mixed with 5 μM , 10 μM , and 15 μM coat-protein dimers (CP ₂). B. A plot of the F-pili fluorescence fraction (a measure of the ratio of the number of particles bound to F-pili to the total number of particles). C. Photographs of a plaque assay of wild-type MS2 (left) and wild-type MS2 added to 15 μM coat-protein dimers (right). There are more plaques (empty regions) in the left image (wild-type MS2), indicating that the additional coat protein reduces the infectivity but does not eliminate it. | 38 |
| 3.6 | Images of plaque assays done with 50 nM wild-type MS2 virus mixed with 0, 5, 10, and 15 μM concentration of coat-protein dimers. We wait 10 min for assembly to occur. These experiments show a decrease in the number of plaques, which indicates a decrease in the overall infectivity. | 39 |
| 3.7 | Diagram showing pathway by which multi-layer MS2 capsids form. | 41 |

| | | |
|-----|---|----|
| 4.1 | Images of agarose gels used to characterize the assembly products of 5 μ M MS2 coat-protein dimers with varying MS2 RNA concentrations ranging from 10 nM to 200 nM. We first stain with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat proteins (bottom image). Lane 1, 6, 7, and 16 all contain 1 kb DNA ladders. Lanes 2, 3, and 5 show controls: MS2 RNA, wild-type MS2 virus, and MS2 RNA treated with RNase. Lanes 9–14 show the assembly products of these experiments. We observe that with increasing RNA concentration the particles that have assembled have the same size as the wild-type MS2 virus. | 48 |
| 4.2 | TEM images of negatively stained samples of assembly products of 5 μ M MS2 coat-protein dimers at varying MS2 RNA concentrations. At all concentrations of MS2 RNA we see well formed MS2 VLPs that resemble the wild-type MS2 virus. . . . | 48 |
| 4.3 | Images of agarose gels used to characterize assemblies formed when we mix 50 nM MS2 RNA with 5 μ M MS2 coat-protein dimers at NaCl concentrations ranging from 100 mM to 1000 mM. Lanes 3, 4, and 6 show the following controls: MS2 RNA, wild-type MS2 virus, and digested RNA. Lanes 1, 11, and 20 contain 1 kb DNA ladders. Lanes 8, 9, and 13–18 contain the results from experiments done at varying NaCl concentration. We observe a broadening of the topmost band with increasing NaCl, indicating that larger structures are forming that contain both RNA and protein. | 50 |
| 4.4 | TEM images of negatively stained samples of assembly products of 50 nM concentration of MS2 RNA mixed with 5 μ M MS2 coat-protein dimers at 100 mM, 300 mM, 600 mM, and 1000 mM NaCl concentrations. At lower concentrations (100 mM and 300 mM NaCl), we see well-formed MS2 VLPs. At higher concentrations, we see aggregates of capsids. At the highest NaCl concentration, 1000 mM, we see large, disordered aggregates that resemble the condensates seen in chapter 2. | 50 |
| 4.5 | Higher-magnification TEM images of the structures formed at 1000 mM NaCl. | 51 |
| 4.6 | Plots of size distribution of wild-type MS2 virus and VLPs assembled <i>in vitro</i> as inferred from dynamic light scattering. We show the size distribution on a volume basis. The light gray peaks indicate eight individual experiments. We observe an increase in average particle size with increasing NaCl concentration. | 52 |
| 4.7 | Image of gel containing 50 nM wild-type MS2 virus in buffers containing 100 mM to 1000 mM NaCl. Lanes 2 and 3 contain controls for MS2 RNA and wild-type MS2 virus at 100 mM NaCl. In addition, lane 19 contains a control for an MS2 VLP assembled in buffer containing 1000 mM NaCl. Lanes 1 and 19 contain DNA ladders. The remaining lanes contain samples of wild-type MS2 virus at varying NaCl concentrations. We observe an overall increase in the size of these particles, suggesting clustering or aggregation. | 52 |
| 4.8 | TEM images of negatively stained samples of 50 nM wild-type MS2 virus in 1000 mM NaCl. We observe clustering of the wild-type capsids. | 53 |

| | | |
|------|---|----|
| 4.9 | Image of gel showing 1 kb DNA ladder (lanes 1 and 13), controls consisting of MS2 RNA (lane 2), wild-type MS2 virus (lane 3), and MS2 RNA at NaCl concentrations from 100 mM to 1000 mM (lanes 7–11). The results show that RNA aggregation may occur with increasing salt concentration. | 54 |
| 4.10 | Diagram showing how ionic strength might affect the pathways and morphologies of assembly. Top diagrams show a pathway for assembly of well-formed VLPs at low NaCl concentration (100 mM). Middle diagrams show a pathway for assembly of clusters of capsids at 600 mM NaCl. Bottom diagrams show a pathway for assembly of condensates at 1000 mM NaCl. The images to the right of each diagram are transmission electron micrographs of negatively stained MS2 VLPs. | 55 |
| 4.11 | Images of agarose gels used to characterize assembly products at varying temperature. Lanes 1 and 20 show a 1 kb DNA ladder. Lanes 2 and 3 show two controls: MS2 RNA (lane 2) and wild-type MS2 virus (lane 3). The other lanes show the results of gel electrophoresis on samples assembled at temperatures ranging from 4 °C to 60 °C. One of the most striking features of this gel is the increase in intensity of the lower band as temperature increases to 32 °C and decrease in intensity above this temperature. In addition, we see a second band appear starting at 37 °C and disappear above 45 °C. Above 45 °C the samples do not travel past the wells of the gel. | 56 |
| 4.12 | TEM images of negatively stained samples of 50 nM MS2 RNA mixed with 5 μM MS2 coat-protein dimers at varying temperatures (°C). | 57 |
| 4.13 | Lower-magnification TEM images of the same samples shown in Fig. 4.12. | 57 |
| 4.14 | Plots of size distributions of MS2 VLPs assembled <i>in vitro</i> at various temperatures from 50 nM free MS2 RNA and 5 μM concentration of MS2 coat-protein dimers. The distributions are inferred from dynamic light scattering measurements and show the size distribution on a volume basis. The light gray peaks show the results of nine individual experiments. | 58 |
| 4.15 | Phase diagrams of MS2 VLP assembly (bottom left) comparing coat-protein dimer concentration, MS2 RNA concentration, NaCl concentration, and temperature. We observe 6 potential morphologies or assembly products that can be explored (top right): well-formed capsids, monsters, condensates, capsids that stick, incomplete or <i>en masse</i> capsids, and aggregates. | 61 |
| 5.1 | Images of agarose gels used to characterize products of assembly at increasing concentrations of glycerol. We stain first with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat protein (bottom image). Lanes 1 and 19 contain 1 kb DNA ladders. Lanes 2, 3, and 5 contain the following controls: MS2 RNA, wild-type MS2 virus, and digested RNA. The remaining lanes contain 5 μM coat-protein mixed with 50 nM RNA at varying glycerol concentrations ranging from 0% to 99%. | 64 |

I WOULD LIKE TO DEDICATE THIS THESIS TO MY PARENTS. WITHOUT THEM I WOULDN'T BE THE PERSON THAT I AM TODAY. I AM A PRODUCT OF THEIR VALUES, THEIR LOVE, THEIR WISDOM AND THEIR SACRIFICE.

Acknowledgments

First, I would like to thank my thesis advisor Professor Vinny Manoharan. Vinny has been very supportive over the years, but I am most grateful that he reminded me that part of this process is supposed to be fun. Without that reminder, I don't believe I would have continued my PhD nor would I have found a way to replenish my love for science. Vinny provided me a level of independence that allowed me to cultivate my own ideas and pursue my own pathway. At each step, he supported my research and my well-being. For that I am truly grateful.

Next, I would also like to thank the rest of my thesis committee: Professor Daniel Needleman and Professor Michael Brenner. Dan provided a level of very thoughtful feedback that I truly appreciated. My conversations with him helped me realize that I wasn't replenishing my need for scientific community and debate in ways that were fulfilling to me. His support and conversation were much appreciated. Michael provided a level of energy and enthusiasm that I truly appreciated. His candor was always very refreshing. I can honestly say that any conversation or interaction I had with Michael, left me feeling excited as well! Each of my committee members contributed to my journey in very unique ways and I am thankful for each of them.

I would also like to thank the Manoharan lab as a whole for their support. In particular, Dr. Aaron Goldfain and Professor Rees Garmann. Aaron and Rees taught me everything that I know about MS2 virus self-assembly and I honestly couldn't have had a better pair of people to learn virus assembly from. I would also like to thank Dr. Tim Chiang and Amelia Paine for their support and engagement. It has been a pleasure working with them over the past few years. In addition, I would like to thank Professor Danai Montalvan and Professor Nabila Tanjeem for their constant support during their time at Harvard. The first half of my PhD would have looked very differently without them. I will be forever grateful for their sisterhood. Nabila was such an inspiration for me during my PhD and I learned so much about soft condensed matter from her. Danai taught me so much about all of biology (including viruses!) and I'm so grateful that she is one of my closest friends.

Next, I would like to thank the faculty and staff who have supported me within Harvard University's Physics Department. In particular, Dr. Jacob Barandes (who would often pop up in the most timely ways to support me), Lisa Cacciabauda (for introducing me to Harvard and giving me candy), Carol Davis (for all of the laughs and Black joy), Dionne Clarke (for being such a great friend), Hannah Belcher (for her humor and candor even when times were tough), and the entire Harvard University Physics staff. I would also like to thank Professor Jenny Hoffman and Professor Bob Westervelt for their continued support in my outreach efforts over the years to increase the number of underrepre-

sented minority women in STEM through the Women+ of Color Project.

Harvard's Center for Nanoscale Systems (CNS) staff has also played a huge role success. In particular, I would like to thank the Executive Director Dr. Bill Wilson. Unfortunately, Bill will now need to find a replacement graduate student to randomly knock on his door and bother him in his office after I leave. In addition, I would like to thank Dr. Arthur McClelland, Nicki Watson, and the rest of the CNS staff. Without CNS, many of experiments in this thesis would not have been done. In particular, without Nicki I wouldn't have such nice TEMs! I would also like to thank Dr. Bridget Queenan, the Executive Director of the Quantitative Biology Initiative (QBio) at Harvard, for consistently supporting me during my PhD. She has been a champion from day one.

I would like to thank my mentors over the years for their continued support. In particular, Professor Nia Imara. As a postdoc at Harvard, she was the only other Black woman in physics and astronomy that I knew on campus and she has had a huge influence on my journey. I would also like to thank Dr. Christopher Moore. Both Chris and Nia reminded me that I didn't need to compromise my mind, my culture, and my values to fit in. In addition, I'd like to thank Professor Devin Walker who was the first African American physicist to graduate from Harvard University. He is also a Memphis native like myself! On the days I felt that someone from my background could not make it through, I remembered Professor Devin Walker. In addition, I would like to thank the broader Black physicist community for their continued support over the years.

I would like to thank the many faculty and staff members from my alma mater – Wesleyan University – as well. In particular I'd like to thank Professor Christina Othon, my first research advisor during my undergraduate and the only person willing to take me into her lab. Without her, I would not have become a scientist. Thank you for taking a chance on me. In addition, I'd like to thank Professor Illesanmi Adeboya and Professor Meredith Hughes. You both took time to help me get better at math and physics so that I could pursue my goals. I'd also like to thank Ronnie Hendrix and the Ronald E. McNair Program. I would not have had the support and funding needed to pursue research at Wesleyan without this program. And lastly, I'd like to thank Clifford Thornton and Dr. Renee Johnson Thornton. When my parents were not around, you were essentially my extended family in undergraduate. Thank you for all of the support, thank you for the encouragement, thank you for the trips to church on Sundays, thank you for the love and support you have constantly given me. I would also like to thank my master's thesis advisor from Fisk University, Professor Arnold Burger, for supporting my goals no matter how big they were.

Lastly, I'd like to thank my family for their support: they are my core. My younger sister, Marinda, for always reminded me to "never take anything from anyone." My older brother, Marion, for constantly reminding me not let anyone place fear in me and deter me from my goals. My mother, Nell, for reminding me every week not to "let the devil steal my joy." And my father, Marion LaTroy Alexander Williams, who taught me to always do the right thing even if no one else is. I want to thank my parents especially because they grew up during times with fewer opportunities available to them because of the color of their skin. I often think about what could have been different for them had they been provided the same opportunities that I was provided. I am thankful for their contributions to our society and their sacrifices over the years to ensure I became the first in our family to receive my PhD. Words cannot express my level of gratitude for all of the support I have had over the years.

1

Introduction

1.1 VIRUS SELF-ASSEMBLY AND STRUCTURE

Self assembly is believed to be an important part of the life-cycle of small icosahedral RNA viruses like bacteriophage MS2. These viruses replicate themselves by first attaching themselves to a host cell and releasing their RNA. The host then produces more viral RNA synthesized by replicase enzymes that the viral RNA encodes. Coat proteins are then translated from these RNAs by the ribosomes

within the host cell. The coat protein and RNA assemble into viruses that then escape the cell and infect others. The process by which well-formed capsids form from a random distribution or “soup” of RNAs and coat proteins within a crowded cell is not well understood^{42,29}.

If we take the virus and separate it into its native coat protein and RNA, then mix them outside of the host, we can often produce virus-like particles (VLPs)⁴⁷. Because this assembly process happens *in vitro*, it appears that the viral capsid represents a free-energy minimum. Minimization of free energy explains *why* capsids are able to form *in vitro* but not *how* they assemble around their RNA. This process is interesting because the structures that are formed after assembly are non-trivial. Wild-type MS2 contains 178 identical protein monomers in addition to the maturation protein. Each of these monomers has to assemble into either hexamers and pentamers, as in a soccer ball, to enclose the RNA. The resulting shape has icosahedral symmetry and can be described by a triangulation number $T = 3$ ²⁷. When MS2 coat protein and RNA assemble *in vitro*, the MS2 virus-like particle (VLP) that is formed contains 180 identical coat protein monomers (or 90 identical coat protein dimers) in quasi-equivalent positions around the viral genome. The structures of these VLPs also have icosahedral symmetry with pentameric and hexameric configurations.

Previous experiments using interferometric scattering have shown that the assembly pathway for MS2 VLPs follows classical nucleation theory (CNT)¹⁵. It was shown that when there is a sufficient amount of coat protein available, a critical nucleus forms on the surface of tethered MS2 RNA before fully enclosing the strand of RNA. These experiments have also shown that at increasing coat protein concentration structures consisting of multiple partial capsids grow on the surface of a single strand of RNA. Although these experiments give insight into the assembly pathway, they were done with RNA tethered to a surface. The RNA may therefore have had a restricted conformational space. Furthermore, this study provided limited structural information about the assemblies.

There are many theories for how viruses avoid malformed structures, but the first step towards testing these theories is to understand how assembly can go wrong. Doing this requires us to modify

variables such as the coat protein and RNA concentrations, as well as the ionic strength and temperature, which affect the protein-protein interactions and protein-RNA interactions. In general these interactions have been found to play an important role in the assembly of MS2 VLPs^{3,39}. However, little is known about how the assembly products vary with these parameters.

My aim is to understand how different buffer conditions affect the products of assembly. With this understanding, we can infer how MS2 VLPs can minimize their free energy under favored buffer conditions. The structures that form may provide insight into intermediate assembly steps or additional configurations that may be explored (and potentially avoided) within a host cell. These insights may also shed light on the pathway that RNA viruses use to assemble themselves and avoid malformed structures.

1.2 OVERVIEW

This thesis demonstrates how dynamic light scattering, gel electrophoresis, transmission electron microscopy, and other complementary methods that include course-grained simulation, zeta-potential analysis, and plaque assays can reveal a phase diagram for MS2 VLP self-assembly and characterize the various configurations that MS2 VLPs may explore. In addition, we uncover information about what physical parameters matter most in MS2 VLP assembly. I begin by exploring the role that coat protein concentration plays in MS2 VLP bulk assembly (Chapter 2). We conclude in this chapter that the nucleation rate is controlled by the coat protein and increasing concentration results in structures that are different than the well-formed VLPs found in other studies⁴⁷. I then go on to further explore how assembly may be driven by electrostatic interactions (Chapter 3). We conclude in this chapter that electrostatic interactions might act as a driver for nucleation and growth. I then explore the effects of RNA concentration, ionic strength, and temperature on the structure of MS2 VLPs (Chapter 4).

2

Effect of Capsid Protein Concentration on MS₂ Coat Protein Assembly Around RNA

In this chapter I describe our experiments on the assembly of bacteriophage MS₂ coat protein around the RNA genome of MS₂ in bulk. These experiments represent a portion of the main focus for my thesis work. I further explore the potential for malformation under other buffer conditions in Chapter

3.

The work described in this chapter was conducted jointly with Rees F. Garmann and Andreas Neophytou.

2.1 INTRODUCTION

For positive-sense RNA viruses to replicate, coat proteins must assemble around the viral RNA to form new virus particles⁵. Certain features of this assembly process can be replicated *in vitro*, in the absence of host-cell factors^{12,2,21}. For example, virus-like particles (VLPs) can be assembled from solutions of the coat protein and RNA of bacteriophage MS2. Though these VLPs lack the maturation protein required for infectivity, they nonetheless have the same structure and size as wild-type MS2 virus particles⁴⁷, lending support to the premise that virus assembly is driven by free-energy minimization.

However, the assembly process itself and the conditions under which it leads to well-formed structures are not yet well understood. In MS2, most previous work on this question has focused on the role of specific interactions between coat protein and the viral RNA and, in particular, on a portion of the RNA called the translational operator²³. Studies on R17, a virus closely related to MS2, have shown that the overall yield of assembled VLPs decreases if the RNA does not contain this translational operator^{3,39}. Nonetheless, assembly proceeds in its absence, perhaps due to non-specific interactions of the coat protein with the RNA³. Therefore, specific RNA-protein interactions might affect the assembly rate and yield but do not seem to be essential to the assembly process.

While these studies have established the importance of RNA-protein interactions to the assembly process, they did not directly reveal the assembly pathway itself. More recent work involving interferometric scattering microscopy (iSCAT), a technique that can image individual VLPs as they form, shows that MS2 VLPs assemble by a nucleation-and-growth pathway¹⁵ at near-neutral pH, salt concentrations on the order of 100 mM, and micromolar coat-protein concentrations. In this pathway,

a critical nucleus of proteins must form on the RNA before the capsid can grow to completion. The size of the critical nucleus, estimated to be less than six coat-protein dimers, is associated with a free-energy barrier. Taken together with the previous experiments on the role of the RNA sequence, these results show that MS2 assembly is a heterogeneous nucleation process, in which the nucleation rate is likely controlled by two factors: RNA-protein interactions and the coat-protein concentration.

Arguably, the coat-protein concentration has a larger role in controlling the morphology of the VLPs than does the nature of the RNA-protein interactions (at least in *in vitro* experiments, where the coat-protein concentration is typically fixed). The iSCAT experiments¹⁵ showed that very few VLPs are formed at low ($1 \mu\text{M}$) concentration of MS2 coat-protein dimers, while well-formed capsids form at higher concentrations, and so-called "monster" capsids, consisting of multiple partially formed capsids on a single strand of RNA, form at even higher concentrations (several μM). These results suggest that the nucleation barrier, which controls the nucleation rate, depends sensitively on the coat-protein concentration. At low concentration, the nucleation rate is too small for capsids to form within the experimental time frame; at high concentration, the nucleation rate is so high that multiple nuclei can form on a single RNA strand, resulting in monster capsids. However, these studies examined only a few protein concentrations, and they were performed at low RNA concentration relative to protein.

