

GENOME PACKAGING IN SINGLE-STRANDED RNA VIRUSES

A Dissertation

by

JENG-YIH CHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Junjie Zhang
Committee Members,	Thomas Meek
	Gregory Reinhart
	Guan Zhu
Head of Department,	Joshua Wand

May 2021

Major Subject: Biochemistry

Copyright 2021 Jeng-Yih Chang

ABSTRACT

Single-stranded RNA viruses (ssRNAs) are a group of contagious microbes that have compact structures, containing a protein capsid and a genomic RNA (gRNA) packaged inside. Genome packaging is an essential part of the viral infectious cycle, yet this step is not fully understood. To know the packaging of ssRNA viruses, having a gRNA structure is needed. In this work, we combined cryo-electron microscopy (cryo-EM) and molecule modeling to determine the structures of packaged gRNAs.

Compared to the capsid, the gRNA is structurally flexible, limiting the resolution of its structure solved by cryo-EM. To efficiently refine RNA structures into middle-resolution cryo-EM maps, we developed the Hierarchical Natural Move Monte Carlo (HNMMC) simulation, which takes advantage of collective motions for groups of nucleic acid residues to reduce the computational resources and also preserve atomic details in the models. After successfully refining several cryo-EM models, we used HNMMC simulation to model atomic details and flexibility for two distinct conformations of the gRNA packaged inside the bacteriophage MS2.

We also presented the complete gRNA model of the bacteriophage Q β revealing that the 3' untranslated region binds to the maturation protein and the 4,127 nucleotides fold domain-by-domain. 33 operator-like RNA stem-loops serving as package signals are located, suggesting a pathway for the assembly of Q β virion. Besides, from overexpressing the Q β capsid proteins or through the infection by wild-type Q β phages, we have discovered various forms of the virus-like particles (VLPs), including the

regular $T=3$ icosahedral, larger $T=4$ icosahedral, prolate, oblate forms, and a small prolate form elongated along the 3-fold axis. Overexpressing the shorter RNA fragments increased the percentages of the smaller VLPs, which may be sufficient to encapsidate a shorter RNA.

With the help of cryo-EM and molecule modeling, the structure of a packaged gRNA can be solved and the interaction between the capsid protein and the RNA can also be revealed. These techniques provide a new direction to study the viral RNA packaging in structural virology. Understanding the packaging of ssRNA viruses has an implication of using them for biotechnology, such as RNA perseverance and delivery.

DEDICATION

To my parents whose support made this possible

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Junjie Zhang, who is energetic, passionate, and knowledgeable. It is a great honor to pursue my Ph.D. in his lab. With his help, I have a chance to learn the latest techniques in the field of cryo-electron microscopy (cryo-EM). He always supports me to explore my ideas and provides directions and different views whenever I need them. Without Dr. Zhang, I would not be able to accomplish any of these works.

I am privileged to have Dr. Thomas Meek, Dr. Gregory Reinhart, and Dr. Guan Zhu on my advisory committee. I appreciate helpful discussions and comments. They have taught me how to ask and answer questions in a scientific way. They are always patient and ready to help.

I am very fortunate to have the greatest lab mates who have helped me in every aspect of my lab life. Dr. Kailu Yang is a great person to work with. We learned how to operate the microscopy and set up the lab together in my first two years. Dr. Karl Gorzelnik is an expert in phage biology. He provided the phage specimen and spent lots of time answering my questions about phages. I cannot finish this dissertation without him. Mengqiu Jiang is a very good microscopist. She set up the automatic data collection in the microscopy and gave me lots of advice on the cryo sample preparation. Zhicheng Cui always shared new knowledge in the field of cryo-EM and discussed it with me. Ran Meng and Jirapat Thongchol gave me a lot of help in the wet lab experiments.

I also have some awesome collaborators. Dr. Peter Minary in the Department of Computer Science at the University of Oxford helped me develop the software. Dr. Kevin Week in the Department of Chemistry at the University of North Carolina at Chapel Hill shared the knowledge and the specimen of satellite tobacco mosaic viruses. Dr. Karen-Beth G. Scholthof in the Department of Plant Pathology and Microbiology at Texas A&M University provided the specimen of satellite panicum mosaic viruses.

I would like to thank High-Performance Research Computing at Texas A&M University who provided a lot of help in my computational works. The staff, especially Francis Dang, helped me install the software in the computer cluster and fix several errors when I ran my jobs by the cluster. Also, I have to thank the staff in the Microscopy and Imaging Center. They helped me a lot in the microscopy operation and cryo sample preparation.