Here we use bulk assembly experiments to determine the assembly products of MS2 coat protein and MS2 RNA as a function of coat-protein concentration. We characterize the assembly products using three techniques: gel electrophoresis, dynamic light scattering (DLS), and transmission electron microscopy (TEM). This three-pronged approach allows us to corroborate results and test hypotheses about how the assembly products form. Gel electrophoresis and TEM provide qualitative data that we use to determine the size and structure of the assembly products, and DLS provides quantitative information about their size distributions. With these methods, we show that as protein concentration increases, the morphologies transition from well-formed VLPs to "monster" capsids to RNA-protein

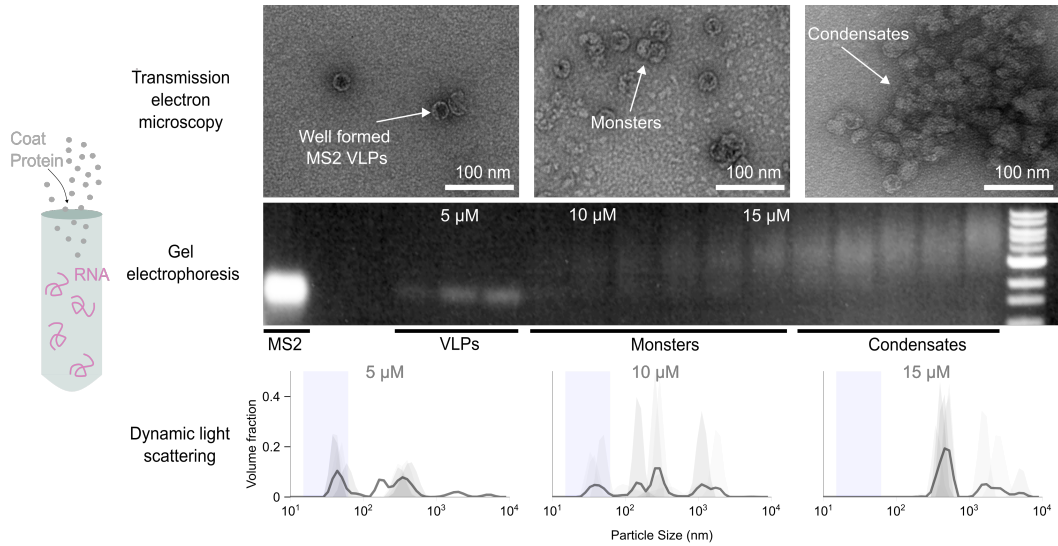


Figure 2.1: Overview of experiments and results. We mix MS2 coat protein with MS2 RNA to make a solution with 50 nM RNA concentration and varying coat-protein concentration. The transmission electron microscopy images, gel electrophoresis measurements (full image is shown in Fig. 2.2), and dynamic light scattering results demonstrate a transition from well-formed MS2 VLPs to monster capsids to RNA-protein condensates with increasing coat protein concentration. The main text and subsequent figures elaborate on all of these results.

condensates consisting of large networks of RNA and protein. We explain these results, which are summarized in Fig. 2.1 and discussed in more detail below, in terms of a nucleation-and-growth model for capsid assembly.

2.2 RESULTS

2.2.1 OVERVIEW OF EXPERIMENTAL APPROACH

Briefly, our experimental procedure consists of combining 50 nM MS2 RNA with purified MS2 coat-protein dimers at concentrations ranging from 2.5 to 30 μM (see Section 2.5 for full details). For reference, a full VLP has an icosahedral capsid with a triangulation number of 3, corresponding to 180 coat proteins. At 50 nM RNA concentration, a coat-protein dimer concentration of 5 μM therefore corresponds approximately to the stoichiometric ratio of coat proteins to RNA in a full VLP. We

work with dimer concentrations instead of monomer concentrations because MS2 coat proteins are typically dimerized in solution³⁰. After mixing the RNA and coat protein, we then wait 10 min to allow assembly to occur, after which we add RNase to digest any excess MS2 RNA that is not encapsidated. We then characterize the resulting assembly products with gel electrophoresis, DLS, and TEM (see Section 2.5).

2.2.2 RESULTS FROM GEL ELECTROPHORESIS

We first qualitatively characterize the size and composition of the assembly products using agarose gel electrophoresis. We use both Southern blotting to detect RNA and western blotting to detect coat protein in our samples. For comparison, we also characterize wild-type MS2, MS2 RNA, and digested MS2 RNA (see Section 2.5).

The most striking feature of the gel is a band that runs at the same position as wild-type MS2 but with a brightness that increases from 2.5 to 7.5 μM coat-protein dimers and then suddenly decreases at 8.7 μM (see highlighted region in Fig. 2.2). We interpret this increase and sudden decrease as follows. Near the stoichiometric ratio (approximately 5 μM dimers to 50 nM RNA), well-formed VLPs assemble, with more VLPs forming at higher protein concentration. Above 7.5 μM , the sharp decrease in brightness indicates that far fewer well-formed MS2 VLPs assemble. Instead, as indicated by the spreading of the band toward to the upper part of the gel, the assembly products at dimer concentrations greater than 7.5 μM are larger than the wild-type particles, and they contain both RNA and protein.

We also see that at coat-protein dimer concentrations higher than 7.5 μM , the intensity of the diffuse band increases with increasing concentration (Fig. 2.2). The increase in brightness and change in the center position of this band suggest that the amount of large assembly products increases at the expense of the wild-type-sized products. At 15 μM , the diffuse band no longer overlaps with the band corresponding to wild-type-size VLPs. For dimer concentrations beyond 15 μM , some of the

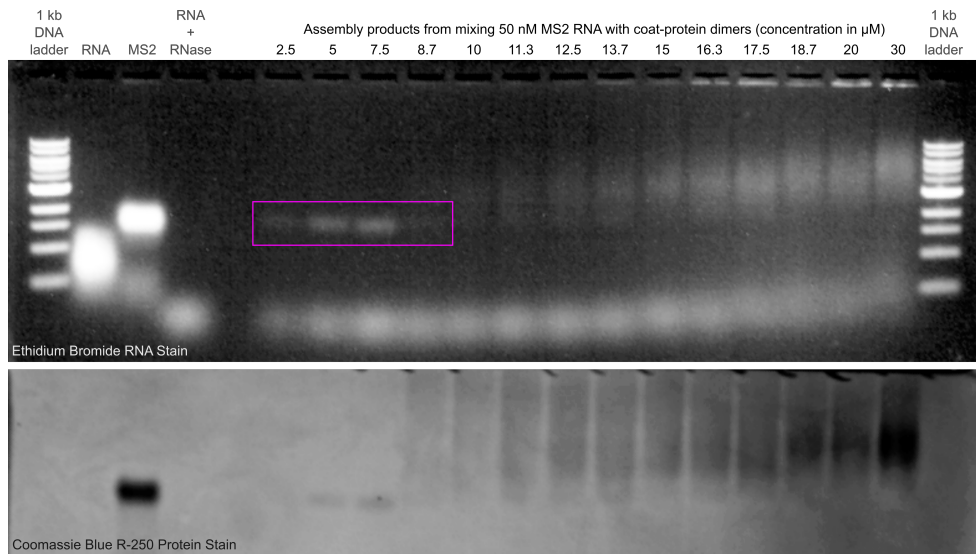


Figure 2.2: Images of agarose gels used to characterize the assembly products. We first stain with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat protein (bottom image). Lanes 1 (leftmost lane) and 20 (rightmost lane) show a DNA ladder. Lanes 2–4 show three controls: MS2 RNA (lane 2), wild-type MS2 capsids (lane 3), and MS2 RNA treated with RNase (lane 4). The other lanes show the results of gel electrophoresis on samples prepared with coat-protein dimer concentrations ranging from 2.5 μM to 30 μM . The region highlighted in purple shows that the amount of wild-type-sized products increases as dimer concentration increases from 2.5 to 7.5 μM and then decreases sharply at 8.7 μM .

assembly products are so large that they are trapped near the top of the agarose gel.

The transition from a bright to a diffuse band might represent a transition from well-formed VLPs to either malformed structures or aggregates of capsids. The gels by themselves cannot confirm either hypothesis, since they reveal only that the assembly products all contain RNA and that they increase in size with increasing coat-protein concentration. We therefore turn to dynamic light scattering and transmission electron microscopy experiments, as described below.

2.2.3 RESULTS FROM DYNAMIC LIGHT SCATTERING

To quantify the sizes of the assembly products, we use DLS with numerical inversion methods. These methods yield the size distributions of assembly products in both number and volume bases (see Section 2.5).

At coat-protein dimer concentrations $7.5 \mu\text{M}$ and below, we observe in both the number and volume distribution a peak at or near the size of wild-type MS2 particles (see shaded bands in Fig. 2.3; we expect some variation in the location of this peak because the inversion of the autocorrelation function is sensitive to noise). This peak is accompanied by peaks at larger sizes, unlike the size distribution for wild-type MS2, which consists of only one peak. At coat-protein dimer concentrations above $7.5 \mu\text{M}$, the peak corresponding to size of wild-type MS2 particles decreases until it disappears (in the volume-basis distributions) at $12.5 \mu\text{M}$. At concentrations of 15 and $20 \mu\text{M}$, we observe a single peak corresponding to much larger assembly products. Overall, we observe that the average size of the assembly products increases with increasing protein concentration (Fig. 2.3).

The DLS data support our interpretation of the gel-electrophoresis data. Specifically, both the DLS and gel data show that the proportion of VLPs with sizes corresponding to the wild-type size decreases with concentration above $7.5 \mu\text{M}$ coat-protein dimer concentration, whereas only larger products form at high concentration. The DLS data additionally show that the size of these larger products is on the order of several hundred nanometers.

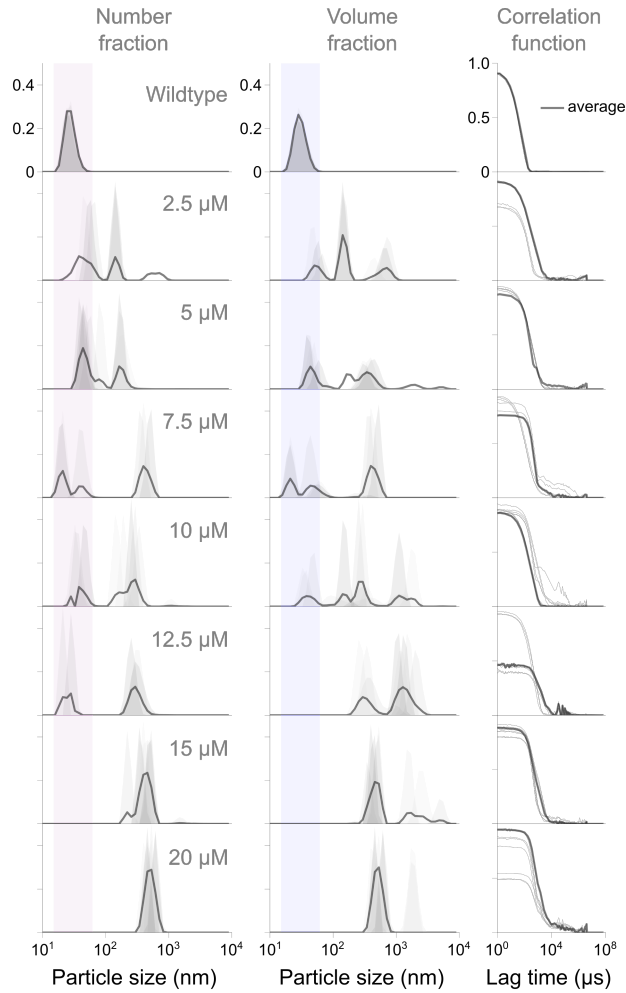


Figure 2.3: Plots of size distributions of wild-type MS2 virus particles and VLPs assembled *in vitro* at 50 nM concentration of free RNA and varying coat-protein dimer concentrations. The distributions are inferred from dynamic light scattering measurements. The first column shows the size distribution on a number basis, the second column shows the size distribution on a volume basis, and the third column shows the measured autocorrelation functions. Light gray peaks in the distributions show the results from eight individual experiments. Dark gray peaks show the results inferred from the average autocorrelation function. The autocorrelation functions for each individual measurement are shown in light grey in the plots at right, and the average is shown in dark gray.

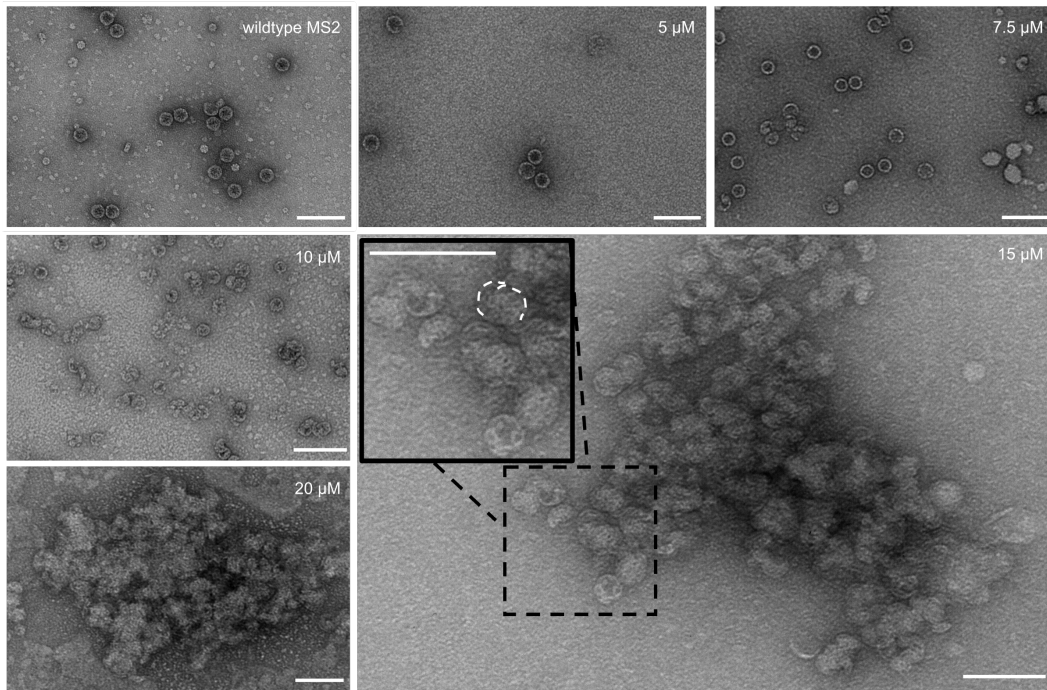


Figure 2.4: TEM images from negatively stained samples of wild-type MS2 particles and products of assembly at varying coat-protein dimer concentrations. At concentrations less than $10\ \mu\text{M}$, most particles have the shape and size of wild-type capsids. At higher concentrations, we observe clusters of partially formed capsids that increase in size with concentration. The dotted line in the inset of the $15\ \mu\text{M}$ image shows the outline of one such partial capsid. All scale bars are 100 nm.

However, the DLS data also show peaks corresponding to particles larger than wild-type at concentrations less than $10\ \mu\text{M}$. We do not see evidence of such particles in the gel data. These peaks may correspond to weakly-bound clusters of well-formed MS2 VLPs that are observable in the DLS experiments but fall apart during gel electrophoresis (see Fig. 2.2). Because DLS does not provide any structural information, we turn to TEM to test this hypothesis and characterize the structures of the assembly products.

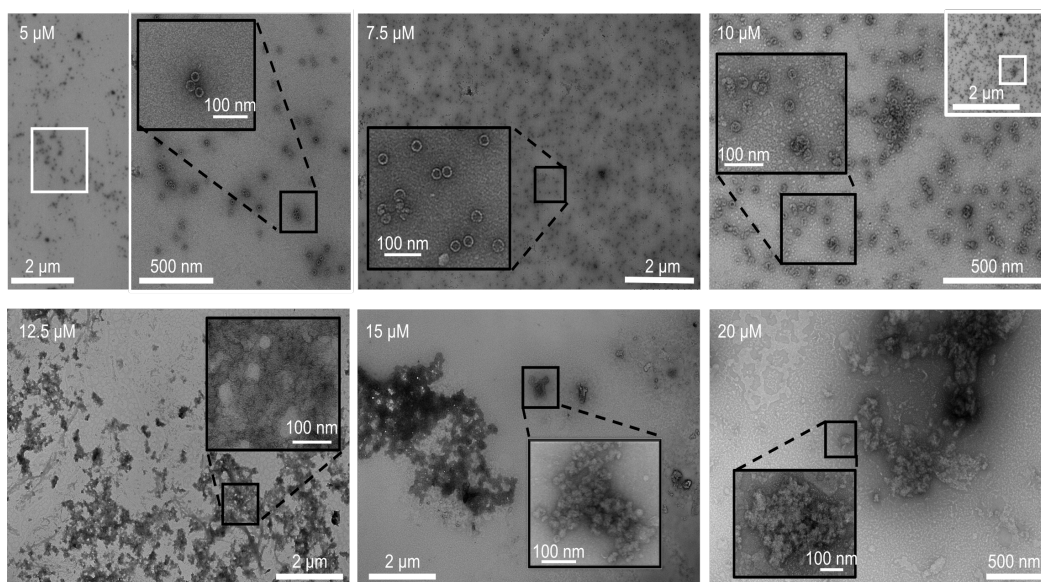


Figure 2.5: TEM images of the same samples as shown in Fig. 2.4, but at different magnifications for each sample. The low magnification images of the samples at 12.5 μM coat-protein dimer concentration and higher show that the assembled structures are several micrometers in size.

2.2.4 TRANSMISSION ELECTRON MICROSCOPY (TEM) EXPERIMENTS

TEM images of negatively stained samples show that most of the assembly products at dimer concentrations 7.5 μM and below are well-formed MS2 VLPs (see Figs. 2.4 and 2.5), with some malformed VLPs and clusters of MS2 VLPs, consistent with the larger sizes present in the DLS-derived size distributions. At a concentration of 10 μM , we observe malformed particles that consist of partially formed capsids. These structures are similar to the so-called “monster” particles observed in turnip-crinkle-virus assemblies⁴⁵ and, more recently, in MS2 assembly experiments¹⁵. At concentrations above 15 μM we observe what appear to be large aggregates of partially formed capsids (Fig. 2.4). These aggregates are micrometer-sized (Fig. 2.5), comparable to the sizes seen in the DLS distributions (Fig. 2.3).

2.3 DISCUSSION

Our measurements show that coat-protein concentration plays an important role in the morphology of the assembly products of MS2 RNA and coat protein. At low coat-protein dimer concentrations (less than $7.5 \mu\text{M}$), gel electrophoresis, DLS, and TEM all point to the formation of MS2 VLPs that are the same size as wild-type MS2. These structures appear to be well-formed, consistent with previous studies⁴⁷. At higher concentrations (between 7.5 and $10 \mu\text{M}$), we observe monster particles consisting of a few partial capsids. At even higher concentration ($12.5 \mu\text{M}$), results from gel electrophoresis, DLS, and TEM point to the formation of large aggregates several hundred nanometers in size and containing many partial capsids.

Whereas the observation of well-formed VLPs and even monsters is consistent with previous studies, the observation of large aggregates at high protein concentrations has not, to our knowledge, been studied in detail. These structures are interesting not only because they contain many partially formed capsids, but also because they contain RNA, as shown by our gel electrophoresis measurements. Because the structures contain both RNA and protein, we term them “condensates.” Below, we consider several hypotheses that might explain the formation of the condensates, with the aim of understanding what they reveal about the assembly pathway of the virus.

One hypothesis is that the condensates arise primarily by aggregation of coat proteins. However, gel electrophoresis, DLS, and TEM experiments show no evidence of coat-protein aggregation in the absence of RNA, even at $15 \mu\text{M}$ dimer concentration. In addition, gel electrophoresis data at high coat-protein concentrations show that the condensates contain both RNA and coat protein. While such structures might arise if the aggregation of the coat proteins were rapid, trapping the RNA inside, the absence of aggregation of coat protein at high concentrations is evidence against this hypothesis.

Another hypothesis is that the RNA-protein condensates arise from an *en masse* pathway⁴⁰, in which the interactions between the coat proteins and RNA are strong compared to the inter-protein

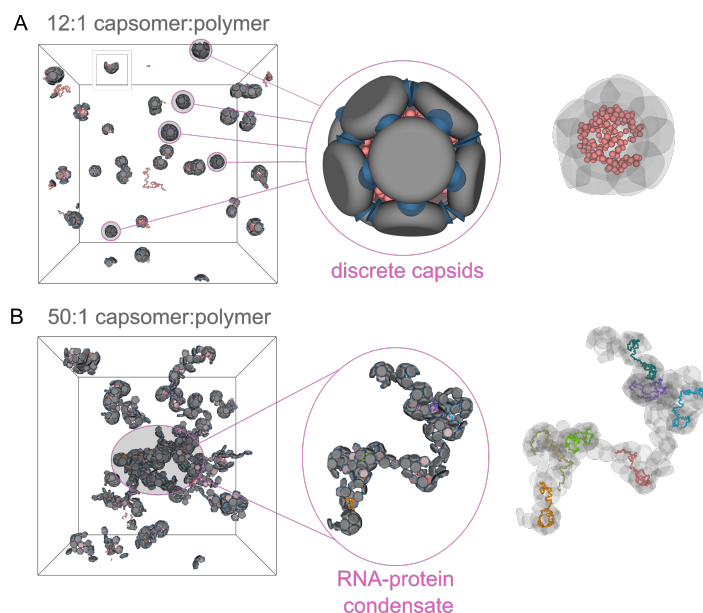


Figure 2.6: Representative snapshots from Monte Carlo simulations of capsomer-and-polymer systems with (A) a 12:1 ratio of capsomer to polymer and (B) a 50:1 ratio. The volumes of the simulation cells and reduced temperature of the simulations are identical (see Section 2.5). Each capsomer is modelled as a hard disk with five sticky patches on its rim to mediate capsomer-capsomer interactions, and a large sticky patch on its face to mediate capsomer-polymer interactions. At the low capsomer:polymer ratio, discrete particles form containing 12 capsomers surrounding a polymer (a select few are highlighted in panel A). At the high capsomer:polymer ratio, extended clusters form, mediated by capsomer-capsomer interactions, that contain multiple polymer chains and which resemble the RNA-protein condensates found in the experiments. In the images at right, the capsomers are made transparent so that the polymer conformations are more visible.

interactions. In this scenario, coat proteins would first decorate the RNA, potentially leading to a heteroaggregate of RNA and protein. This scenario would account for the presence of RNA in the condensates. However, it is at odds with the observation of a nucleation-and-growth pathway¹⁵ at lower concentrations. If nucleation and growth happens at low concentrations, we expect that increasing the protein concentration should not cause a transition to an *en masse* pathway but should instead primarily change the nucleation rate.

We therefore consider the hypothesis that a nucleation-and-growth pathway is operative at all coat-protein concentrations. At low concentrations, where the assembly products are well-formed VLPs, our study provides no direct evidence for this pathway, but as noted above, previous direct imag-

ing measurements have shown that the assembly is nucleated. The nucleation-and-growth pathway does, however, account for the monster particles seen at intermediate protein concentrations. These structures, which consist of multiple partial capsids, can form when more than one nucleation event happens on a single RNA strand; indeed, we expect that the probability of multiple nucleation events should increase with the coat-protein concentration.

The question then is how the condensates form with the RNA trapped inside. To understand whether and how a nucleated pathway might lead to such condensates, we turn to simulations. We perform coarse-grained patchy-particle simulations in which the capsomers are represented as patchy hard disks, and the RNA is represented as a free polymer approximately 14 times the length of each capsomer (see Fig. 2.6 and Section 2.5), such that each polymer can be encapsidated by 12 capsomers. Although the experimental system is more complex – specifically, an MS2 VLP consists of 90 coat-protein dimers, yielding a $T = 3$ structure – the simulation is designed to test the hypothesis that nucleation and growth can lead to a condensate. To this end, we tune the interactions so that the assembly is nucleated, as seen in Fig. 2.6.

Whereas at low capsomer concentrations the simulations show the assembly of well-formed capsids containing polymer, at high capsomer concentrations they show the assembly of large networks of polymers and partial capsids, just as in the experiments. Interestingly, the simulations show that these networks consist of multiple polymer strands that are bridged by networks of partial capsids (Fig. 2.6). A partial capsid attached to one polymer molecule can connect, through other capsomers, to a partial capsid attached to a different polymer molecule. This observation provides a plausible explanation for why the condensates seen in the experiments can grow to be so large even in the absence of significant coat-protein aggregation: the coat proteins may be able to bridge partial capsids on different RNA molecules.

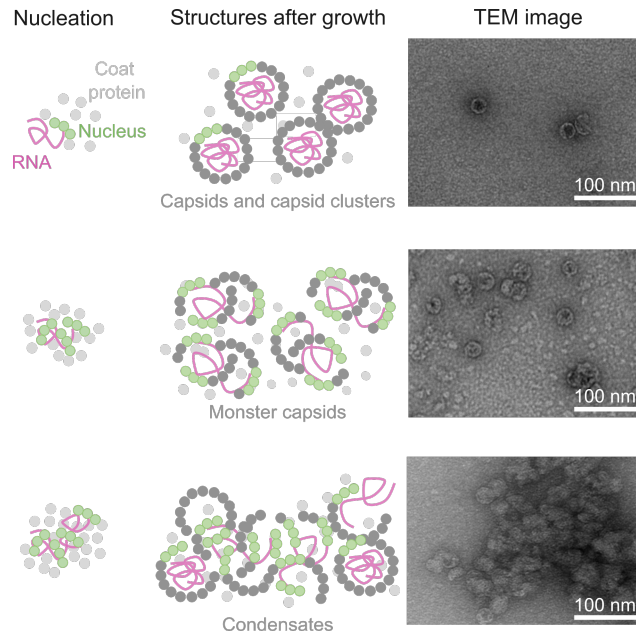


Figure 2.7: Overview of the experiments and results, showing how coat protein concentration affects the structures of MS2 coat protein assembly around RNA. We mix 50 nM MS2 RNA with varying amounts of MS2 coat protein. We find three kinds of assembly products. The first (top row), which forms at 5 μM coat protein, consists of well-formed capsids that in some cases appear to stick together. The second (middle row), which forms at 10 μM protein concentration, consists of monster particles: multiple partial capsids that decorate the RNA. The third (bottom row), which forms at 15 μM protein concentration, consists of condensates of RNA and coat protein. Diagrams at left and middle illustrate potential pathways for formation of these products. Images at right are from transmission electron microscopy (TEM) of negatively stained MS2 VLPs.

2.4 CONCLUSIONS

Our experiments and simulations show that all the morphologies we observe as a function of coat-protein concentration – well-formed capsids, monster capsids, and RNA-protein condensates – can be understood as outcomes of a nucleation-and-growth process. Other hypotheses, including coat-protein aggregation and *en masse* assembly, do not account for all of our results.

In a nucleation-and-growth assembly pathway, the primary effect of increasing the coat-protein concentration is to increase the nucleation rate. If the growth rate depends more weakly on concentration than does the nucleation rate, we can explain the formation of all the structures we observe as follows. If the timescale of nucleation is short compared to the time for a nucleus to grow into a full capsid, multiple nuclei can form on the same RNA strand. When these nuclei grow, they do not in general form a closed capsid, but instead form partial capsids. At moderate concentrations, these partial capsids remain disconnected as they grow, leading to monster capsids. At high concentration, coat proteins can bridge partial capsids on different RNA molecules, leading to the formation of large RNA-protein condensates.

There remain a few questions to be resolved in future studies. One question is how the RNA is spatially distributed in the condensates, and specifically whether the bridging mechanism observed in the simulations is operative in the experiments. Another question is what happens at concentrations between those at which well-formed capsids form and monster capsids form. DLS and TEM experiments suggest that at these intermediate protein concentrations, small clusters of well-formed capsids are present. The driving force for the formation of these clusters is not clear, but they might arise when a single RNA molecule spawns multiple nuclei that each form a full (or nearly full) capsid. In this situation, the RNA would connect the capsids into a “multiplet” structure¹⁴. This hypothesis would explain why we do not see evidence for small clusters of well-formed capsids in the gel data; in the gel experiments, we treat the assembled samples with RNase, which could cleave any RNA link

between capsids. Fluorescent microscopy experiments could help answer all of these questions.

Our work might also inform models of the assembly pathway, particularly those based on the law of mass action^{55,54,36,53,50}, in which the concentration of coat proteins plays a critical role. Further experiments that quantify how the nucleation rate depends on the coat-protein concentration would help connect these models to the morphological observations we present here. From a more practical perspective, our work helps establish constraints on concentration for the production of MS2 VLPs form. Such VLPs are used to encapsulate materials for drug delivery^{28,13,20}.

2.4.1 ACKNOWLEDGMENTS

We thank Amy Barker and Peter Stockley at the University of Leeds for initial stocks of MS2 and *E. coli* cells. We thank Tim Chiang, Amelia Paine, Aaron Goldfain, and Danai Montalvan for helpful scientific discussions. This research was partially supported by a National Science Foundation (NSF) Graduate Research Fellowship under grant number DGE-1745303 and by NSF through the Harvard University Materials Research Science and Engineering Center under NSF grant number DMR-2011754, by the National Institute of General Medical Sciences of the National Institutes of Health under grant K99GM127751, by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University under NSF grant number 1764269, and by the Harvard Quantitative Biology Initiative. This work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF grant number ECCS-2025158. The work was also performed in part at the Harvard University Bauer Core Facility. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

2.5 MATERIALS AND METHODS

All materials were used as received. Buffers were prepared as follows:

- Assembly buffer: 42 mM Tris, pH 7.5; 84 mM NaCl; 3 mM acetic acid, 1 mM EDTA
- TNE buffer: 50 mM Tris, pH 7.5; 100 mM NaCl, 1 mM EDTA
- TE buffer: 10 mM Tris, pH 7.5; 1 mM EDTA
- TAE buffer: 40 mM Tris-acetic acid, pH 8.3; 1 mM EDTA

2.5.1 VIRUS GROWTH, CULTIVATION, AND STORAGE

We purify wild-type bacteriophage MS2 as described by Strauss and Sinsheimer⁴⁶. In brief, we grow MS2 virus particles by infecting *E. coli* strain C3000 in minimal LB Buffer, and we remove *E. coli* cell debris by centrifugation at 16700g for 30 min. We then use chloroform extraction to purify the solute containing the virus. We extract the purified virus particles by density gradient centrifugation in a cesium chloride gradient. We store the purified virus at 4 °C at a concentration of 10¹¹ plaque-forming units (pfu) in Tris-NaCl-EDTA or TNE buffer (50 mM Tris, 100 mM NaCl, 5 mM EDTA) at pH 7.5. We determine the concentration of virus by UV-spectrophotometry (NanoDrop 1000, Thermo Scientific) using an extinction coefficient of 8.03 mL/mg at 260 nm.

2.5.2 COAT-PROTEIN PURIFICATION AND STORAGE

MS2 coat-protein dimer is purified as described by Sugiyama, Herbert, and Hartmant⁴⁷. Wild-type bacteriophage MS2 is suspended in glacial acetic acid for 30 min to denature the capsid, separate it into protein dimers, and precipitate the RNA. We then centrifuge the sample at 10000g and collect the supernatant, which contains coat-protein dimers. We filter out the glacial acetic acid with 20 mM

acetic acid buffer through 3-kDa-MWCO sterile centrifugal filters (Millipore Sigma, #UFC500324) five times. This process removes the glacial acetic acid to prevent further denaturing of the coat-protein dimers. We then determine the concentration of our coat-protein dimers by measuring the absorbance with the Nanodrop Spectrophotometer (Thermo Fisher) at 280 nm. We store the MS2 coat protein at 4°C in a 20 mM acetic acid buffer. We measure the absorbance at 260 nm as well to detect residual RNA. In our experiments, we use only purified protein with an absorbance ratio (protein:RNA) above 1.5 to avoid RNA contamination.

2.5.3 RNA PURIFICATION AND STORAGE

Wild-type MS2 RNA is purified using a protocol involving a Qiagen RNeasy Purification Kit Mini (Qiagen, #7400450). We take 100 µL of MS2 stored in TNE buffer and mix with 350 µL of buffer RLT (a lysis buffer) to remove the coat-protein shell. We add 250 µL of ethanol to our sample and mix to precipitate the RNA. We then transfer our sample to a 2 mL RNeasy Mini spin column (provided by the Qiagen Purification Kit) that is placed in a collection tube. We then centrifuge at 10000g for 15 s and discard the flow through. We add 500 µL of buffer RPE (to remove traces of salts) to the spin column and centrifuge for 15 s at 10000g. We discard the flow-through. We then add 500 µL of buffer RPE once more to the spin column and centrifuge for 2 min at 10000g. We place the spin column upside down into in a fresh 1.5 mL collection tube (provided in the purification kit) to collect the RNA trapped in the spin column. We add 50 µL of TE buffer to the spin column and centrifuge at 10000g for 1 min to collect the RNA. We measure the RNA concentration using a Nanodrop spectrophotometer by measuring the absorbance at 260 nm and using an extinction coefficient of 25.1 mL/mg. We store the purified MS2 RNA at -80°C in Tris-EDTA (TE) buffer at neutral pH (7.5).

2.5.4 GEL ELECTROPHORESIS AND ANALYSIS

For gel electrophoresis experiments, we mix 15 μL of sample with 4 μL of glycerol and load into a 1% agarose gel in assembly buffer consisting of 5 parts Tris-NaCl-EDTA (TNE) buffer (50 mM Tris, 100 mM NaCl, 10 mM EDTA, pH 7.5) to 1 part 20 mM acetic acid buffer. We use Southern blotting with Ethidium Bromide (EtBr) to stain the RNA and to detect the presence of MS2 RNA. Western blotting with Coomassie Blue R-250 is used to detect the presence of MS2 coat protein. The combination of these staining methods allow us to confirm the presence of both MS2 RNA and MS2 coat protein within the resulting assemblies. We place three control samples in lanes 2 through 4 that include MS2 RNA at 50 nM concentration (lane 2), wild-type MS2 at 50 nM concentration (lane 3), and 50 nM concentration of digested MS2 RNA genome (lane 4) resulting from the addition of RNase A. These controls allow us to compare the sizes of our assembly products to systems of known sizes. We can also determine whether the samples consist of MS2 VLPs formed during assembly or excess strands of MS2 RNA. We place our assembly products in lanes 6 through 19. These samples are loaded and run at 21 $^{\circ}\text{C}$ at 100 V voltage for 40 min and visualized using a Biosystems UV Imager (Azure, #AZ1280).

2.5.5 DYNAMIC LIGHT SCATTERING (DLS) AND ANALYSIS

We use dynamic light scattering (Malvern ZetaSizer Nano ZS by Malvern Panalytical) to determine the size distribution of particles that assemble at 50 nM MS2 RNA concentration and coat-protein dimer concentrations of 2.5, 5, 7.5, 10, 12.5, 15, and 20 μM . We also characterize the wild-type samples. The size distributions are determined using the regularization method provided by the software provided with the instrument³³.

2.5.6 TRANSMISSION ELECTRON MICROSCOPY (TEM) AND ANALYSIS

For transmission electron microscopy, we negatively stain the samples with 2% aqueous uranyl acetate on 200 mesh carbon-coated copper TEM grids (Polyscience, TEM-FCF200CU), then image them with the Hitachi 7800 TEM provided by the Center for Nanoscale Systems at the Science and Engineering Complex (CNS-SEC) at Harvard University. Each image was taken at 20 kV, 50 kV, and 100 kV voltage.

As a control, we mix 15 μM MS2 coat-protein dimers in assembly buffer. This control is done to ensure that capsid-like or VLP-like structures do not form in the absence of MS2 RNA.

2.5.7 RNA AND COAT-PROTEIN BULK ASSEMBLY EXPERIMENTS

Wild-type MS2 RNA genome at a concentration of 50 nM is mixed with varying concentrations of MS2 coat-protein dimers ranging from 2.5 μM to 30 μM and left at room temperature (21 °C) for 10 min. The assembled virus-like particles are then characterized using gel electrophoresis, dynamic light scattering (DLS), and transmission electron microscopy (TEM).

2.5.8 COARSE-GRAINED MODEL FOR CAPSID ASSEMBLY

To model the assembly of the capsids and condensates, we developed a patchy particle model that captures the essential features of the system. The model consists of capsomers and a polymer chain, which is used to model the RNA. A capsid is constructed from 12 subunits, each having C_{5v} symmetry, where the center of each subunit sits on the vertex of an icosahedron.

CAPSOMER-CAPSOMER INTERACTIONS

We coarse-grain the capsomeric building blocks as hard disks decorated with five attractive patches arranged such that they form a pentagon, giving the capsomers C_{5v} symmetry. We model the hard core

of the capsomers using the hard oblate spherocylinder model previously used to investigate the phase behavior of discotic liquid crystals⁷. In this model, the particles have a cylindrical core with diameter σ and thickness L . Additionally, there is a toroidal rim with a tube diameter equal to that of the thickness of the cylindrical core. We are then able to define the total diameter of the particles as $D = \sigma + L$, and their aspect ratio as $L^* = L/D$. The pair interaction between two spherocylinders is infinite if the shortest distance between the cylindrical cores of the particles is less than L , and zero otherwise. We compute the shortest distance between two hard oblate spherocylinders using the algorithm outlined in Ref. 7.

We model the attractive patches by adapting the Kern-Frenkel model²⁶, where each attractive site involves a square-well attraction modulated by an angular factor between the patches, with a half-angle θ . The angular factor is unity only when the patches are oriented such that the vector connecting the centers of the two particles passes through the patches on their surfaces, and zero otherwise. The width of the square well, δ_{cap} , determines the range of the attraction between the patches relative to the particle diameter. The depth of the square well, ε_{cap} , governs the strength of the bonds. Owing to the anisotropic geometry of the particles, we must include an additional parameter φ that defines the orientation of the patches relative to the normal of the flat surface of the cylindrical core of the particle.

The total pair potential defining capsomer-capsomer interactions is then

$$v_{ij}^{\text{cap}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = v_{ij}^{\text{ohsc}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) + \sum_{\alpha, \beta}^5 v_{\alpha\beta}^{\text{sw, cap}}(\mathbf{r}_{\alpha\beta}) f(\mathbf{r}_{\alpha\beta}, \hat{\mathbf{n}}_{i,\alpha}, \hat{\mathbf{n}}_{j,\beta}), \quad (2.1)$$

where $r_{ij} = |\mathbf{r}_{ij}|$ is the center-to-center distance between particles i and j , Ω_i and Ω_j define the orientation of the cylindrical core of particles i and j , $\hat{\mathbf{n}}_{i,\alpha}$ is a normalized vector defining the orientation of patch α on particle i (similarly, $\hat{\mathbf{n}}_{j,\beta}$ is a normalized vector defining the orientation of patch β on particle j), and $\mathbf{r}_{\alpha\beta}$ is the separation vector connecting patches α and β .

The term u_{ij}^{ohsc} is the hard oblate spherocylinder pair potential

$$v_{ij}^{\text{ohsc}}(r_{ij}, \Omega_i, \Omega_j) = \begin{cases} \infty & \text{if } d_{ij} < L \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

with d_{ij} representing the shortest distance between the two particles. The term $v_{\alpha\beta}^{\text{sw,cap}}$ is a square-well potential:

$$v_{\alpha\beta}^{\text{sw,cap}}(r_{\alpha\beta}) = \begin{cases} -\varepsilon_{\text{cap}} & \text{if } r_{\alpha\beta} \leq (1 + \delta_{\text{cap}})\sigma \\ 0 & \text{otherwise,} \end{cases} \quad (2.3)$$

and $f(\mathbf{r}_{\alpha\beta}, \hat{\mathbf{n}}_{i,\alpha}, \hat{\mathbf{n}}_{j,\beta})$ is the angular modulation factor,

$$f(\mathbf{r}_{\alpha\beta}, \hat{\mathbf{n}}_{i,\alpha}, \hat{\mathbf{n}}_{j,\beta}) = \begin{cases} 1 & \text{if } \hat{\mathbf{n}}_{i,\alpha} \cdot \hat{\mathbf{r}}_{\alpha\beta} > \cos \theta \text{ and } \hat{\mathbf{n}}_{j,\beta} \cdot \hat{\mathbf{r}}_{\beta\alpha} > \cos \theta \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

The reference orientation of particle i is taken to be $\Omega_i = (0, 0, 1)$; that is, the normal of the flat face of the oblate spherocylinder is aligned with the z -axis of the global coordinate frame. We then define the reference position of the first patch on particle i as $p_{i,1} = (\sigma/2, 0, 0)$ and the position of each other patch as a rotation about the z -axis of the local coordinate frame of the particle such that $p_{i,n} = R_z(n2\pi/5) \cdot p_1$, where $n = 2, 3, 4, 5$ and R_z is a rotation matrix defining a clockwise rotation of angle $\phi = n2\pi/5$ about Ω_i . The orientation of patch α on particle i is then $\hat{\mathbf{n}}_{i,\alpha} = \sin(\phi)\Omega_i + (2 \cos(\phi)/\sigma)p_\alpha$, where ϕ is the angle between $\hat{\mathbf{n}}_{i,\alpha}$ and the plane containing the flat face of the oblate spherocylinder.