Finally, I want to thank my family. I would not be here without their inspiration and love. They have provided all the possible support for me to explore my interests and to pursue a career that excites me.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Junjie Zhang, Thomas Meek, and Gregory Reinhart of the Department of Biochemistry and Biophysics, and Professor Guan Zhu, of the Veterinary Medicine & Biomedical Sciences.

The software in Chapter II was developed by Jeng-Yih Chang and Dr. Junjie Zhang. The software was developed based on the software, MOSAICS from Dr. Peter Minary in the Department of Computer Science, University of Oxford. The result was published in RNA in 2020. Karl Gorzelnik purified the specimen used in Chapter III. The Electrophoretic mobility shift assay in Chapter III was done by Jirapat Thongchol. Some of the data in Chapter IV was collected by Jeng-Yih Chang and Karl Gorzelnik at Baylor College of Medicine. The specimens used in Chapter IV were from Dr. Kevin Week in the Department of Chemistry at the University of North Carolina at Chapel Hill and Dr. Karen-Beth G. Scholthof in the Department of Plant Pathology and Microbiology at Texas A&M University. Karl Gorzelnik and Jirapat Thongchol are both members of the Department of Biochemistry and Biophysics at Texas A&M University.

Joanita Jakana is a member of the NCMI, at Baylor College of Medicine. She helped align the microscope when we went.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

The National Center for Macromolecular Imaging (NCMI) at Baylor College of Medicine, where the JEM3200 is located, is funded by NIH grants P41GM103832 and U24GM116787. I was also supported by grants to Dr. Junjie Zhang, a Welch Foundation grant A-1863, the NSF grant MCB-1902392, and the NIH grants R21AI137696 and P01AI095208. While there is nothing controversial in this dissertation, its contents are solely the responsibility of the author and do not necessarily represent the official views of the funders.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
CONTRIBUTORS AND FUNDING SOURCES.....	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES.....	xii
LIST OF TABLES	xv
CHAPTER I INTRODUCTION	1
1.1 Genome packaging by ssRNA viruses	1
1.1.1 Packaging signals	3
1.1.2 The size of the packaged RNA.....	8
1.1.3 The model of an ssRNA virus assembly	9
1.2 Cryo-electron microscopy as a versatile technique for structural virology	10
1.2.1 Virus structures beyond icosahedral symmetry.....	15
1.3 Outline of this Thesis	17
CHAPTER II HIERARCHICAL NATURAL MOVE MONTE CARLO REFINEMENTS OF FLEXIBLE RNA STRUCTURES WITH CRYO-EM DENSITIES.....	19
2.1 Introduction	19
2.2 Results	24
2.2.1 Optimize HNMMC refinement using the simulated data of tRNA.....	24
2.2.2 Refine the specificity domain (S-domain) of Bacillus subtilis ribonuclease P into the experimental cryo-EM map	28
2.2.3 Model the large-scale conformational flexibility of the “handle” in Mycobacterium tuberculosis (Mtb) ribosomes.....	34
2.2.4 Building atomic models of viral gRNA in the virion	39
2.3 Discussion	47
2.4 Materials and methods	50

2.4.1 Multiscale Natural Move Representation of Macromolecules	50
2.4.2 Temperature-modulated MC Simulation.....	51
2.4.3 Cryo-EM Data Collection for MS2	52
2.4.4 Cryo-EM Data Processing for MS2	52
2.5 Data availability	53
CHAPTER III STRUCTURAL ASSEMBLY OF Q β VIRION AND ITS DIVERSE FORMS OF VIRUS-LIKE PARTICLES	54
3.1 Introduction	54
3.2 Results	58
3.2.1 The complete gRNA model inside the Q β virion.....	58
3.2.2 Operator-like RNA stem-loops dominantly support the pentamers of the CPs.....	65
3.2.3 Q β virions are being made along with VLP of different forms.....	71
3.2.4 The plasticity in the hexamers supports the formation of different VLPs	76
3.3 Discussions.....	82
3.4 Materials and Methods	88
3.4.1 Image processing of the Q β gRNA	88
3.4.2 Modeling of wild-type gRNA	88
3.4.3 Purification of Q β and VLPs using gel-filtration	90
3.4.4 Purification of Q β and VLPs using CsCl gradient	91
3.4.5 Electrophoretic mobility shift assay	92
3.4.6 Cryo-EM specimen preparation and data collection	92
3.4.7 Cryo-EM data processing.....	93
3.4.8 Modeling of VLP capsids.....	94
CHAPTER IV STRUCTURAL ASSEMBLY OF SATELLITE PLANT VIRUSES	96
4.1 Introduction	96
4.2 Results	98
4.2.1 Icosahedral Structure of SPMV and STMV	98
4.2.2 The asymmetric structure of SPMV and STMV	101
4.2.3 Focused Classification of the gRNA inside SPMV and STMV.....	103
4.3 Discussion	106
4.4 Materials and Methods	108
4.4.1 Grid Preparation and Cryo-EM Data Collection.....	108
4.4.2 Cryo-EM image processing.....	109
CHAPTER V CONCLUSIONS AND FUTURE DIRECTIONS.....	111
REFERENCES	118
APPENDIX A THE STRUCTURE OF THE HOLIN-ANTIHOLOIN COMPLEX	140