POLYMER-POLYMER INTERACTIONS

Each RNA molecule is modeled as a flexible self-avoiding polymer – that is, as a chain of hard-spheres, where neighboring beads in the chain are connected by a harmonic spring^{10,25,24}:

$$v_{\text{poly}}(r_{ij}) = \kappa(r_{ij} - \sigma_b l_b)^2, \quad (2.5)$$

where r_{ij} is the distance between beads i and j (where $j = i - 1, i + 1$), κ sets the strength of the harmonic spring, σ_b is the hard-sphere diameter of the beads in the polymer chain, and l_b sets the equilibrium bond length between neighboring beads.

CAPSOMER-POLYMER INTERACTIONS

Finally, we allow for interaction between the capsomers and the polymer *via* an attractive patch on the surface of the capsomer. The orientation of the patch is aligned with that of the oblate spherocylinder. The beads of the polymer and the capsomer then interact *via* an attractive square-well interaction, plus a hard-core repulsion between their respective cores. The pair interaction when particle i is a capsomer and particle j is a bead of a polymer chain is

$$v_{ij}^{\text{cap-pol}}(\mathbf{r}_{ij}, \Omega_i) = v_{ij}^{\text{hc}}(\mathbf{r}_{ij}, \Omega_i) + v_{ij}^{\text{sw, cap-pol}}(\mathbf{r}_{ij})g(\mathbf{r}_{ij}, \Omega_i), \quad (2.6)$$

where v_{ij}^{hc} is the hard-core interaction

$$v_{ij}^{\text{hc}}(\mathbf{r}_{ij}, \Omega_i) = \begin{cases} \infty & \text{if } d_{ij} < (L + \sigma_b)/2 \\ 0 & \text{otherwise,} \end{cases} \quad (2.7)$$

where d_{ij} is the shortest distance between the capsomer and polymer bead. The steps to compute this distance are as follows:

1. Compute the projection of the polymer bead onto the plane spanned by the cylindrical core of the capsomer: $\mathbf{r}_j^{\text{proj},i} = \mathbf{r}_{ij} - (\mathbf{r}_{ij} \cdot \Omega_i)\Omega_i$
 - (a) If $r_j^{\text{proj},i} \leq \sigma/2$ then the bead lies over the cylindrical core of the capsomer, so the shortest distance vector between the two particles is then $\mathbf{d}_{ij} = \mathbf{r}_{ij} - \mathbf{r}_j^{\text{proj},i}$.
 - (b) Otherwise, the closest point of the capsomer to the bead lies on its edge. The shortest distance vector between the two particles is then $\mathbf{d}_{ij} = \mathbf{r}_{ij} - (\sigma/2)\hat{\mathbf{r}}_j^{\text{proj},i}$.

The term $v_{ij}^{\text{sw,cap-pol}}$ is the square-well interaction between the patch on the face of the capsomer and the polymer bead:

$$v_{ij}^{\text{sw,cap-pol}}(\mathbf{d}_{ij}) = \begin{cases} -\varepsilon_{\text{cap-pol}} & \text{if } d_{ij} \leq (1 + \delta_{\text{cap-pol}})\sigma \\ 0 & \text{otherwise} \end{cases}, \quad (2.8)$$

and $g(\mathbf{r}_{ij}, \Omega_i)$ is the angular modulation factor for the attractive capsomer-polymer interaction:

$$g(\mathbf{r}_{ij}, \Omega_i) = \begin{cases} 1 & \text{if } \cos^{-1}(\mathbf{r}_{ij} \cdot \Omega_i / r_{ij}) < \pi/2 \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

2.5.9 MONTE CARLO SIMULATIONS

We carry out two sets of Monte Carlo simulations in the NVT ensemble using the model outlined above. For both simulations, we set the volume to be $V = 600000\sigma^3$, the reduced temperature to be $k_B T / \varepsilon_{\text{cap}} = 0.12$ (where k_B is the Boltzmann constant, which is taken to be equal to one), and the

number of polymer chains $N_{\text{poly}} = 30$, with each polymer chain consisting of $l_{\text{poly}} = 150$ beads. In one simulation, there are $N_{\text{cap}} = 360$ capsomers, and in the other, there are $N_{\text{cap}} = 1500$ capsomers.

We take σ to be the unit of length and ε_{cap} to be the unit of energy. We then choose the parameters defining the system to be $L = 0.5\sigma$, $\delta_{\text{cap}} = 0.2$, $\theta = 25^\circ$, $\varphi = 25^\circ$, $\kappa = 100\varepsilon_{\text{cap}}$, $\sigma_b = 0.2\sigma$, $l_b = 1.05\sigma_b$, $\varepsilon_{\text{cap-pol}} = 0.2\varepsilon_{\text{cap}}$, and $\delta_{\text{cap-pol}} = 0.3\sigma$.

We carry out all Monte Carlo simulations with systems contained in a cubic box under periodic boundary conditions, using the minimum image convention. Each capsomer is treated as a rigid body for which the orientational degrees of freedom are represented by quaternions. The potential energy is calculated using a spherical cutoff of 1.7σ , and a cell list is used for efficiency. Each Monte Carlo cycle consists of N translational or rotational single-particle or cluster moves, chosen at random with equal probabilities.

*“Did you ever observe to whom the accidents happen?
Chance favors only the prepared mind.”*

Louis Pasteur

3

The Role of Electrostatics in the Formation of MS₂ Coat Protein Scaffolds

In this chapter, I describe our experiments on the assembly of MS₂ coat protein around wild-type MS₂ virus. These experiments were originally intended to test the aggregation of wild-type MS₂ virus by charge screening from additional coat-protein dimers in Chapter 2, but something much cooler happened!

The work described in this chapter was conducted jointly with Kellianne Kornick and Amelia W. Paine.

3.1 INTRODUCTION

Large icosahedral viruses like the herpes simplex virus (HSV)²² need scaffold proteins to assemble their outer shell. This outer coat-protein shell protects the genome (and other structures) within the virus. Unlike large viruses, small icosahedral viruses such as bacteriophage MS2 do not need these scaffolding proteins to assemble and can spontaneously assemble a single coat-protein shell *in vitro* around their native RNA⁴⁷. The existence of scaffold proteins is believed to be necessary for the assembly of large viruses³². However, little is known about how viral capsids use scaffolds to assemble.

Simulations based on elasticity theory show that larger viruses will not form an additional layer without a scaffold present³². Instead, they assemble into viruses with smaller radii. It has also been shown experimentally that the absence of VP19C and VP23 molecules (known to control scaffolding in HSV capsids) results in the formation of smaller HSV capsids ($T = 7$)⁴³. These studies demonstrate that a scaffold needs to be present for multiple layers of coat protein to form in certain viruses, but they do not explain what causes these coat proteins to assemble in the first place.

In this study, we assemble MS2 coat protein around wild-type MS2 as a function of coat-protein concentration in buffer with a neutral pH of 7. We then characterize the resulting structures using transmission electron microscopy (TEM), gel electrophoresis, plaque assay analysis, confocal microscopy, and dynamic light scattering (DLS). TEM and gel electrophoresis help us determine the size and structure of the assembly products while plaque assays and confocal microscopy help us determine whether the additional coat-protein layer covers the maturation protein of the wild-type virus. DLS provides quantitative information about the size distribution of the assemblies. We measure the zeta potential to estimate the charge on the outside of the wild-type MS2 and understand the electrostatic

interactions. The combination of these methods shows that as we add coat protein, additional (or multiple) capsid shells form around the wild-type MS2. We discuss the importance of electrostatic forces as a potential driver for the nucleation and growth of these additional shells.

3.2 RESULTS

3.2.1 OVERVIEW OF EXPERIMENTAL APPROACH

Our experimental approach consists of combining 50 nM wild-type bacteriophage MS2 with purified MS2 coat-protein dimers at concentrations ranging from 2.5 μM to 20 μM (see Section 3.5 for full details). Wild-type MS2 is known to have icosahedral symmetry with a triangulation number of 3, 178 coat proteins and an additional maturation protein known to assist in infection. This capsid protects the native RNA genome. After mixing the wild-type MS2 virus and coat protein, we wait 10 min to allow assembly to occur. We then characterize the resulting assembly products with TEM, gel electrophoresis, DLS, plaque assay analysis, confocal microscopy, and zeta potential analysis.

3.2.2 TRANSMISSION ELECTRON MICROSCOPY

First we determine structural information about the assembly products using transmission electron microscopy. We negatively stain our sample to visualize the assembly products and characterize the cross section. Surprisingly, we observe an additional layer of coat protein around the wild-type MS2 (see Fig. 3.1) at high protein concentrations. At concentrations below 5 μM coat-protein dimers, we do not see this additional shell form. Above 5 μM , we see that the coat-protein forms a conformal shell around the wild-type virus.

One potential hypothesis for the formation of these additional shells is that the maturation protein drives this process. To test this hypothesis, we perform the same experiments with MS2 virus-like particles (VLPs) that do not contain the maturation protein. We assemble MS2 virus-like particles

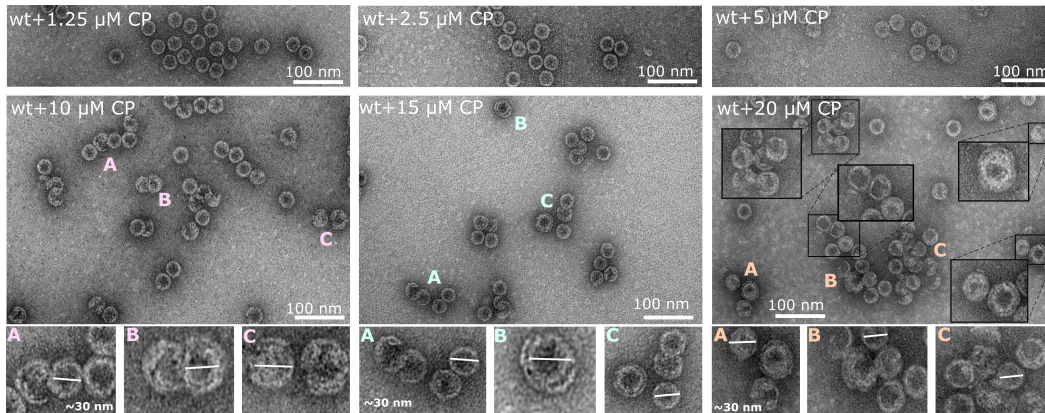


Figure 3.1: TEM images of negatively stained samples of wild-type MS2 particles mixed with varying coat-protein dimer concentrations. At concentrations less than $5 \mu\text{M}$, we begin to see the formation of an additional coat-protein layer and, in some cases, clusters of particles. The number of layers and clusters increases with increasing concentration.

(VLPs) as described in Section 2.5 by taking 50 nM concentration of purified MS2 RNA with $15 \mu\text{M}$ concentration of MS2 coat-protein. As with the wild-type virus, we observe an additional coat-protein layer around the assembled VLPs.

3.2.3 RESULTS FROM GEL ELECTROPHORESIS

We use agarose gel electrophoresis to further characterization the size and composition of the assembly products. As in Chapter 2, we use both Southern blotting to detect the RNA and western blotting to detect the coat protein. As a control, we stain wild-type MS2 (see Section 2.5.4). The gels show an additional sharp band that corresponds to a larger size than the wild-type across all protein concentrations. As we increase protein concentration, we begin to see a spreading of this additional band (see Fig. 3.1).

3.2.4 RESULTS FROM DYNAMIC LIGHT SCATTERING

To quantify the sizes of the assembled particles, we use DLS with numerical inversion methods that provide size distributions of our samples in both intensity and volume bases (See Section 2.5.5). In the

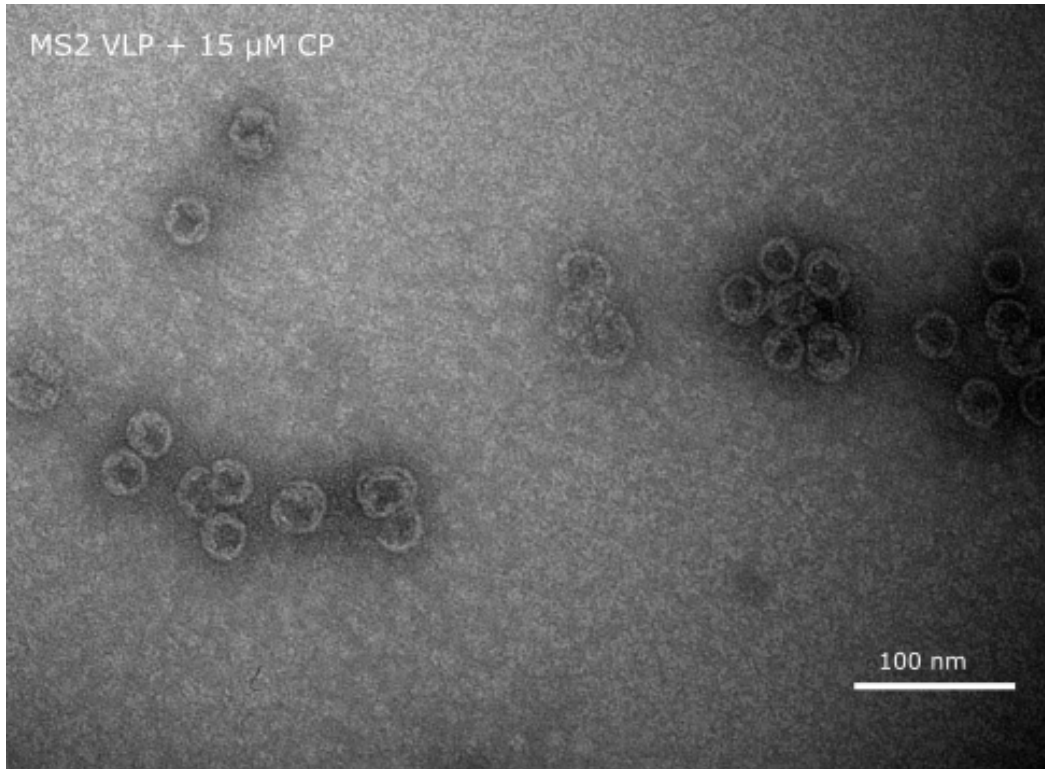


Figure 3.2: TEM image of a negatively stained sample of MS2 virus-like particles (VLP) mixed with 15 μ M coat-protein dimers. Again, we see the formation of an additional coat protein layer that forms around the VLP.

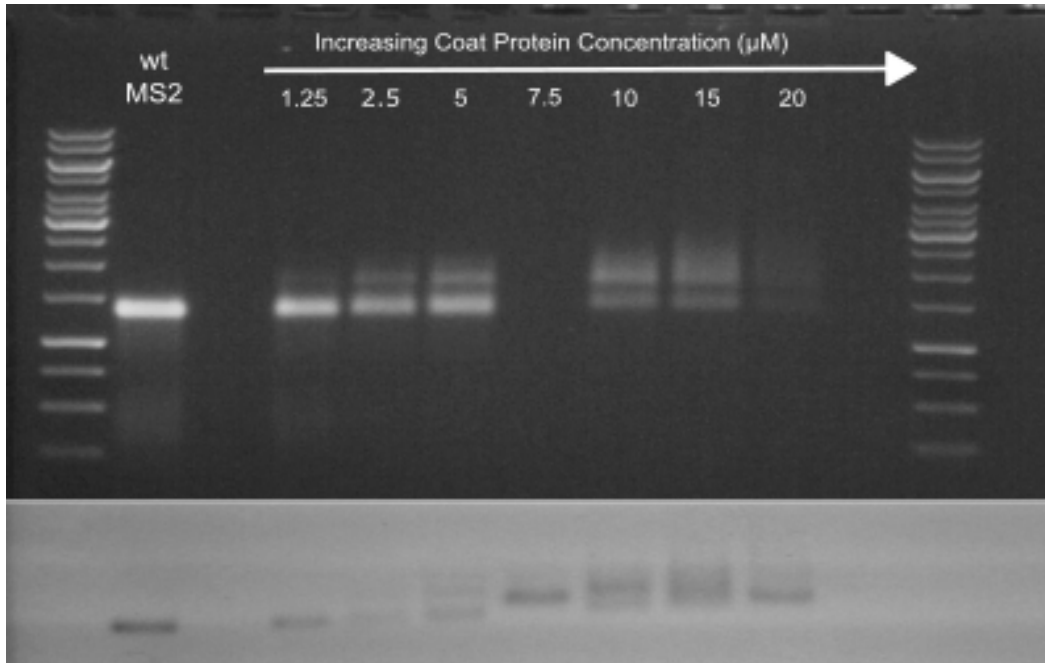


Figure 3.3: Images of agarose gels used to characterize samples. We first stain the samples with ethidium bromide to detect the RNA (top image) and then with Coomassie Blue R-250 to detect protein (bottom image). Lanes 1 (leftmost lane) and 12 (rightmost lane) show a DNA ladder. Lane 2 shows a control sample of wild-type MS2 capsids. Lanes 4–6 and 8–10 show samples prepared with coat-protein dimer concentrations from 1.25 to 20 μM (lane 7 represented a concentration of 7.5 μM but there was a mistake in the preparation of the sample). An additional sharp band appears that broadens with increasing coat-protein concentration.

wild-type sample and in samples at coat-protein concentrations below $5 \mu\text{M}$, we observe two peaks, one corresponding to the size of wild-type particles and the other corresponding to much larger particles, perhaps dust or aggregates (see Fig. 3.4). The presence of the larger particles is evidenced directly by the long-time delay in the autocorrelation functions. At coat-protein concentrations of $5 \mu\text{M}$ and above, we observe an additional shoulder at a size (about 40 nm) that is slightly larger than that of the wild-type particles. In particular, we see two differentiated peaks at $15 \mu\text{M}$ concentration, one at 24 nm and the other at 44 nm, though at $20 \mu\text{M}$ these peaks are not resolved.

We associate the larger peak with clusters of particles containing multiple layers seen in TEM (Fig. 3.1), such as the clusters of two wild-type particles enclosed by an additional layer of coat protein shown in particle B of Fig. 3.1 at $15 \mu\text{M}$. These clusters would also account for the second band seen in the gel electrophoresis measurements.

3.2.5 INFECTIVITY OF MULTI-SHELL CAPSIDS

To test if the additional protein layers fully cover the wild-type virus (including the maturation protein), we use confocal microscopy to image the binding of MS2 to the F-pili of *E. coli*. We expect that a particle in which the maturation protein is covered by an additional shell will not bind.