A.1 Introduction	140
A.2 Results and Discussion.....	142
A.2.1 Single-particle cryo-EM structure of the sT-sRI complex	142
A.2.2 Cryo-EM image of the sT oligomer	144
A.3 Materials and Methods	146
A.3.1 Cryo-EM data collection	146
A.3.2 Cryo-EM image processing and model building.....	146
A.4 References	147

LIST OF FIGURES

	Page
Figure I-1. PS-mediated packaging mechanisms	4
Figure I-2. Icosahedral ssRNA viruses with crystallographic visible RNA structures	11
Figure I-3. Statistics of the EM maps in the EMDB	14
Figure I-4. Asymmetric reconstruction of MS2 and Q β	16
Figure II-1. HNMMC refinement of a deformed tRNA.....	26
Figure II-2. The DOF of the ribonuclease P S-domain	29
Figure II-3. HNMMC refinement of the native structure of the S-domain against the cryo-EM map of the folding intermediate	31
Figure II-4. Dynamics of the handle in the <i>Mtb</i> 70S ribosome depicted by HNMMC....	35
Figure II-5. HNMMC refinement of the handle in <i>Mtb</i> 50S rRNA	36
Figure II-6. HNMMC refinement of the gRNA fragment 2394-2604 inside MS2.....	41
Figure II-7. HNMMC refined models of the complete MS2 gRNA	43
Figure II-8. Two cryo-EM maps of MS2 with different gRNA conformations.....	44
Figure II-9. Two conformations of the gRNA inside MS2	46
Figure II-10. Dynamics of MS2 gRNA at each nucleotide in the two conformations.....	47
Figure II-11. The comparison between the model refinements in HNMMC and MDFF for the handle of the <i>Mtb</i> ribosome.....	50
Figure III-1. Cut-through view of the asymmetry cryo-EM map of Q β	57
Figure III-2. The complete model of the dominant gRNA fold in Q β	59
Figure III-3. The interaction of the Mat and the 3' end of Q β	60
Figure III-4. The 3' UTR of the Q β gRNA	62
Figure III-5. Specific RNA stem-loops at the 3' end of the gRNA required for infectious particle formation.....	63

Figure III-6. Conservation of the two kissing-loops in the Q β -like phages	65
Figure III-7. Operator-like RNA stem-loops in the gRNA that interact with the capsid shell.....	67
Figure III-8. CP-sandwiched and CP-anchored RNA stem-loops	70
Figure III-9. Structures of the Q β VLPs.....	73
Figure III-10. The 2D class averages of Q β VLPs.....	74
Figure III-11. The purification of ssRNA phages and VLPs using gel-filtration chromatography.	74
Figure III-12. The triangulation numbers of the small prolate.....	75
Figure III-13. The extra densities inside the oblate and small prolate VLPs	76
Figure III-14. The comparison of CP dimers in different forms of Q β VLPs.....	78
Figure III-15. The comparison of pentamers in different forms of Q β VLPs.	79
Figure III-16. Plasticity in the hexamers of different Q β VLPs.....	80
Figure III-17. The structure of Q β CP monomer	81
Figure III-18. Proposed model for the co-replicative assembly of the Q β virion	84
Figure III-19. The geometrical relationship between the oblate, $T=3$, and prolate VLPs	85
Figure III-20. The Fourier Shell Correlation (FSC) curves	94
Figure IV-1. The structures of SPMV and STMV	99
Figure IV-2. The interaction of the CP and the RNA in SPMV and STMV	100
Figure IV-3. The composition of the SPMV and STMV EM maps	101
Figure IV-4. Asymmetric reconstruction of SPMV and STMV	103
Figure IV-5. Focused classification with capsid-signal subtraction	104
Figure IV-6. Focused classification of the SPMV gRNA.....	105
Figure IV-7. Focused classification of the STMV gRNA.....	106

Figure V-1. The RNA densities inside different forms of VLPs	115
Figure A-1. The structure of sT-sRI complex.	143
Figure A-2. Cryo-EM images of sT oligomers	145

LIST OF TABLES

	Page
Table II-1. The C4' RMSD of the tRNA models refined by different EM weights against simulated density maps at different resolutions	27
Table II-2. 18 representative models of the S-domain	33
Table II-3. The refinements of the handle in 40 classified cryo-EM density maps of the <i>Mtb</i> 70S ribosome	38
Table III-1. Operator-like RNA stem-loops	68
Table III-2. CP-sandwiched RNA stem-loops.	69
Table III-3. CP-anchored RNA stem-loops.....	69
Table III-4. Percentages of different Q β VLPs via different purification methods	73
Table V-1. The internal volumes of different VLPs	114

CHAPTER I

INTRODUCTION

1.1 Genome packaging by ssRNA viruses

Single-stranded RNA (ssRNA) viruses infect all domains of living organisms, including bacteria, plants, and animals (Baker, Olson, & Fuller, 1999; Gorzelnik et al., 2016; Kolakofsky, 2015; Rao, 2006; White, 1990). One key step in the viral infectious cycle is to correctly package the genomic RNA (gRNA) into the capsid to form an infectious nucleocapsid. Different from double-stranded DNA and RNA viruses that hydrolyze ATP by enzymes to package their gRNA into a well-formed empty capsid (Aksyuk & Rossmann, 2011; Gaussier, Yang, & Catalano, 2006; Hohn, 1976; Kainov, Tuma, & Mancini, 2006), the gRNA packaging of ssRNA viruses is a spontaneous process without an ATP requirement. The nucleocapsid assembly and RNA packaging in the ssRNA virus are coupled with each other and depend on RNA-protein and protein-protein interactions (Bruinsma, Comas-Garcia, Garmann, & Grosberg, 2016; Cadena-Nava et al., 2012; Comas-Garcia et al., 2014; Dykeman, Stockley, & Twarock, 2013a; Elrad & Hagan, 2008; Garmann, Comas-Garcia, Gopal, Knobler, & Gelbart, 2014; Perlmutter & Hagan, 2015).

Generally, the nucleocapsid assembly of an ssRNA virus requires two things: (1) multiple copies of the capsid protein, also known as coat protein (CP); and (2) one copy of the gRNA. Plant viruses basically follow this simple organization (Rao, 2006). The

composition becomes slightly complex in ssRNA bacteriophages (phages) which have an additional maturation protein embedded in the capsid for the host infection (Dai et al., 2017; Gorzelnik et al., 2016). Animal viruses are even more complex. Most animal viruses contain glycoproteins to identify the appropriate receptors from the host and mechanically help the virus enter the cell (Baker et al., 1999; White, 1990). Although the formation of the nucleocapsids is variant, all these nucleocapsids have three primary biological imperatives: (1) to stably survive in the external environment as an infectious particle for transmission between host individuals; (2) to optimally interact with the host machinery to trigger a successful infection; and (3) to be easily disassembled after the infection to facilitate the release of the infectious genome (Rao, 2006).

The packaging of the genome in ssRNA viruses is a complex story. This process includes several interactions and factors (Comas-Garcia, 2019). One of the reasons that make the study of ssRNA viruses so interesting is that even though viruses are very far away from each other in evolution, they still share a large number of similarities (e.g., physical, biochemical, or biological properties). Therefore, the finding from one particular virus is easy to apply to all members of its family, or even to all viruses. However, the details, like packaging specificity, assembly pathways, and nucleation sites, can be very different even between closely related viruses.

1.1.1 Packaging signals

The RNA contains a large number of negative charges that can neutralize the positively charged surfaces seen on many viral CPs (Callanan et al., 2018; Koonin, Dolja, & Krupovic, 2015; Losdorfer Bozic & Podgornik, 2018). With non-self or random RNAs, *in vitro* experiments have shown show that CPs can correctly co-assemble and still form virus-like particles (VLPs) (Cadena-Nava et al., 2012; Comas-Garcia, Cadena-Nava, Rao, Knobler, & Gelbart, 2012). The lack of sequence specificity of the packaged RNA shows that assembly can be driven by electrostatics. However, these *in vitro* results contradict the apparent discovery *in vivo* that almost all the virions encapsidate only the viral gRNA (Routh, Domitrovic, & Johnson, 2012).