We mix 50 nM MS2 labeled with Alexa Fluor 647 (AF 647) in assembly buffer (see Section 2.5) with 0, 5, 10, and $15 \mu\text{M}$ concentration of MS2 coat-protein dimers. We wait 10 min to ensure shells grow on the labeled wild-type virus. We then take $2 \mu\text{L}$ of the sample and mix it with $10 \mu\text{L}$ of dilute *E. coli* cells before imaging with confocal microscopy. Our raw confocal images show labeled MS2 bound to *E. coli* cells, while the histogram-adjusted data shows the unbound MS2 particles (Fig. 3.5A). The ratio of the intensity of the bound particles to the total image intensity yields the *fluorescence fraction ratio*, which quantifies the binding efficiency. We plot the fluorescence fraction ratio as a function of the concentration of coat-protein dimers added to the assembly reaction, and we observe that the ratio decreases with increased concentration (Fig. 3.5B). This result suggests that the fraction of particles in

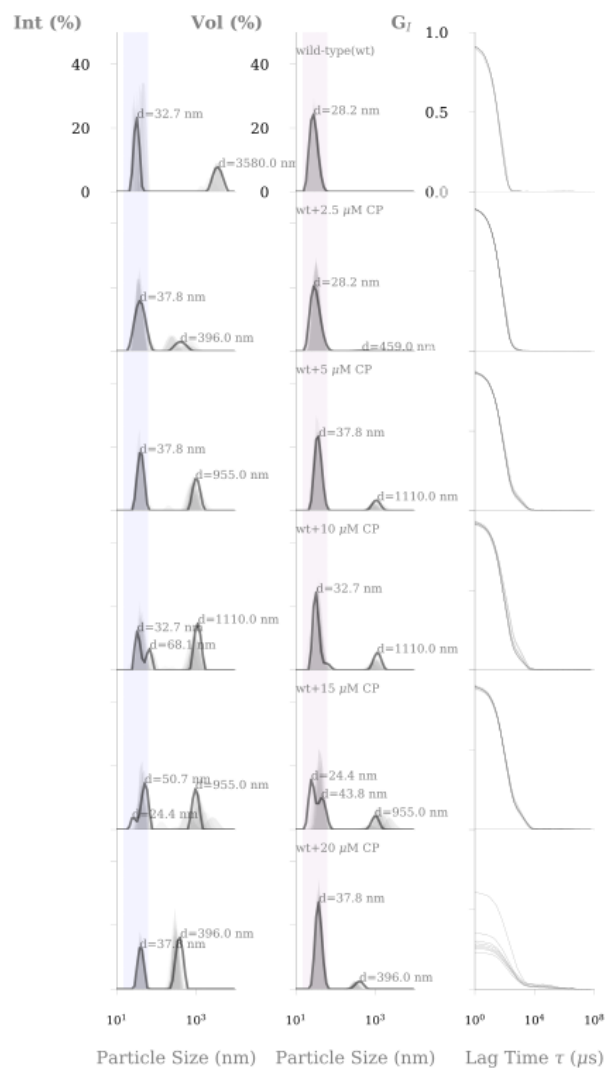


Figure 3.4: Plots of size distributions of wild-type MS2 virus (top row) and of 50 nM MS2 wild-type virus mixed with varying concentrations of coat-protein dimers (bottom five rows). The distributions are inferred from dynamic light scattering measurements. The first column shows the size distribution on an intensity basis, the second column shows the size distribution on a volume basis, and the third column shows the measured autocorrelation functions. The light grey peaks in the distribution show the results from eight individual experiments. Dark gray peaks show the results inferred from the average autocorrelation function. The autocorrelation function for each individual measurements is shown in light gray. The diameter d corresponding to each peak is noted on each distribution.

| Experiment | Zeta Potential (ζ) | Error (mV) |
|-------------------------------|----------------------------|------------|
| 15 μ M Coat Protein (CP) | 6.97 mV | ± 0.23 |
| wild-type MS2 | -16.88 mV | ± 1.28 |
| wild-type MS2 + 15 μ M CP | -9.44 mV | ± 0.12 |

Table 3.1: Table of zeta potential measurements of MS2 coat-protein dimers, wild-type MS2, and wild-type MS2 mixed with 15 μ M coat-protein dimers.

which the additional layer covers the maturation protein increases with increasing coat-protein dimer concentration.

The confocal images quantify the fraction of particles bound to F-pili but not the infectivity, which also depends on the maturation protein. To measure the infectivity we perform a plaque assay on our samples (see Section 3.5 for details). We observe that increasing protein concentration results in a decrease in the number of plaques that are formed (Figs. 3.5C and 3.6). The decrease in plaques indicates a decrease in the overall infectivity, which likely arises because the additional shell prevents the maturation protein from binding to the F-pili. The results of the plaque assays agree qualitatively with the results of our confocal analysis.

3.2.6 ELECTROPHORETIC MOBILITY MEASUREMENTS

To test whether electrostatic interactions might drive the assembly of the additional coat-protein layers, we measure the zeta potential of the wild-type virus and the wild-type virus after the addition of 15 μ M coat-protein dimers (Table 3.1). We find that the wild-type virus carries a negative charge on its surface, despite the RNA being fully encapsulated by positively charged coat proteins. Furthermore, the addition of coat protein to the wild-type structures results in particles that still retain a negative surface charge.

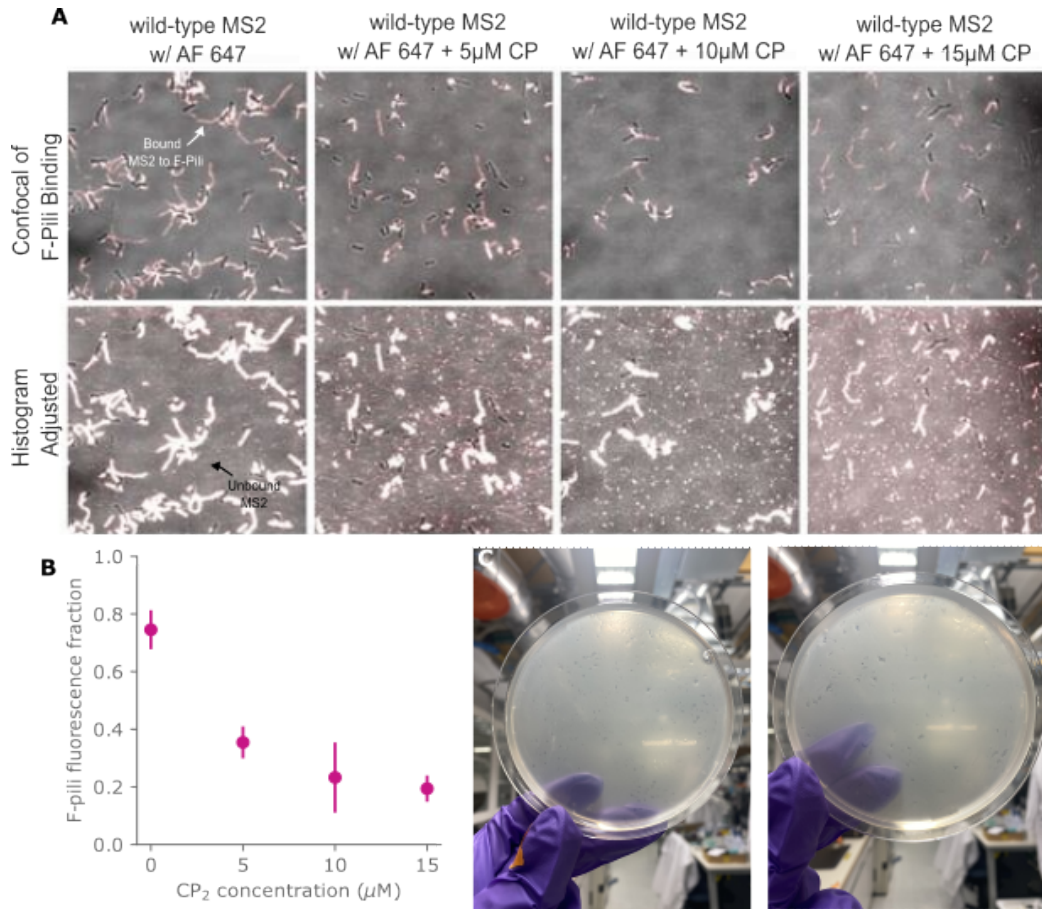


Figure 3.5: Analysis of infectivity of wild-type MS2 and wild-type virus with added coat protein. A. Confocal micrographs (raw images and histogram adjusted images to show individual pixels) of wild-type MS2 labeled with AF 647 mixed with 5 μ M, 10 μ M, and 15 μ M coat-protein dimers (CP₂). B. A plot of the F-pili fluorescence fraction (a measure of the ratio of the number of particles bound to F-pili to the total number of particles). C. Photographs of a plaque assay of wild-type MS2 (left) and wild-type MS2 added to 15 μ M coat-protein dimers (right). There are more plaques (empty regions) in the left image (wild-type MS2), indicating that the additional coat protein reduces the infectivity but does not eliminate it.

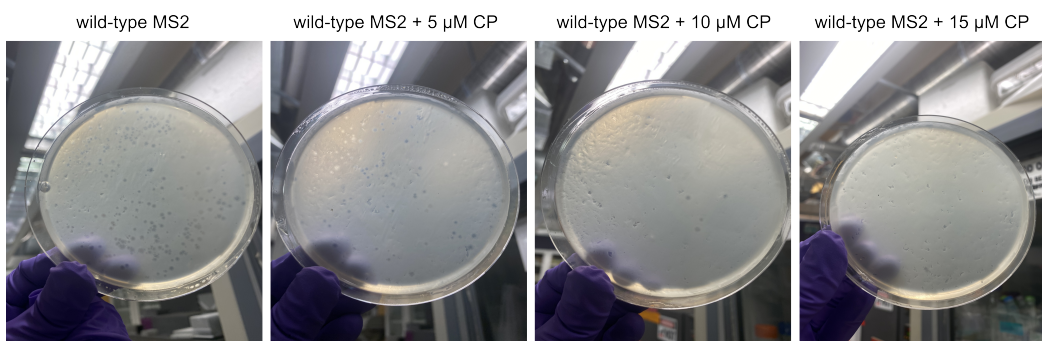


Figure 3.6: Images of plaque assays done with 50 nM wild-type MS2 virus mixed with 0, 5, 10, and 15 μM concentration of coat-protein dimers. We wait 10 min for assembly to occur. These experiments show a decrease in the number of plaques, which indicates a decrease in the overall infectivity.

3.3 DISCUSSION

Overall, our measurements show that MS2 coat-protein dimers can form additional layers around wild-type MS2. In addition, our experiments show that the assembly of these shells is likely driven by electrostatic interactions between the coat proteins, which are positively charged, and the wild-type virus, which is negatively charged. The formation of multiple coat-protein layers driven by electrostatic interactions has also been observed in *in vitro* experiments on cowpea chlorotic mottle virus (CCMV)³¹. However, these experiments were performed at a pH near the isoelectric point of CCMV. In addition, CCMV has been shown to fully shield the overall charge of the RNA at neutral pH⁵¹. By contrast, our experiments, which involve MS2 coat protein and MS2 RNA, are done at a neutral pH of 7. Wild-type MS2 is known to have a different isoelectric point (at pH = 3.9) than CCMV³⁴ and to contain less positively charged coat proteins⁴. Furthermore, the observation of a threshold protein concentration (5 μM dimers) for the formation of additional layers suggests that these layers assemble through a nucleation-and-growth mechanism, just as VLPs do¹⁵.

The formation of these additional coat-protein layers could reveal additional information about the assembly pathway of the MS2 virus. Below, we discuss several hypotheses that may explain the

formation of multilayer MS2 capsids. Our aim is to understand the most important driving forces for the assembly of the shells, in addition to revealing a potential driver for scaffolding in virus self-assembly. One hypothesis is that the maturation protein acts as a nucleation site. However, TEM experiments with MS2 VLPs have shown that this is not the case. MS2 VLPs do not contain the maturation protein⁴⁷ yet at 15 μ M concentration of coat-protein dimers, we observe the formation of additional layers around the VLPs (Fig. 3.2). Furthermore, multilayered wild-type particles are still infectious, albeit to a lower extent than the native capsid, suggesting that the maturation protein is not covered in all assemblies. Furthermore, we note that before adding coat protein to the VLPs, we treat the VLPs with RNase, which would remove any excess RNA that, if unremoved, might help nucleate additional layers. We therefore conclude that the assembly of the additional layers proceeds by nucleation and growth and is driven purely by electrostatic interactions (Fig. 3.7). Compared to the assembly of MS2 coat proteins around MS2 RNA, the assembly of MS2 coat proteins around wild-type MS2 capsids cannot involve specific RNA-protein interactions since the RNA is completely encapsidated. Instead, the assembly is likely driven by the electrostatic attraction between the coat proteins, which are positively charged, and the capsid, which is negatively charged. The negative charge on the wild-type capsid may be a result of the encapsidated, negatively charged RNA being incompletely screened^{9,37}.

3.4 CONCLUSION

Our experiments show that electrostatic interactions drive the nucleation and growth of additional coat-protein layers around wild-type MS2 viruses and MS2 VLPs. These results highlight the importance of electrostatics in the encapsidation process. Because the formation of additional layers does not involve any specific RNA-protein interactions, it is plausible that the assembly of wild-type MS2 capsids and VLPs also relies on electrostatics to a larger extent than previously appreciated. Further-

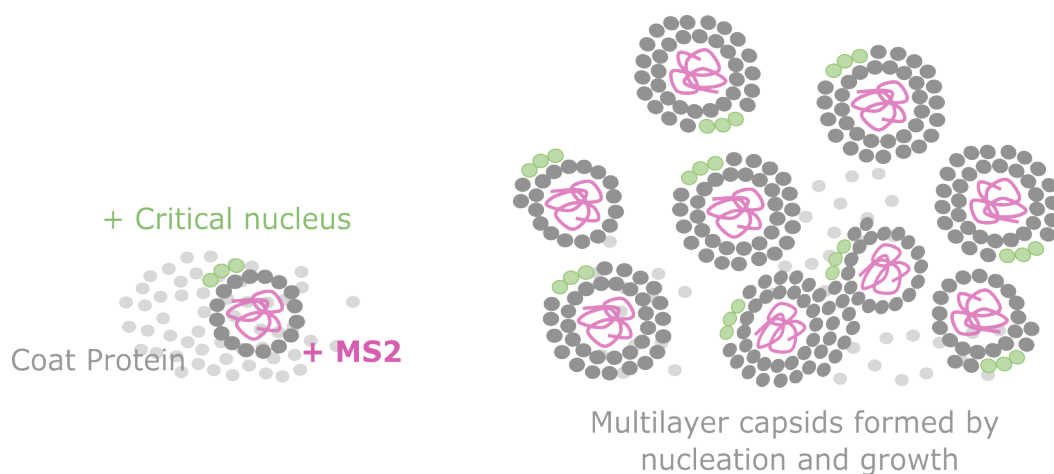


Figure 3.7: Diagram showing pathway by which multi-layer MS2 capsids form.

more, electrostatic interactions are a potential candidate for understanding how larger viruses drive assembly around their scaffolds. A better understanding of how electrostatic interactions control nucleation and growth of capsids may lead to more effective encapsidation in applications such as drug-delivery²⁰ and the fabrication of nano-materials¹. One could imagine the usage of viral scaffolds to form large VLP structures for applications in medicine and environmental science.

3.4.1 ACKNOWLEDGMENTS

We thank Amy Barker and Peter Stockley at the University of Leeds for initial stocks of MS2 and *E. coli* cells. We thank Tim Chiang, Amelia Paine, Aaron Goldfain, and Danai Montalvan for helpful scientific discussions. This research was partially supported by a National Science Foundation (NSF) Graduate Research Fellowship under grant number DGE-1745303 and by NSF through the Harvard University Materials Research Science and Engineering Center under NSF grant number DMR-2011754, by the National Institute of General Medical Sciences of the National Institutes of Health under grant K99GM127751, by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University under NSF grant number 1764269, and by the Harvard Quantitative

Biology Initiative. This work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF grant number ECCS-2025158. The work was also performed in part at the Harvard University Bauer Core Facility. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

3.5 MATERIALS AND METHODS

Buffers were prepared as follows:

- Bottom agar: 3 g Tryptone; 0.3 g yeast extract; 2.4 g sodium chloride; 4.5 g agar; 300 mL Milli-Q water
- Top agarose: 2 g Tryptone; 0.2 g yeast extract; 2.92 g sodium chloride; 0.9 g agarose; 200 mL Milli-Q water

Other buffers are described in Section 2.5.

See Section 2.5 for details of virus growth, cultivation, and storage; coat-protein purification and storage, RNA purification and storage, gel electrophoresis measurements and analysis, dynamic light scattering measurement and analysis, and transmission electron microscopy.

3.5.1 WILD-TYPE AND COAT-PROTEIN BULK ASSEMBLY EXPERIMENTS

Wild-type MS2 at a concentration of 50 nM is mixed with varying concentrations of MS2 coat-protein dimers ranging from 1.25 μ M to 20 μ M and left at room temperature (21 °C) for 10 min. The assembled virus-like particles are then characterized using gel electrophoresis, DLS, and TEM, confocal microscopy, plaque assay, and zeta potential.

3.5.2 ZETA POTENTIAL MEASUREMENTS

Zeta potentials are measured using a Malvern Zetasizer Nano. We prepare three 1 mL samples consisting of (1) 15 μ M coat-protein dimers, (2) 50 nM wild-type bacteriophage MS2, and (3) 50 nM wild-type bacteriophage MS2 mixed with 15 μ M coat-protein dimers.

3.5.3 FLUORESCENT LABELING AND CONFOCAL MICROSCOPY

Wild-type MS2 was fluorescently labeled with Alexa Fluor 647 succinimidyl ester (Thermo Fisher Scientific, #A37573). We incubated the virus for 1.5 hr in a 0.1 M sodium bicarbonate solution and dye at a concentration of 1 mg/mL. The dyed MS2 was then washed with TNE buffer using 100-kDa-MWCO centrifugal filters (Millipore Sigma, UFC100824) and stored at 4 °C.

To image binding of the virus to the host, we placed *E. coli* C3000 cells on a base-washed number 1 coverslip (Globe Scientific, 1419-10), and we added 2 μ L of labeled MS2 particles to 10 μ L of cells. We placed a 2% agarose pad made with PBS buffer on top of the sample to immobilize the cells. Samples were prepared with 0, 5, 10, and 15 μ M of added MS2 coat-protein dimers and imaged on a Leica TCS SPF confocal microscope with a 63 \times water immersion objective (NA 1.2) and a 633 nm excitation wavelength.

3.5.4 PLAQUE ASSAY

For plaque assays, we mix 50 nM wild-type MS2 with coat-protein dimers at 0, 5, 10, and 15 μ M. We then dilute each mixture by a factor of 10^6 . We mix 200 μ L *E. coli* into each of the diluted samples, add 10 mL of top agarose and pour into a 90 mm by 17 mm cell culture plate (VWR, 10062-878) containing 15 μ L of bottom agar. We let the plates cool for 5 min and place them upside down into an incubator at 37 °C overnight. During this time, the *E. coli* grows on top of the bottom agar. Simultaneously, the samples containing wild-type MS2 particles infect the *E. coli* and create a plaque (or hole)

representing a single virus infection.

4

A Phase Diagram of MS₂ Virus Self-Assembly

In this chapter, I describe experiments exploring the effects of RNA concentration, ionic strength, and temperature on VLP assembly. Our goal is to create a phase diagram representing the morphologies of MS₂ VLPs in an attempt to understand the physics that governs the assembly pathways.

The temperature experiments done in this chapter were done in conjunction with Benjamin “Adam”

4.1 INTRODUCTION

Since the discovery of the tobacco mosaic virus in the late 1800s¹⁹, an open question has remained: how do simple RNA viruses assemble? Crick and Watson noted that many small viruses come in the forms of rods and spheres⁶, an observation that raised the question of what physical constraints govern virus structure. They hypothesized, based on the apparently low amount of genetic information in a virus, that the viral capsid consists of many copies of identical, small subunits. Casper and Klug explained how spherical viruses could package larger genomes while still maintaining icosohedral symmetry, which allows the subunits to minimize strain in the capsid^{5,27}. They introduced the concept of quasi-equivalence, which describes a pattern of organizing subunits into pentamers and hexamers, the number of which is related to the triangulation number T . Experimental methods such as X-ray diffraction and electron microscopy have verified many of the structural predictions of Casper and Klug^{52,49,17,38}. However, they did not resolve the question of how these structures form.

The observation that certain viruses, such as bacteriophage MS2⁴⁷, can form virus-like particles (VLPs) from their constituent proteins and RNA outside of the host cell shows that they self-assemble. This observation has led to further studies that use *in vitro* measurements to understand the assembly pathway of a virus. For viruses such as MS2 and BMV, it has been found that the pathway to assembly is one that follows nucleation and growth, as discussed in this thesis^{15,16}. Other studies have explored the importance of protein-protein and protein-RNA interactions on the assembly pathway^{41,40,18,3}. Here, we explore the effect of several variables that could affect the assembly pathway in MS2 that include RNA concentration, ionic strength, and temperature.

Our goal is to build a phase diagram that captures the effects of these variables (in addition to coat-protein concentration from Chapter 2) so that we can shed further light on the interactions and path-

ways that control the assembly of MS₂. We do this by mixing MS₂ coat protein with MS₂ RNA under different conditions and characterizing the resulting assemblies with transmission electron microscopy (TEM), gel electrophoresis, and dynamic light scattering (DLS). As seen in chapter 2, TEM provides structural information about the assembly products, gel electrophoresis provides qualitative information about the size and composition, and DLS provides us with quantitative information about the size. The combination of these methods allows us to characterize and confirm the assembly products and, thereby, to create phase diagrams of MS₂ VLP assembly.

4.2 RESULTS AND DISCUSSION

4.2.1 RNA CONCENTRATION

As discussed in Chapter 1, simple RNA viruses such as MS₂ consist of two components: RNA and coat protein (wild-type MS₂ also contains a maturation protein, which is not found in MS₂ VLPs). In chapter 2, we explored the effects that coat-protein concentration has on the assembly of MS₂ VLPs. In this section, we consider the role of RNA concentration. We combine 5 μ M MS₂ coat-protein dimers with varying concentrations of MS₂ RNA ranging from 10 nM to 200 nM. For reference, a well-formed MS₂ VLP and wild-type MS₂ virus have icosahedral capsids with a diameter of about 30 nm. The stoichiometric ratio of coat-protein dimers to RNA is 90:1, corresponding to about 50 nM RNA to 5 μ M protein dimer concentration. After mixing, we wait 10 min to allow assembly to occur, after which we add RNase to digest any excess RNA that is not encapsidated. We then characterize the products using gel electrophoresis and TEM (see Section 2.5).

We take gel electrophoresis measurements of our samples and observe that at each RNA concentration there is a band that runs at the same position as that of the wild-type virus (Fig. 4.1). This result suggests that well-formed capsids assemble irrespective of the RNA concentration. To confirm this hypothesis we take TEM images of negatively stained samples and observe well-formed VLPs at all

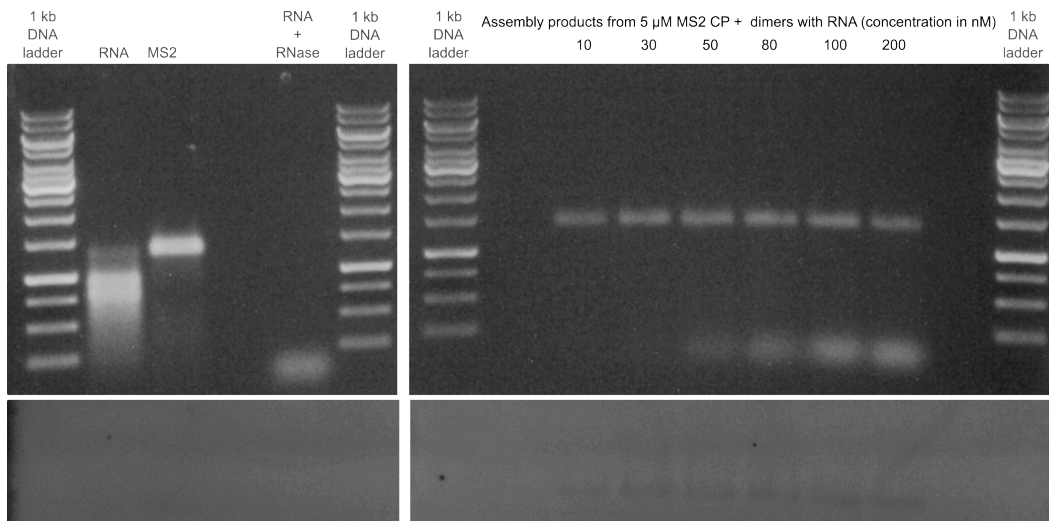


Figure 4.1: Images of agarose gels used to characterize the assembly products of 5 μ M MS2 coat-protein dimers with varying MS2 RNA concentrations ranging from 10 nM to 200 nM. We first stain with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat proteins (bottom image). Lane 1, 6, 7, and 16 all contain 1 kb DNA ladders. Lanes 2, 3, and 5 show controls: MS2 RNA, wild-type MS2 virus, and MS2 RNA treated with RNase. Lanes 9–14 show the assembly products of these experiments. We observe that with increasing RNA concentration the particles that have assembled have the same size as the wild-type MS2 virus.

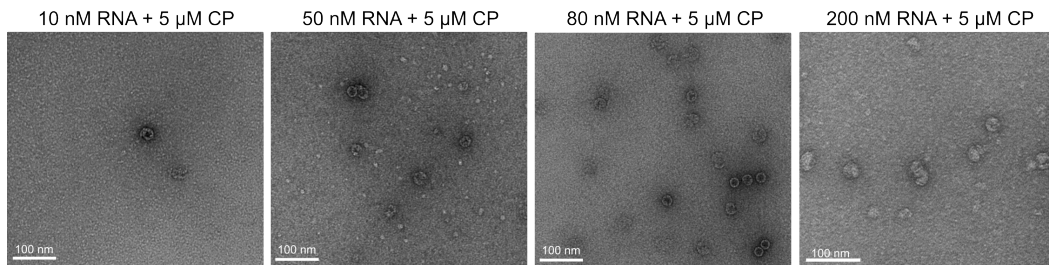


Figure 4.2: TEM images of negatively stained samples of assembly products of 5 μ M MS2 coat-protein dimers at varying MS2 RNA concentrations. At all concentrations of MS2 RNA we see well formed MS2 VLPs that resemble the wild-type MS2 virus.

RNA concentrations (Fig. 4.2). These results are interesting because well-formed VLPs assemble at both sub-stoichiometric and super-stoichiometric concentrations of RNA. In the sub-stoichiometric case, coat proteins are in excess relative to the RNA, but we do not see evidence that multiple nucleation sites are forming on each RNA. At super-stoichiometric ratios, RNA is in excess relative to the coat protein, but we do not see partial capsids form. These results show that at all concentrations of RNA, the nucleation rate is slow in comparison to the growth rate, meaning that the RNA concentration does not change the pathway. Only the coat-protein concentration controls the nucleation rate.

4.2.2 IONIC STRENGTH

Whereas varying concentrations of protein and RNA affects primarily the kinetics of nucleation and growth, varying the ionic strength allows us to explore changes in protein-protein and protein-RNA interactions. To understand the effects of these changes, we assemble VLPs by combining 50 nM MS2 RNA with 5 μ M coat-protein dimers in assembly buffer at varying concentrations of NaCl ranging from 100 mM NaCl to 1000 mM NaCl (see Section 2.5).

We use gel electrophoresis to characterize the size and composition of the assembly products (see Section 2.5). We observe two bands at each concentration, one of which runs at the same position as the digested RNA, and the other of which runs at about the same position as the wild-type virus. With increasing NaCl, the band that runs near the position of the wild-type capsid shifts toward larger sizes and broadens above 300 mM NaCl (Fig. 4.3). This result indicates that at salt concentrations above 300 mM, the assembly products contain VLPs that are larger than the wild-type MS2 virus. TEM images (Figs. 4.4 and 4.5) show well-formed capsids at 100 mM and 300 mM NaCl, aggregates of capsids at 600 mM NaCl, and larger (micrometer-scale) aggregates at 1000 mM NaCl. Gels indicate that these large aggregates contain both RNA and protein. We therefore call them condensates.

To quantify the sizes, we perform DLS measurements that use numerical inversion methods to pro-

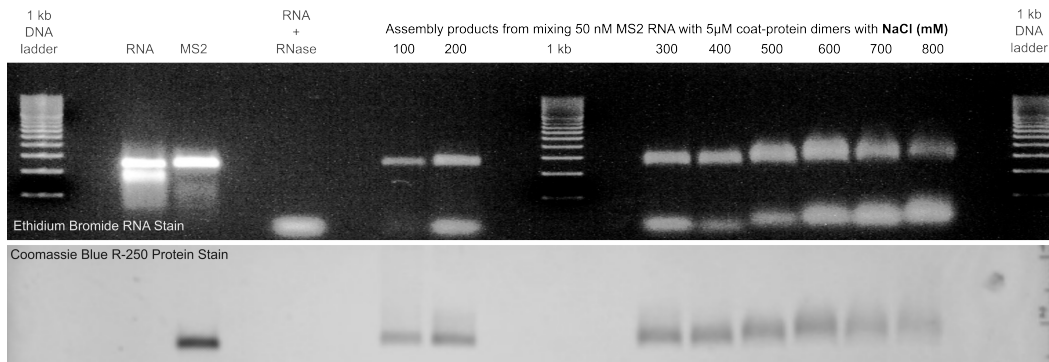


Figure 4.3: Images of agarose gels used to characterize assemblies formed when we mix 50 nM MS2 RNA with 5 μM MS2 coat-protein dimers at NaCl concentrations ranging from 100 mM to 1000 mM. Lanes 3, 4, and 6 show the following controls: MS2 RNA, wild-type MS2 virus, and digested RNA. Lanes 1, 11, and 20 contain 1 kb DNA ladders. Lanes 8, 9, and 13-18 contain the results from experiments done at varying NaCl concentration. We observe a broadening of the topmost band with increasing NaCl, indicating that larger structures are forming that contain both RNA and protein.

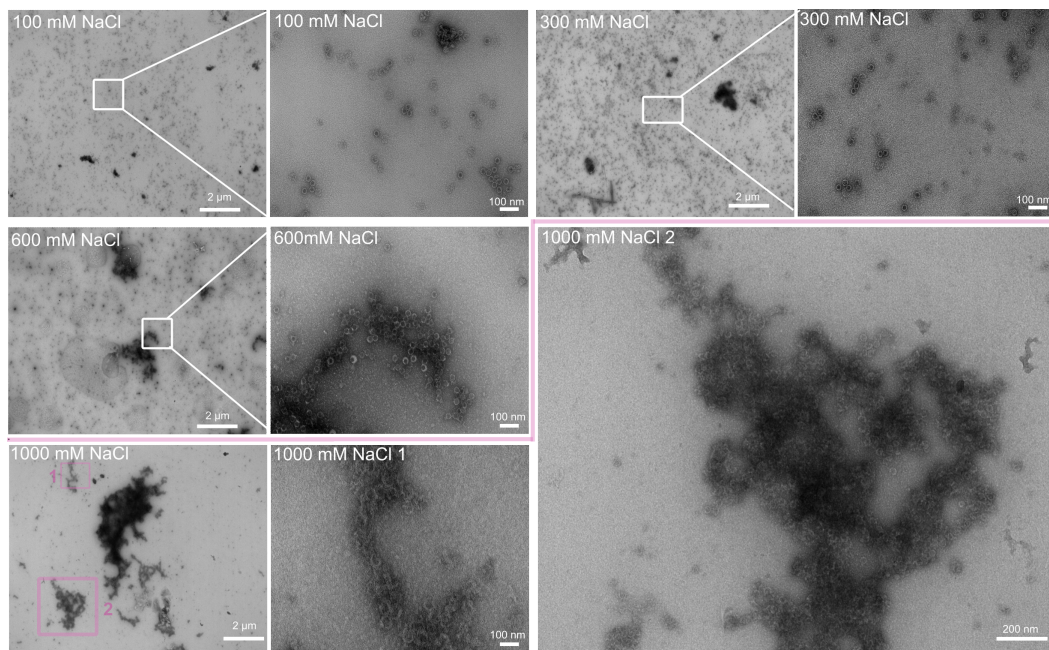


Figure 4.4: TEM images of negatively stained samples of assembly products of 50 nM concentration of MS2 RNA mixed with 5 μM MS2 coat-protein dimers at 100 mM, 300 mM, 600 mM, and 1000 mM NaCl concentrations. At lower concentrations (100 mM and 300 mM NaCl), we see well-formed MS2 VLPs. At higher concentrations, we see aggregates of capsids. At the highest NaCl concentration, 1000 mM, we see large, disordered aggregates that resemble the condensates seen in chapter 2.

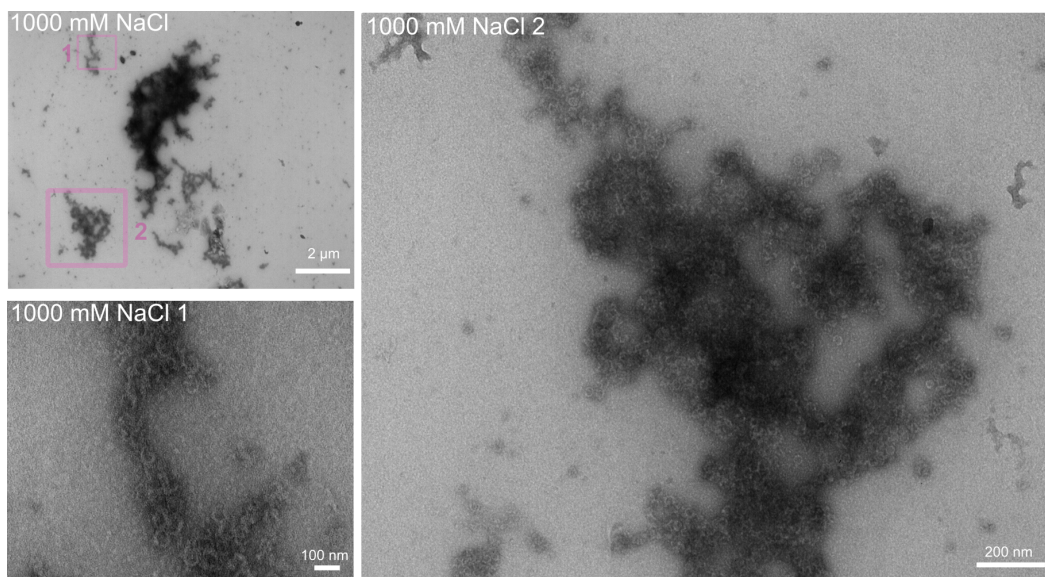


Figure 4.5: Higher-magnification TEM images of the structures formed at 1000 mM NaCl.

vide size distributions of the samples on a volume basis (see Section 2.5.5). At concentrations below 600 mM NaCl, we see evidence of particles that are the same size as the wild-type virus in addition to what may be aggregates or dust (Fig. 4.6). At 600 mM we observe that the secondary peak increases in intensity, indicating that larger particles are being formed. At 1000 mM NaCl concentration, we observe that both peaks no longer exist within the size range of well-formed MS2 VLPs.

We associate the peaks at 600 mM and 1000 mM NaCl concentration with the clusters of VLPs and condensates seen in the TEM data (Fig. 4.5). These structures would account for the increase in the size and broadening of the band seen in the gel electrophoresis measurements. These results suggest that increasing ionic strength affects the overall size of the assemblies by screening charges, leading to aggregation. However, we must still determine what components are aggregating and how that aggregation affects assembly.

A hypothesis for why clusters of VLPs form at intermediate concentrations of NaCl (above 100 mM and below 1000 mM) is that the VLPs aggregate after the capsids fully assemble. Indeed, when we

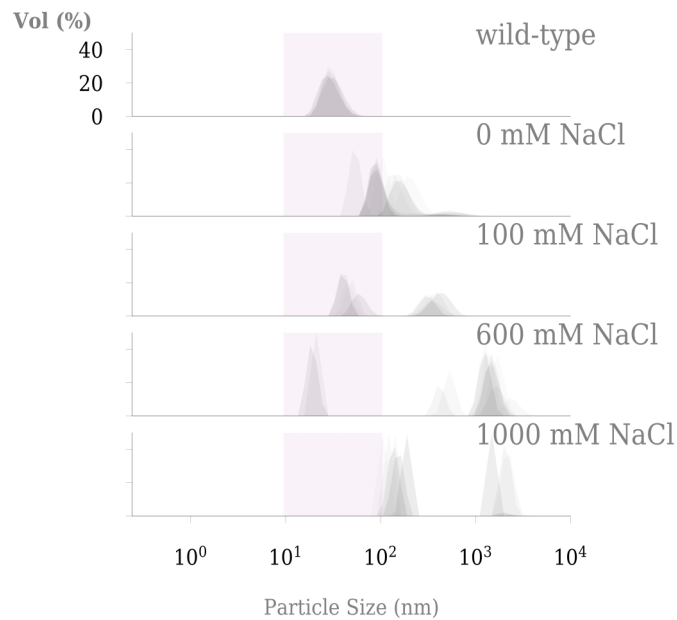


Figure 4.6: Plots of size distribution of wild-type MS2 virus and VLPs assembled *in vitro* as inferred from dynamic light scattering. We show the size distribution on a volume basis. The light gray peaks indicate eight individual experiments. We observe an increase in average particle size with increasing NaCl concentration.

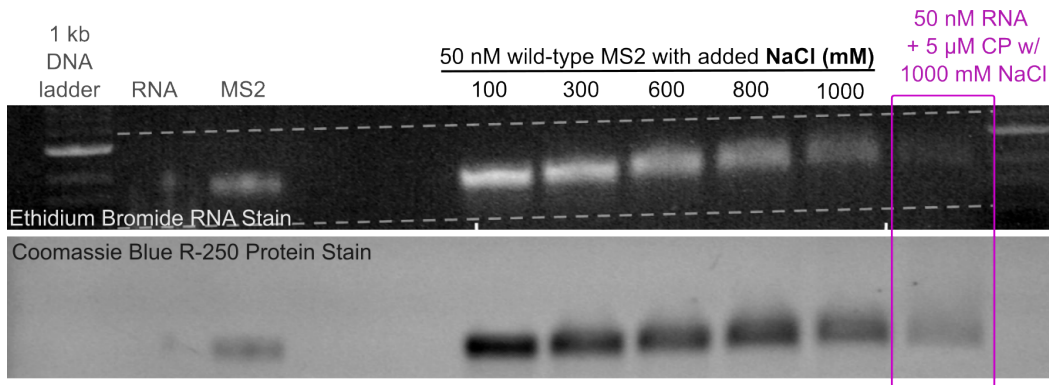


Figure 4.7: Image of gel containing 50 nM wild-type MS2 virus in buffers containing 100 mM to 1000 mM NaCl. Lanes 2 and 3 contain controls for MS2 RNA and wild-type MS2 virus at 100 mM NaCl. In addition, lane 19 contains a control for an MS2 VLP assembled in buffer containing 1000 mM NaCl. Lanes 1 and 19 contain DNA ladders. The remaining lanes contain samples of wild-type MS2 virus at varying NaCl concentrations. We observe an overall increase in the size of these particles, suggesting clustering or aggregation.