Genome packaging should be a highly specific process. The viral gRNA must be distinguished from other cellular RNA molecules when the assembly takes place. Several experiments have demonstrated that the RNA packaging in an ssRNA virus is selected by some highly specific interactions between the gRNA and the CP (Borodavka, Tuma, & Stockley, 2012; Ford et al., 2013; Patel et al., 2017; Sorger, Stockley, & Harrison, 1986). The RNA segment with a particular sequence/structure responsible for this high-affinity interaction is known as a “packaging signal” (PS). These PSs not only serve to designate ‘self’ RNA for the viral coat proteins but also make the assembly process more efficient (Dykeman, Stockley, & Twarock, 2010; Sorger et al., 1986). The PSs with high CP affinity serve as nucleation sites to initiate the formation of capsids (Stockley et al., 2013). The PSs contribute to the gRNA packaging in a variety of

mechanisms, for example, attracting the CPs, inducing the conformational switching of the CPs for a higher affinity of CP-CP interactions, or both (Figure I-1).

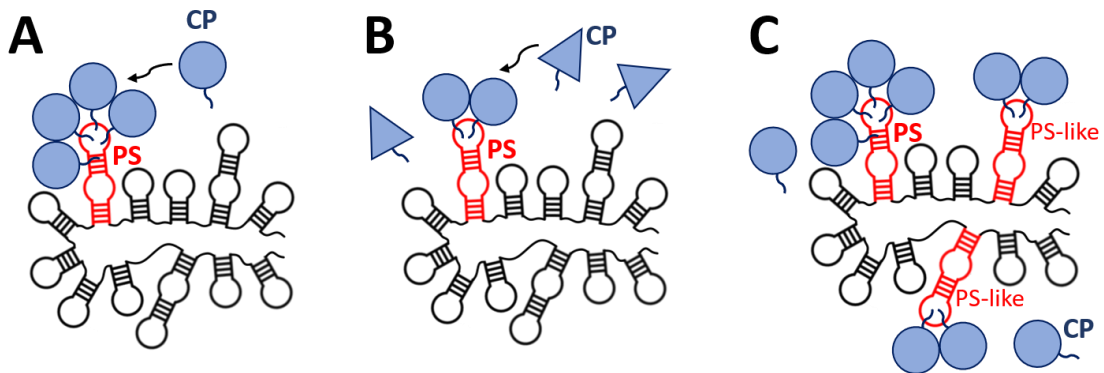


Figure I-1. PS-mediated packaging mechanisms

(A) High-affinity binding mechanism. Selective packaging of the gRNA is mediated by the high-affinity interaction between the CP and the PS. The PS is one RNA fragment with a specific sequence present only in the gRNA. (B) PS allosteric mechanism. The CP binds to the PS resulting in a conformational switch of the CP from one with weak CP-CP interactions to strong interactions. (C) Multiple PS-like fragments. There are one PS and a series of PS-like fragments distributed along with the gRNA. All of them can bind to the CP with high-affinity.

The tobacco mosaic virus (TMV), a helical ssRNA plant virus, is one of the best-studied viruses in regards to PS and the assembly processes (Butler, Finch, & Zimmern, 1977; Fraenkel-Conrat & Williams, 1955; Zimmern, 1977; Zimmern & Butler, 1977). The specific interaction between the CP and a 69-nt RNA fragment at the open reading frame of the gene of movement protein has been shown to cause the specificity of virion assembly (Zimmern, 1977). However, under non-physiological conditions, TMV CPs can be still assembled into VLPs in the absence of RNAs (Durham & Klug, 1971; Mandelkow, Holmes, & Gallwitz, 1976; Mandelkow, Stubbs, & Warren, 1981). These

experiments indicate that the capsid assembly does not necessarily require the PS or the RNA *in vitro*, but one thing has to be noticed that the solution conditions used in the experiments are different from a physiological condition. These non-physiological conditions may change the CP-CP interaction to overcome the absence of a nucleic acid template. Besides, TMV can also be assembled when heterologous RNAs are provided, but the assembly by these heterologous RNAs is less efficient (Fraenkel-Conrat & Williams, 1955). Compared to the homologous RNA, the RNA from yeast or polyA slow down the growth of the viral helical rod (Butler, 1972; Fraenkel-Conrat & Williams, 1955). These results support the conclusion that the PS is a nucleation site for the assembly and this RNA fragment can lower the activation energy of this process (Stockley et al., 2013). Moreover, the nucleocapsid assembly can be inhibited or slowed down by disrupting the PS (Turner & Butler, 1986). The binding of small organic molecules (antofine analogs) to the PS is sufficient to block the formation of the nucleocapsid (Gao, Zhang, Yu, & Xi, 2012). The inhibition mediated by the drug shows that TMV assembly highly relies on the presence of a PS.