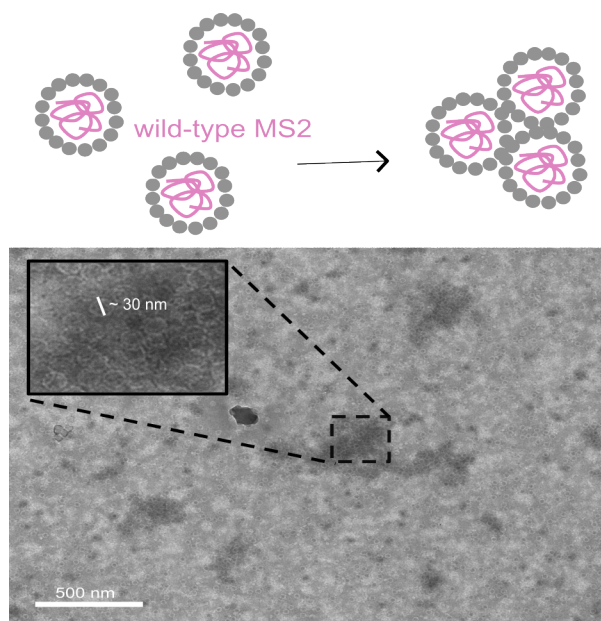


Figure 4.8: TEM images of negatively stained samples of 50 nM wild-type MS2 virus in 1000 mM NaCl. We observe clustering of the wild-type capsids.

subject wild-type MS2 to increasing salt concentration, we see an increase in particle size in gel electrophoresis experiments (Fig. 4.7). TEM images of wild-type MS2 at 1000 mM NaCl show that the wild-type particles form large aggregates (Fig. 4.8).

Another hypothesis is that the RNAs themselves aggregate before the capsids fully assemble. To test the possibility of RNA aggregation, we perform gel electrophoresis experiments on free MS2 RNA at varying NaCl concentrations (see Section 4.4). We observe that the band corresponding to RNA shifts to larger sizes with increasing salt concentration. This result is evidence of RNA aggregation. If RNA aggregation were to happen before assembly or concurrently with assembly, we would expect that aggregation would prevent capsids from growing to full size, leading to the formation of condensates containing many partially formed capsids. The TEM images (Fig. 4.5) do show at least some partial capsids, lending some support for this hypothesis. It is important to note, though, that changing the ionic strength should affect both the nucleation and growth rates, and might even change the assembly

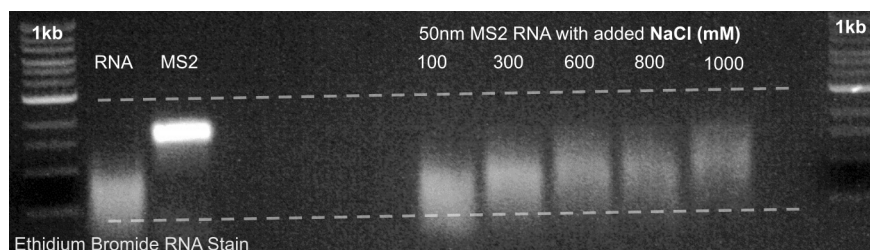


Figure 4.9: Image of gel showing 1 kb DNA ladder (lanes 1 and 13), controls consisting of MS2 RNA (lane 2), wild-type MS2 virus (lane 3), and MS2 RNA at NaCl concentrations from 100 mM to 1000 mM (lanes 7–11). The results show that RNA aggregation may occur with increasing salt concentration.

pathway from nucleation and growth to *en masse* assembly. What is clear is that charge screening plays a strong role in the formation of the clusters and condensates, though it is not yet clear when the aggregation occurs relative to when assembly occurs. A diagram of the potential assembly pathways is given in Fig. 4.10.

4.2.3 TEMPERATURE

Next we explore the effect of temperature on the assembly. We know that *E. coli* cells grow well between 21 °C and 49 °C¹¹. In these experiments, we combine 5 μ M MS2 coat-protein dimers to 50 nM MS2 RNA and wait 10 min for assembly to occur at temperatures ranging from 4 °C to 60 °C, after which we add RNase to digest any excess RNA. We characterize the resulting assemblies using TEM, gel electrophoresis, and dynamic light scattering.

Gel electrophoresis measurements (Fig. 4.11) show a band that runs at the same position as that of the wild-type virus as we increase temperature from 10 °C to 32 °C and that increases in intensity as temperature increases. At 32 °C, the band has the highest intensity. Above this temperature, the primary band corresponding to the size of the wild-type MS2 virus decreases in intensity, and a sharp secondary band appears at 37 °C. This secondary band continues to increase in intensity as the primary band disappears. At 50 °C, both bands disappear and the samples do not appear to travel past the well.

TEM images (Figs. 4.12 and 4.13) show that at low temperatures (10 °C), the assemblies consist of

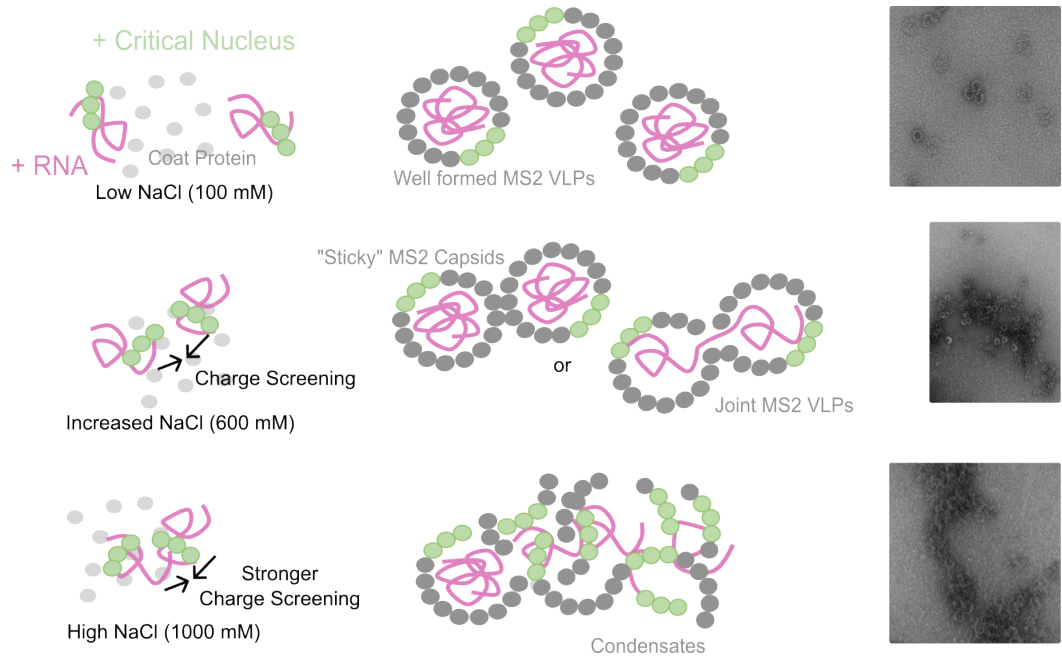


Figure 4.10: Diagram showing how ionic strength might affect the pathways and morphologies of assembly. Top diagrams show a pathway for assembly of well-formed VLPs at low NaCl concentration (100 mM). Middle diagrams show a pathway for assembly of clusters of capsids at 600 mM NaCl. Bottom diagrams show a pathway for assembly of condensates at 1000 mM NaCl. The images to the right of each diagram are transmission electron micrographs of negatively stained MS2 VLPs.

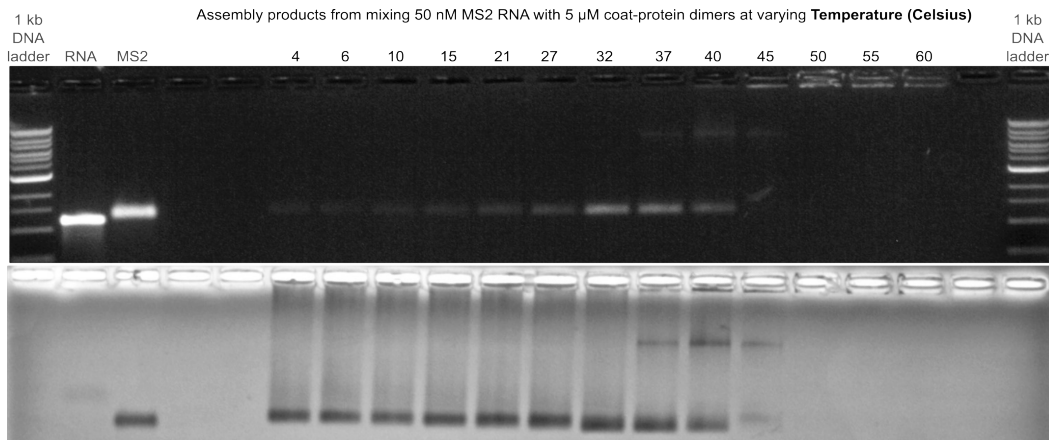


Figure 4.11: Images of agarose gels used to characterize assembly products at varying temperature. Lanes 1 and 20 show a 1 kb DNA ladder. Lanes 2 and 3 show two controls: MS2 RNA (lane 2) and wild-type MS2 virus (lane 3). The other lanes show the results of gel electrophoresis on samples assembled at temperatures ranging from 4 °C to 60 °C. One of the most striking features of this gel is the increase in intensity of the lower band as temperature increases to 32 °C and decrease in intensity above this temperature. In addition, we see a second band appear starting at 37 °C and disappear above 45 °C. Above 45 °C the samples do not travel past the wells of the gel.

incomplete capsids and clusters of proteins that are not quite spherical but are approximately the size of the wild-type virus (Fig. 4.11). As we increase temperature (between 21 °C and 40 °C), we observe well-formed MS2 VLPs. At 50 °C, we see evidence of aggregation. Lower-magnification views of the same samples show that samples between 21 °C and 40 °C are not aggregated (Fig. 4.13).

DLS measurements show a single peak at approximately the size of wild-type capsids for temperatures between 21 °C and 40 °C (Fig. 4.14). At low temperatures (10 °C), we see a single peak corresponding to large particles. Interestingly, gel and TEM data do not show evidence of such large particles at this temperature. This discrepancy might be explained by weak aggregation of capsids or partially formed capsids. Such weakly bound aggregates might fall apart in the gel. Finally at 50 °C, there exists one peak corresponding to a size much larger than the wild-type virus. This result is consistent with what is seen in the gel data (Fig. 4.11) and in the TEM data (Fig. 4.12).

Our experiments show that changes in temperature affect the overall structure of the MS2 VLPs. We attribute the formation of these structures to the many ways that temperature affects the coat

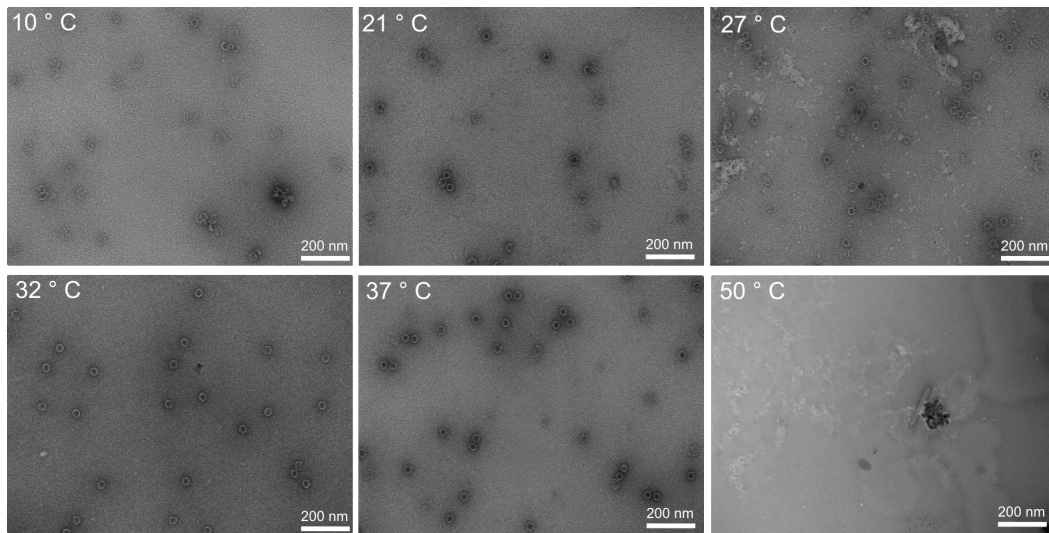


Figure 4.12: TEM images of negatively stained samples of 50 nM MS2 RNA mixed with 5 μ M MS2 coat-protein dimers at varying temperatures ($^{\circ}$ C).

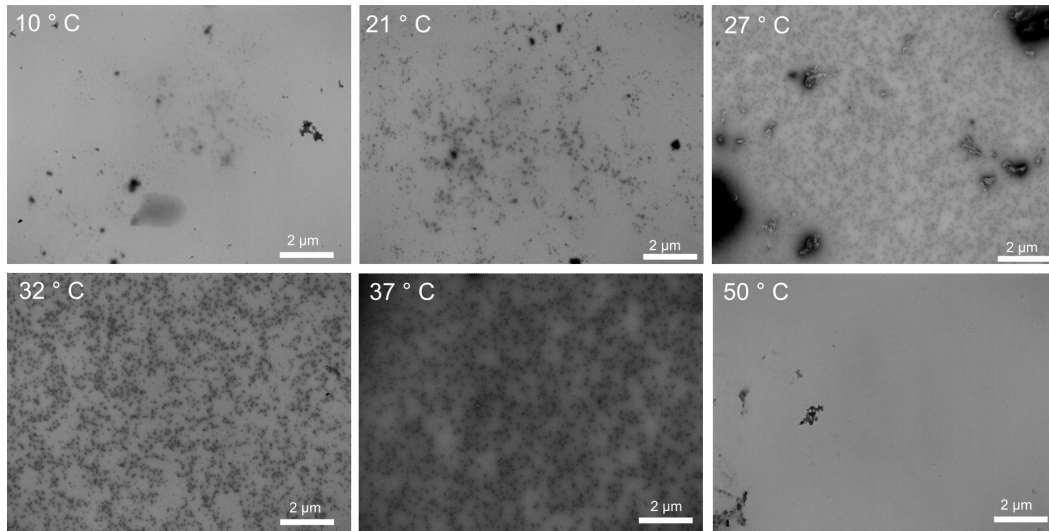


Figure 4.13: Lower-magnification TEM images of the same samples shown in Fig. 4.12.

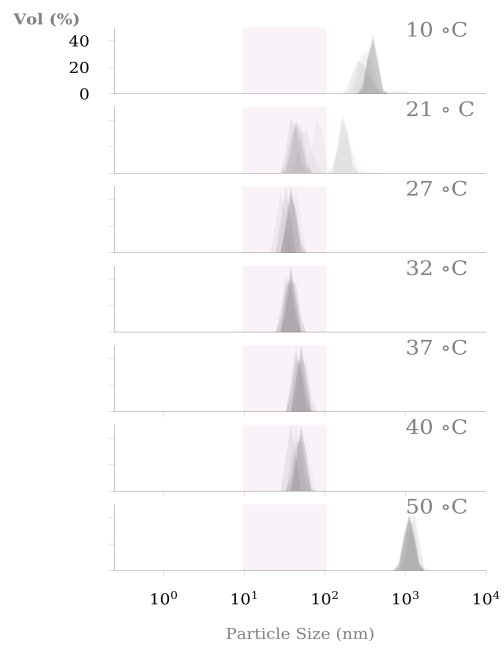


Figure 4.14: Plots of size distributions of MS2 VLPs assembled *in vitro* at various temperatures from 50 nM free MS2 RNA and 5 μ M concentration of MS2 coat-protein dimers. The distributions are inferred from dynamic light scattering measurements and show the size distribution on a volume basis. The light gray peaks show the results of nine individual experiments.

protein, the RNA, and their interactions with each other. At temperatures above 32 °C, the secondary structure of the RNA becomes susceptible to additional changes in temperature⁴⁸. Above 40 °C, complementary bases of the RNA are ruptured, causing the RNA to unfold⁴⁸. Above 50 °C, protein-protein interactions are also disturbed⁴⁸. These changes in interactions and structure might explain why we see multiple changes to the assembly products, particularly in the gel data (Fig. 4.11). However, the most striking feature occurs at lower temperatures, where very few well-formed MS2 VLPs assemble.

At 10 °C, we observe structures that are approximately 30 nm in diameter but are not spherical (or do not have icosahedral symmetry). These structures could be the result of partially formed MS2 VLPs that aggregate as a result of protein-protein interactions. Another potential hypothesis is that these structures are evidence of a transition to an *en masse* assembly pathway⁴⁹. In this context, the coat-protein would decorate the RNA before reorganizing into well-formed capsids. It is possible that at low temperatures, the timescale of our assembly reaction (10 min) is too small for the clusters to organize into well-formed VLPs. Because variations in temperature change many interactions within the system (both specific and non-specific), it is difficult to interpret what causes the formation of these clusters. However, further characterization of the protein-protein and protein-RNA interactions as a function of temperature might shed light on the formation of these structures.

4.3 CONCLUSIONS

Our experiments have shown that there are many parameters that affect the assembly of MS2 VLPs. We have shown that MS2 coat protein and RNA can assemble into fully enclosed capsids that stick, incomplete or *en masse* assembled capsids, and aggregates (Fig. 4.15). These results demonstrate how varying the components (the RNA and coat protein) and varying the interactions (ionic strength and temperature) can cause MS2 assembly to “go right” and “go wrong”. Our results illustrate the impor-

tance of charge screening, which can lead to aggregation of capsids and possibly RNA with increasing ionic strength. We have also shown that changing temperature can cause multiple changes in the structure of the assemblies, likely due to changes in RNA or protein structure or both.

4.3.1 ACKNOWLEDGMENTS

We thank Amy Barker and Peter Stockley at the University of Leeds for initial stocks of MS2 and *E. coli* cells. We thank Tim Chiang, Amelia Paine, Aaron Goldfain, and Danai Montalvan for helpful scientific discussions. This research was partially supported by a National Science Foundation (NSF) Graduate Research Fellowship under grant number DGE-1745303 and by NSF through the Harvard University Materials Research Science and Engineering Center under NSF grant number DMR-2011754, by the National Institute of General Medical Sciences of the National Institutes of Health under grant K99GM127751, by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University under NSF grant number 1764269, and by the Harvard Quantitative Biology Initiative. This work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF grant number ECCS-2025158. The work was also performed in part at the Harvard University Bauer Core Facility. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

4.4 MATERIALS AND METHODS

Buffers, materials, and characterization methods are described in Section 2.5.

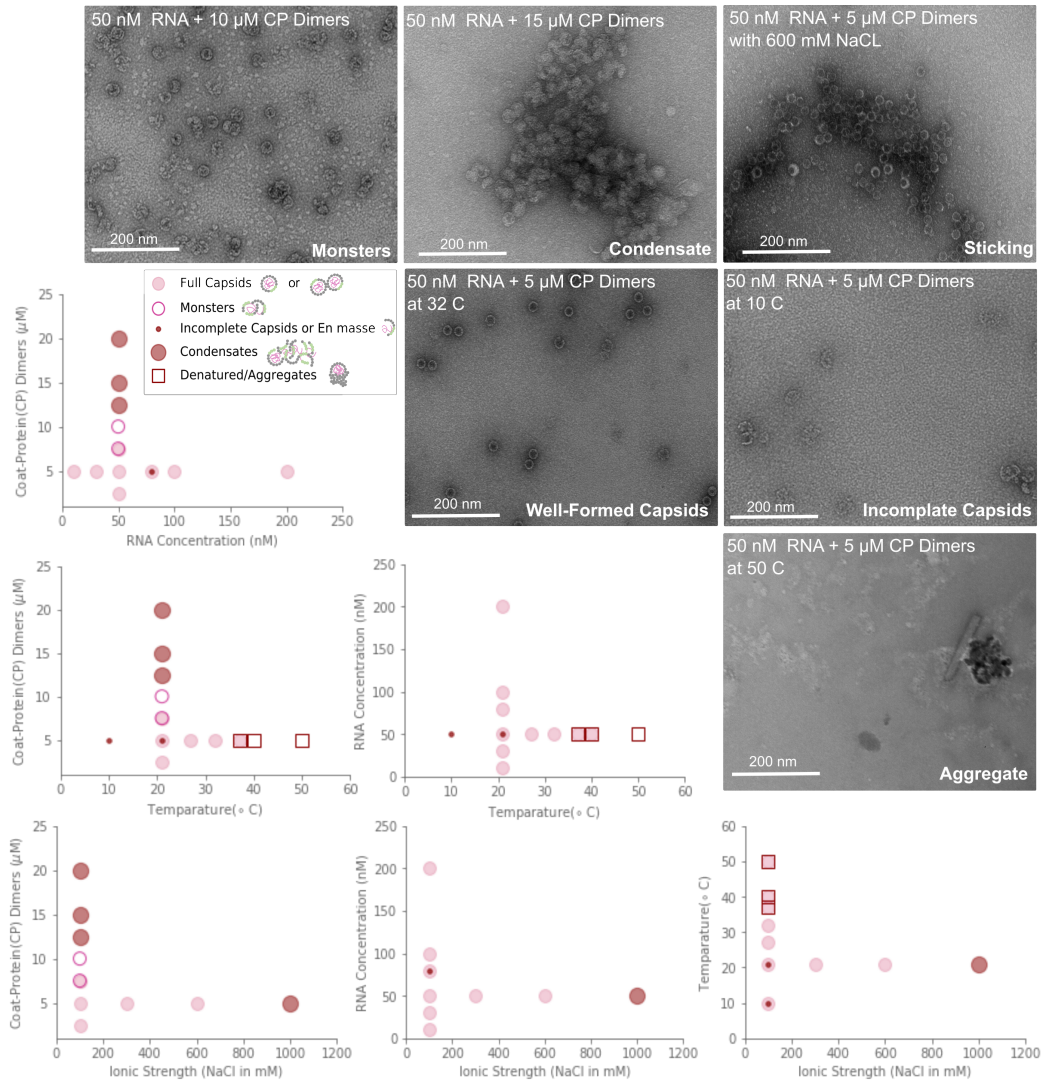


Figure 4.15: Phase diagrams of MS2 VLP assembly (bottom left) comparing coat-protein dimer concentration, MS2 RNA concentration, NaCl concentration, and temperature. We observe 6 potential morphologies or assembly products that can be explored (top right): well-formed capsids, monsters, condensates, capsids that stick, incomplete or *en masse* capsids, and aggregates.