Among ssRNA viruses with icosahedral symmetry, the phage MS2 and the satellite tobacco necrosis virus (STNV) are two of the viruses having characterized PSs. The PS of MS2 is a 19-nt stem-loop, named the ‘operator’ (or sometimes TR), that serves as a translational regulator for the replicase cistron (Valegard, Murray, Stockley, Stonehouse, & Liljas, 1994). The PS of STNV is a 16-nt stem-loop, named B3, at nucleotide positions 57-81 (Bunka et al., 2011). Besides having a high-affinity with the

CPs, both of the PSs are proposed to be able to promote the efficiency of the CP-CP interactions. For MS2, the binding of the TR stem-loop by the CP induces a conformational switch of the CP dimer from a C/C-like (symmetric) to A/B-like (asymmetric)(Dykeman et al., 2010). The assembly process is very slow if only the symmetric or asymmetric dimers exist, however, once both types of dimers are present, the formation of the $T=3$ MS2 capsid will be extremely fast (Basnak et al., 2010; Morton et al., 2010). Similarly, the CP of STNV in solution is a monomer and does not assemble in the absence of RNA (Bunka et al., 2011). Unlike MS2 where there is a conformational change in the CP upon binding RNA, in STNV the RNA B3 stabilizes the N-terminus of the STNV CP to promote the interaction between different CPs (Ford et al., 2013). The helices at the N-terminus from three adjacent CP subunits bind at the three-fold axis of the capsid with the B3 binding. The N-terminus of the STNV CP contains several positively charged amino acids, therefore RNA binding helps the CPs overcome an electrostatic barrier for assembly (Bunka et al., 2011).

While these two RNA-coat protein interactions are very well characterized, these viruses can still package their gRNA even if those sites are not present (Kotta-Loizou, Peyret, Saunders, Coutts, & Lomonossoff, 2019; Peabody, 1997). Besides using one PS, selective packaging can also depend on a series of PS-like fragments distributed within the gRNA. The Stockley and Twarock groups found the PS-like fragments through several *in vitro* approaches (Borodavka et al., 2012; Borodavka, Tuma, & Stockley, 2013; Patel et al., 2017). They first used SELEX (Systematic Evolution of Ligands by

Exponential Enrichment) experiments (Stoltenburg, Reinemann, & Strehlitz, 2007) to test a large library of RNA aptamers and select a small set of sequences with a high binding affinity and specificity for a given CP. Then, they used single-molecule fluorescence correlation spectroscopy (smFCS) to evaluate these selected RNA aptamers through the *in vitro* assembly of CPs (Borodavka et al., 2012). Through smFCS, extremely diluted concentrations of CP and RNA can be used, and therefore non-specific protein-RNA interactions that could drive non-selective packaging could be excluded. Finally, combining with mass spectrometry and bioinformatic approaches, the CP-specific binding RNA fragments can be located within the genome of interest and defined as the PS-like fragments (Patel et al., 2017; Rolfsson et al., 2016). These findings suggest that the assembly is selected by not only one PS but also a series of PS-like fragments distributed along the genome. Also, the virus assembly may have a better modulation through these PS-like fragments. More PS-like fragments may mean that the entire assembly process can accommodate more mutations that may disrupt the packaging if only one PS is used. Although it still needs to be determined if the biological activity of the PS-like fragment is similar to that of the PS, more and more available data agree with the “multiple PS-like” scenario. For example, the putative PSs of the hepatitis C virus (HCV) identified *in vitro* by SELEX experiments can contribute to RNA packaging (Stewart et al., 2016). In addition, the PS for HIV-1 is proposed to be a series of unpaired guanosine residues distributed on the 5'UTR and Gag open reading frame, like multiple PS-like fragments (Comas-Garcia et al., 2018; Keane et al., 2015).

1.1.2 The size of the packaged RNA

Besides specific sequences, the length of the packaged RNA also strongly contributes to the packaging efficiency. One *in vitro* experiment has shown that the ssRNA length strongly contributes to the packaging efficiency of CCMV (Comas-Garcia et al., 2012). Without considering the specific RNA-CP, the relative packaging efficiencies of CCMV CP increase with RNA length, while the highest efficiency is at the RNA with 4.0 kb which is similar to the length of wild-type. Moreover, when the two RNAs have the same charge and overall size, the packaging efficiency will be affected by the RNA secondary or tertiary structures. Similar results of the assembly related to the RNA length can also be found in cultured cells (Krishna, Marshall, & Schneemann, 2003).