5

Conclusions and Future Work: How to Build a Virus and Potential Applications

Prior to this study, we had some understanding of how virus self-assembly (in particular MS2 VLP assembly) can “go right.” The work in this thesis has demonstrated not only that there are many parameters that contribute to the assembly of well-formed VLPs, but that virus self-assembly can in fact “go wrong.” In chapter 2, we showed that increasing coat-protein concentration changes the nucle-

ation rate, resulting in multiple nucleation sites on the RNA and subsequent formation of monster capsids and condensates. In chapter 3, we showed that MS2 coat proteins can assemble around wild-type MS2 and MS2 VLPs at sufficiently high concentrations. The formation of these additional layers appears to be driven by electrostatic interactions between the coat proteins and the pre-assembled capsids. In chapter 4, we showed how other parameters such as RNA concentration, ionic strength, and temperature affect the assembly products and found that there are many ways assembly can go wrong, leading to the formation of fully enclosed capsids that stick, incomplete capsids, and aggregates. The formation of many, though not all, of these assembly products can be related to a nucleation-and-growth assembly pathway.

From these experiments, we now have an understanding of the various assembly products of MS2 RNA and coat protein. In addition, we have some understanding of the physics of how these parameters affect the assembly pathways. Knowing what drives these structures to form allows us to infer what parameters are most important and eventually control that pathway. The work shown in chapter 4 in particular leaves a lot of room for further exploration into a phase diagram for MS2 VLP assembly in addition to a deeper understanding of the physical principles that may govern virus self-assembly.

5.1 A PHASE DIAGRAM OF VIRUS SELF-ASSEMBLY AND BEYOND

5.1.1 FURTHER WORK WITH MS2

Another important parameter that we were unable to fully explore was viscosity and its impact on MS2 VLP assembly. The viscosity inside of an *E. coli* cell (the native host of bacteriophage MS2) is 0.95 Pa-s³⁵. We perform our experiments in assembly buffer (see Section 2.5) which has a viscosity of 0.001 Pa-s, similar to water. The inside of an *E. coli* cell is much more viscous than water or buffer and would therefore affect the dynamics of the coat-protein dimers and the RNA. We attempted to explore the effects of increasing viscosity by assembling 50 nM MS2 RNA with 5 μ M concentration of

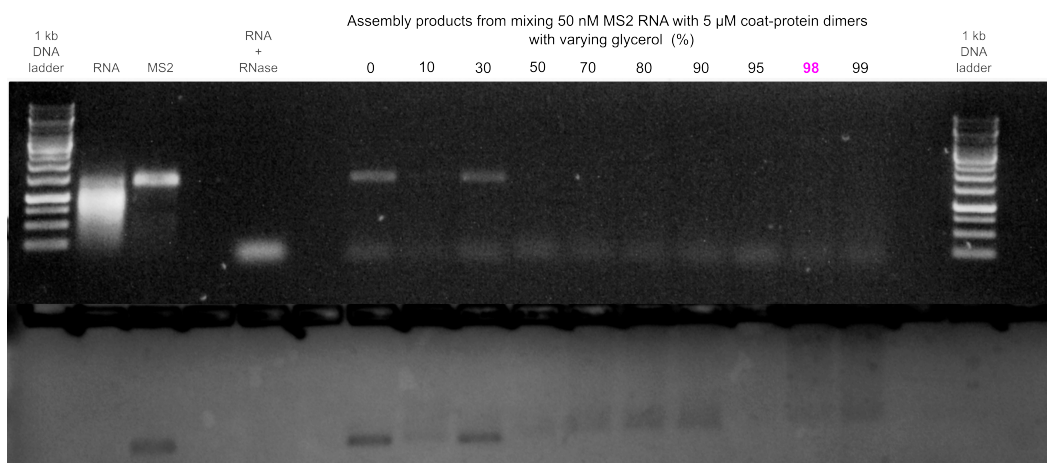


Figure 5.1: Images of agarose gels used to characterize products of assembly at increasing concentrations of glycerol. We stain first with ethidium bromide to detect the RNA (top image) and then stain with Coomassie Blue R-250 to detect MS2 coat protein (bottom image). Lanes 1 and 19 contain 1 kb DNA ladders. Lanes 2, 3, and 5 contain the following controls: MS2 RNA, wild-type MS2 virus, and digested RNA. The remaining lanes contain 5 μ M coat-protein mixed with 50 nM RNA at varying glycerol concentrations ranging from 0% to 99%.

coat-protein dimers (as we have done in Chapter 2 and Chapter 4) at varying concentrations of glycerol (Fig. 5.1). The concentrations of glycerol ranged from 0 to 99% with concentrations between 95% and 99% having a similar viscosity to the cytoplasm inside an *E. coli* cell⁴⁴. From gel electrophoresis alone we were able to capture a change in the number and brightness of bands in the gel with increasing glycerol concentration. These changes may be due to changes in the rate at which proteins arrive at the RNA or in the rate at which the RNA folds. The results might eventually give insight into the prefactor for the nucleation rate and might help us understand the rate of RNA folding as well.

There are many other experiments to be done to further explore the phase diagram in chapter 4 (Fig. 4.15). For example, measurements of the protein-protein and RNA-protein binding affinities as a function of temperature might give insight into the assemblies observed at different temperatures. Also, exploring other areas of our phase diagram by combining parameters would assist in determining where phase transitions might occur. For example, we might vary both ionic strength and temperature or other combinations of parameters to further explore this space.

Although MS₂ has been an model system to study VLP assembly, there are other viruses that have not been studied to the same extent. Exploring other viruses (including human viruses such as HSV) in a similar manner would allow us to compare phase diagrams between viruses and determine which parameters are most important across many viruses. Such studies could lead to the discovery of general principles governing virus assembly. In addition, these experiments could be helpful in choosing parameters for further simulations of these types of systems or assembling viral capsids for drug delivery more effectively.

5.2 MS₂ COAT-PROTEIN ASSEMBLY AROUND SCAFFOLDS

In chapter 2 we explored the effect that electrostatics have on MS₂ coat-protein assembly around wild-type viruses and showed that electrostatics play an important role in the nucleation and growth of a conformal shell around a scaffold. Future work might focus on experiments that probe the effects of ionic strength, which would help us understand the electrostatic interactions in more detail. In addition, our discovery that wild-type MS₂ has a negative charge that makes it potentially susceptible to the formation of additional coat protein layers might make MS₂ an ideal system to study viruses with multiple coat-protein shells. Our experiments also lead us to believe that MS₂ coat-proteins may be able to nucleate and grow capsids on other negatively charged particles such as gold in the same way that has been demonstrated with CCMV⁸. With the knowledge from our MS₂ VLP phase diagram, we could also potentially produce MS₂ VLPs *in vitro* on fast timescales and at higher yields for applications in medicine and environmental science.

Overall, the goal of this thesis was to learn about the conditions that promote virus self-assembly and the ones that do not. And in turn we've learned about how viruses might build themselves using nucleation and growth. We've learned that the coat protein (and the concentration) matters. We've learned that ionic strength matters. We've learned that temperature matters. And we've also learned

that the RNA itself might not matter as much, despite what previous studies have shown. With this knowledge and studies that build upon it, not only should we be able to assemble VLPs more efficiently, we might also be able to assemble more complex viruses from their basic components – the way nature does.

References

- [1] Aniagyei, S. E., DuFort, C., Kao, C. C., & Dragnea, B. (2008). Self-assembly approaches to nanomaterial encapsulation in viral protein cages. *Journal of Materials Chemistry*, 18(32), 3763–3774.
- [2] Bancroft, J. B. & Hiebert, E. (1967). Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. *Virology*, 32(2), 354–356.
- [3] Beckett, D., Wu, H.-N., & Uhlenbeck, O. C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *Journal of Molecular Biology*, 204(4), 939–947.
- [4] Božič, A. L. & Podgornik, R. (2017). pH dependence of charge multipole moments in proteins. *Biophysical Journal*, 113(7), 1454–1465.
- [5] Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 27 (pp. 1–24).
- [6] Crick, F. H. & Watson, J. D. (1956). Structure of small viruses. *Nature*, 177, 473–475.
- [7] Cuetos, A. & Martínez-Haya, B. (2008). Columnar phases of discotic spherocylinders. *The Journal of Chemical Physics*, 129(21), 214706.
- [8] Durán-Meza, A. L., Escamilla-Ruiz, M. I., Segovia-González, X. F., Villagrana-Escareño, M. V., Vega-Acosta, J. R., & Ruiz-García, J. (2020). Encapsulation of different plasmonic gold nanoparticles by the CCMV CP. *Molecules*, 25(11), 2628.
- [9] Duval, J. F. & Ohshima, H. (2006). Electrophoresis of diffuse soft particles. *Langmuir*, 22(8), 3533–3546.
- [10] Elrad, O. M. & Hagan, M. F. (2010). Encapsulation of a polymer by an icosahedral virus. *Physical Biology*, 7(4), 045003.
- [11] Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N., & Timmis, K. N. (2003). Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nature Biotechnology*, 21(11), 1266–1267.

- [12] Fraenkel-Conrat, H. & Williams, R. C. (1955). Reconstitution of active Tobacco Mosaic Virus from its inactive protein and nucleic acid components. *Proceedings of the National Academy of Sciences*, 41(10), 690–698.
- [13] Galaway, F. A. & Stockley, P. G. (2013). MS2 viruslike particles: A robust, semisynthetic targeted drug delivery platform. *Molecular Pharmaceutics*, 10(1), 59–68.
- [14] Garmann, R. F., Comas-Garcia, M., Gopal, A., Knobler, C. M., & Gelbart, W. M. (2014). The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *Journal of Molecular Biology*, 426(5), 1050–1060.
- [15] Garmann, R. F., Goldfain, A. M., & Manoharan, V. N. (2019). Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proceedings of the National Academy of Sciences*, 116(45), 22485–22490.
- [16] Garmann, R. F., Goldfain, A. M., Tanimoto, C. R., Beren, C. E., Vasquez, F. F., Villarreal, D. A., Knobler, C. M., Gelbart, W. M., & Manoharan, V. N. (2022). Single-particle studies of the effects of RNA–protein interactions on the self-assembly of RNA virus particles. *Proceedings of the National Academy of Sciences*, 119(39), e2206292119.
- [17] Golmohammadi, R., Valegård, K., Fridborg, K., & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *Journal of Molecular Biology*, 234(3), 620–639.
- [18] Hagan, M. F. (2009). A theory for viral capsid assembly around electrostatic cores. *The Journal of Chemical Physics*, 130(11), 114902.
- [19] Harrison, B. & Wilson, T. (1999). Milestones in research on tobacco mosaic virus. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 354(1383), 521–529.
- [20] Hartman, E. C., Jakobson, C. M., Favor, A. H., Lobba, M. J., Álvarez-Benedicto, E., Francis, M. B., & Tullman-Ercek, D. (2018). Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nature Communications*, 9(1), 1385.
- [21] Hiebert, E., Bancroft, J., & Bracker, C. (1968). The assembly in vitro of some small spherical viruses, hybrid viruses, and other nucleoproteins. *Virology*, 34(3), 492–508.
- [22] Homa, F. L. & Brown, J. C. (1997). Capsid assembly and DNA packaging in herpes simplex virus. *Reviews in Medical Virology*, 7(2), 107–122.
- [23] Johansson, H. E., Liljas, L., & Uhlenbeck, O. C. (1997). RNA recognition by the MS2 phage coat protein. *Seminars in Virology*, 8(3), 176–185.
- [24] Joseph, J. A., Espinosa, J. R., Sanchez-Burgos, I., Garaizar, A., Frenkel, D., & Collepardo-Guevara, R. (2021). Thermodynamics and kinetics of phase separation of protein-RNA mixtures by a minimal model. *Biophysical Journal*, 120(7), 1219–1230.

- [25] Kampmann, T. A., Boltz, H.-H., & Kierfeld, J. (2015). Monte Carlo simulation of dense polymer melts using event chain algorithms. *The Journal of Chemical Physics*, 143(4), 044105.
- [26] Kern, N. & Frenkel, D. (2003). Fluid-fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *Journal of Chemical Physics*, 118, 9882–9889.
- [27] Klug, A. & Caspar, D. (1961). The structure of small viruses. In *Advances in Virus Research*, volume 7 (pp. 225–325). Elsevier.
- [28] Kovacs, E. W., Hooker, J. M., Romanini, D. W., Holder, P. G., Berry, K. E., & Francis, M. B. (2007). Dual-surface-modified bacteriophage MS2 as an ideal scaffold for a viral capsid-based drug delivery system. *Bioconjugate Chemistry*, 18(4), 1140–1147.
- [29] Krahn, P., O’Callaghan, R., & Paranchych, W. (1972). Stages in phage R17 infection: VI. Injection of A protein and RNA into the host cell. *Virology*, 47(3), 628–637.
- [30] Lalwani Prakash, D. & Gosavi, S. (2021). Understanding the folding mediated assembly of the bacteriophage MS2 coat protein dimers. *The Journal of Physical Chemistry B*, 125(31), 8722–8732.
- [31] Lavelle, L., Gingery, M., Phillips, M., Gelbart, W., Knobler, C., Cadena-Nava, R., Vega-Acosta, J., Pinedo-Torres, L., & Ruiz-Garcia, J. (2009). Phase diagram of self-assembled viral capsid protein polymorphs. *The Journal of Physical Chemistry B*, 113(12), 3813–3819.
- [32] Li, S., Roy, P., Travasset, A., & Zandi, R. (2018). Why large icosahedral viruses need scaffolding proteins. *Proceedings of the National Academy of Sciences*, 115(43), 10971–10976.
- [33] Malvern Instruments (2013). *Zetasizer Nano Series User Manual*. Malvern Instruments Ltd, Enigma Business park, Govewood Road, Malvern Worcester WR14 1XZ, UK, 11 edition.
- [34] Michen, B. & Graule, T. (2010). Isoelectric points of viruses. *Journal of Applied Microbiology*, 109(2), 388–397.
- [35] Mika, J. T., Thompson, A. J., Dent, M. R., Brooks, N. J., Michiels, J., Hofkens, J., & Kuimova, M. K. (2016). Measuring the viscosity of the *Escherichia coli* plasma membrane using molecular rotors. *Biophysical Journal*, 111(7), 1528–1540.
- [36] Morozov, A. Y., Bruinsma, R. F., & Rudnick, J. (2009). Assembly of viruses and the pseudo-law of mass action. *Journal of Chemical Physics*, 131, 155101.
- [37] Nguyen, T. H., Easter, N., Gutierrez, L., Huyett, L., Defnet, E., Mylon, S. E., Ferri, J. K., & Viet, N. A. (2011). The RNA core weakly influences the interactions of the bacteriophage MS2 at key environmental interfaces. *Soft Matter*, 7(21), 10449–10456.

- [38] Ni, C.-Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D. S., & Ely, K. R. (1995). Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure*, 3(3), 255–263.
- [39] Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *The EMBO Journal*, 12(2), 595–600.
- [40] Perlmutter, J. D., Perkett, M. R., & Hagan, M. F. (2014). Pathways for virus assembly around nucleic acids. *Journal of Molecular Biology*, 426(18), 3148–3165.
- [41] Perlmutter, J. D., Qiao, C., & Hagan, M. F. (2013). Viral genome structures are optimal for capsid assembly. *eLife*, 2, e00632.
- [42] Ricciuti, C. P. (1972). Host-virus interactions in Escherichia coli: Effect of stationary phase on viral release from MS2-infected bacteria. *Journal of Virology*, 10(1), 162–165.
- [43] Saad, A., Zhou, Z. H., Jakana, J., Chiu, W., & Rixon, F. J. (1999). Roles of triplex and scaffolding proteins in herpes simplex virus type 1 capsid formation suggested by structures of recombinant particles. *Journal of Virology*, 73(8), 6821–6830.
- [44] Shankar, P. & Kumar, M. (1994). Experimental determination of the kinematic viscosity of glycerol-water mixtures. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 444(1922), 573–581.
- [45] Sorger, P., Stockley, P., & Harrison, S. (1986). Structure and assembly of turnip crinkle virus: II. Mechanism of reassembly in vitro. *Journal of Molecular Biology*, 191(4), 639–658.
- [46] Strauss Jr., J. H. & Sinsheimer, R. L. (1963). Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *Journal of Molecular Biology*, 7(1), 43–54.
- [47] Sugiyama, T., Hebert, R., & Hartman, K. (1967). Ribonucleoprotein complexes formed between bacteriophage MS2 RNA and MS2 protein in vitro. *Journal of Molecular Biology*, 25(3), 455–463.
- [48] Thomas Jr., G., Prescott, B., McDonald-Ordzie, P. E., & Hartman, K. (1976). Studies of virus structure by Laser-Raman spectroscopy: II. MS2 phage, MS2 capsids and MS2 RNA in aqueous solutions. *Journal of Molecular Biology*, 102(1), 103–124.
- [49] Valegård, K., Liljas, L., Fridborg, K., & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, 345(6270), 36–41.
- [50] van der Holst, B., Kegel, W. K., Zandi, R., & van der Schoot, P. (2018). The different faces of mass action in virus assembly. *Journal of Biological Physics*, 44, 163–179.

- [51] Vega-Acosta, J., Cadena-Nava, R., Gelbart, W., Knobler, C., & Ruiz-García, J. (2014). Electrophoretic mobilities of a viral capsid, its capsid protein, and their relation to viral assembly. *The Journal of Physical Chemistry B*, 118(8), 1984–1989.
- [52] Verdaguer, N., Garriga, D., & Fita, I. (2013). X-Ray crystallography of viruses. In *Structure and Physics of Viruses: An Integrated Textbook*, volume 68 of *Subcellular Biochemistry* (pp. 117–144). Springer.
- [53] Zandi, R. & van der Schoot, P. (2009). Size regulation of ss-RNA viruses. *Biophysical Journal*, 96(1), 9–20.
- [54] Zandi, R., van der Schoot, P., Reguera, D., Kegel, W., & Reiss, H. (2006). Classical nucleation theory of virus capsids. *Biophysical Journal*, 90(6), 1939–1948.
- [55] Zlotnick, A. (1994). To build a virus capsid: An equilibrium model of the self assembly of polyhedral protein complexes. *Journal of Molecular Biology*, 241(1), 59–67.