The physical size of the capsid is another reason that constrains the packaging of the gRNA. The dimension of the capsid imposes an upper limit on the amount of nucleic acid that can be accommodated. The packaging capability of turnip crinkle virus (TCV) demonstrates the size limitation of a viral gRNA. TCV is a $T=3$ icosahedral virus with a 30-nm diameter and the gRNA of 4054-nt is encapsidated in the purified TCV virion. When the wild-type TCV and a tomato bushy stunt virus (TBSV)-TCV chimera were co-transfected to the protoplasts, the isolated virions of TCV contained only the 2044-nt sub-genomic RNA of the chimera instead of the larger full-length gRNA of 4665 nts (Qu & Morris, 1997). However, similar co-transfection assays performed with a truncated version of the TBSV-TCV chimera resulted in the packaging of 4133-nt gRNA. These

observations suggest that, even with the required PS, TCV cannot accommodate RNA larger than 4600 nts.

1.1.3 The model of an ssRNA virus assembly

The self-assembly of a regularly shaped capsid with RNA packaged involves a fine-tuned range of interactions between the CPs and the RNA to create this highly ordered, symmetrical molecule. The mechanism of assembly of an ssRNA virus was proposed for the phage MS2 by two steps: (1) nucleation-and-growth and (2) extension (Borodavka et al., 2012). Before assembly, the protein-free RNA exists in an ensemble of conformations that holds well folded secondary structures that favor assembly. In the first step, CPs bind to high-affinity RNA stem-loops to initiate the assembly, while the RNA-induced conformational change increases the CP-CP binding resulting in a compact initiation complex. This stage mainly depends on the specific recognition of viral RNA sequences and the following CP-CP interactions. In the second stage, other CP subunits are recruited to grow the shell. This recruiting process basically relies on the CP concentration. The two-stage mechanism can be applied to many ssRNA viruses that have high-affinity PSs. On the other hand, for nonspecific RNAs, the assembly may use a different mechanism that bypasses the first step and forms heterogeneous initiation complexes often leading to fused or partial capsids.

1.2 Cryo-electron microscopy as a versatile technique for structural virology

The key to understanding the mechanism of virus assembly is to know the structures involved. X-ray crystallography has been one of the major approaches in the studies of viral structures. The first high-resolution structures of icosahedral viruses from X-ray crystallography were solved around 1980 (Abad-Zapatero et al., 1980; Harrison, Olson, Schutt, Winkler, & Bricogne, 1978; Rossmann et al., 1985). These early viral structures presented the first atomic view of viral CPs and how these CPs are assembled into an icosahedral capsid. The structures of these CPs show high similarities between viruses from different species, particularly in the interior scaffold of β -strands, which was known as the “jelly-roll fold”. Later, lots of the icosahedral viruses have been determined to atomic resolutions by crystallography (Rossmann, 2013). Some of them show a small portion of the gRNA density that lies on the edge of the capsid (Figure I-2), such as Satellite Tobacco Mosaic Virus (STMV) (Larson, Day, Greenwood, & McPherson, 1998), Bean Pod Mottle Virus (BPMV) (Chen et al., 1989), and Pariacoto Virus (PaV) (L. Tang et al., 2001).

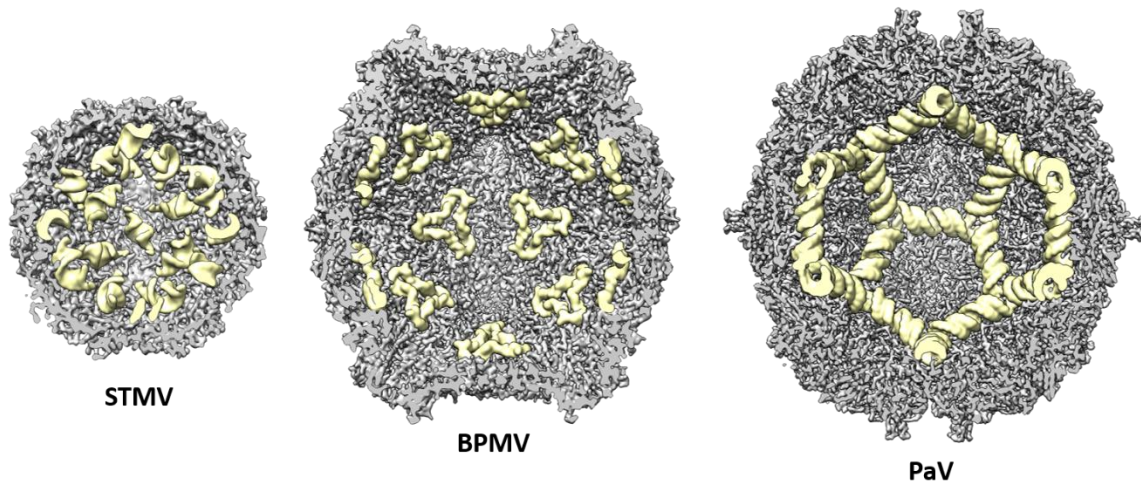


Figure I-2. Icosahedral ssRNA viruses with crystallographic visible RNA structures Cut-through view of STMV, BPMV, and PaV shows the distribution of icosahedral ordered internal RNA. The capsid and RNA are represented in grey and yellow, respectively.

However, virus crystallography has several limitations: (1) Crystallography requires crystallization, but some viral particles or viral complexes are hard to be crystallized, especially the complexes of viruses and their receptors; (2) The high-resolution crystal requiring good diffraction could only be produced in some icosahedral viruses, which are usually small and non-enveloped, but many of human infectious viruses are lipid enveloped and/or pleomorphic, such as retroviruses (HIV), flaviviruses (Dengue virus), coronaviruses (SARS-CoV); (3) The symmetries of structural features in the virus may not be consistent. The information of non-icosahedral proteins or RNAs would be smeared in crystal packing dominated by the icosahedral capsids; (4) Crystallography can only present one conformation of the virus that may have different packaging in native.

Alternatively, the single-particle cryo-electron microscopy (cryo-EM) has been a technique widely used in studying the three-dimensional (3D) structure of biological specimens under near-native conditions (Glaeser, 2007). In cryo-EM, the biological specimen is frozen quickly enough at an extremely low temperature (-180 °C), so ice crystals do not have time to form, resulting in a thin vitrified ice layer embedding the sample. These ice-embedded specimens then are imaged in a transmission electron microscope with a low temperature kept (below -160 °C). By combining the images of many identical particles in different orientations, the structure of interest can be reconstructed to a high resolution.

Cryo-EM started in the mid-1970s (Frank, 1975; Henderson & Unwin, 1975; Taylor & Glaeser, 1976). The procedure of a rapid plunge-freezing was improved during the 1980s (Adrian, Dubochet, Lepault, & McDowell, 1984; Chiu, 1986). The first “near-atomic” resolution (3.8 Å) of cryo-EM was obtained from rotavirus in 2008 by Nikolaus Grigorieff and his colleagues (X. Zhang et al., 2008). For the icosahedral EM map at near-atomic resolution, most of the amino acids of the inner capsid can be built de novo. The molecular interactions and predict functionally active regions can also be determined.

In 2013, the introduction of direct electron detectors (DEDs) led to a high-resolution revolution in cryo-EM. Compared to the traditional photographic film and charge-coupled device (CCD), the DED has very high sensitivity and can collect high-

quality data in movies (X. Li et al., 2013). The frames of each movie acquired in a very short period of time (usually 0.2 seconds/frame) enable researchers to calculate and compensate for the beam-induced motion. Moreover, the DED can count the total electron dose used in the image collection. It largely limits the radiation damage on the susceptible specimens, like small biological molecules. The new detector helps cryo-EM be able to easily apply to the smaller biological structures that cannot be solved in crystallography, such as membrane proteins. For example, the work of a transient receptor potential channel (TRPV1) determined at 3.4 Å resolution is the first time that such high resolution could be achieved on a membrane protein with low-order symmetry by single-particle cryo-EM (Liao, Cao, Julius, & Cheng, 2013).

Besides the breakthrough of the detector, the development of automated data collection (Mastronarde, 2005; Tan, Cheng, Potter, & Carragher, 2016) and data processing software (Punjani, Rubinstein, Fleet, & Brubaker, 2017; Scheres, 2012; G. Tang et al., 2007b; Zheng et al., 2017) also reduces the time required for structural determination by cryo-EM. By these state-of-art techniques, cryo-EM has a rapid rise in structural biology. In the early of 2020, there have been more than 10,000 EM maps deposited in the EM Data Bank (<http://www.ebi.ac.uk/pdbe/emdb>) while 40% of them have a resolution higher than 5 Å (Figure I-3). The highest resolution of the deposited map is also improved every year. Recently, the single-particle cryo-EM has been shown to determine the structures of apoferritin, a stable iron-storing protein around 505kDa, to a resolution of approximately 1.2 Å (Nakane et al., 2020; Yip, Fischer, Paknia, Chari, &

Stark, 2020). The data quality of the EM map is already high enough to resolve the individual atoms in apoferritin.

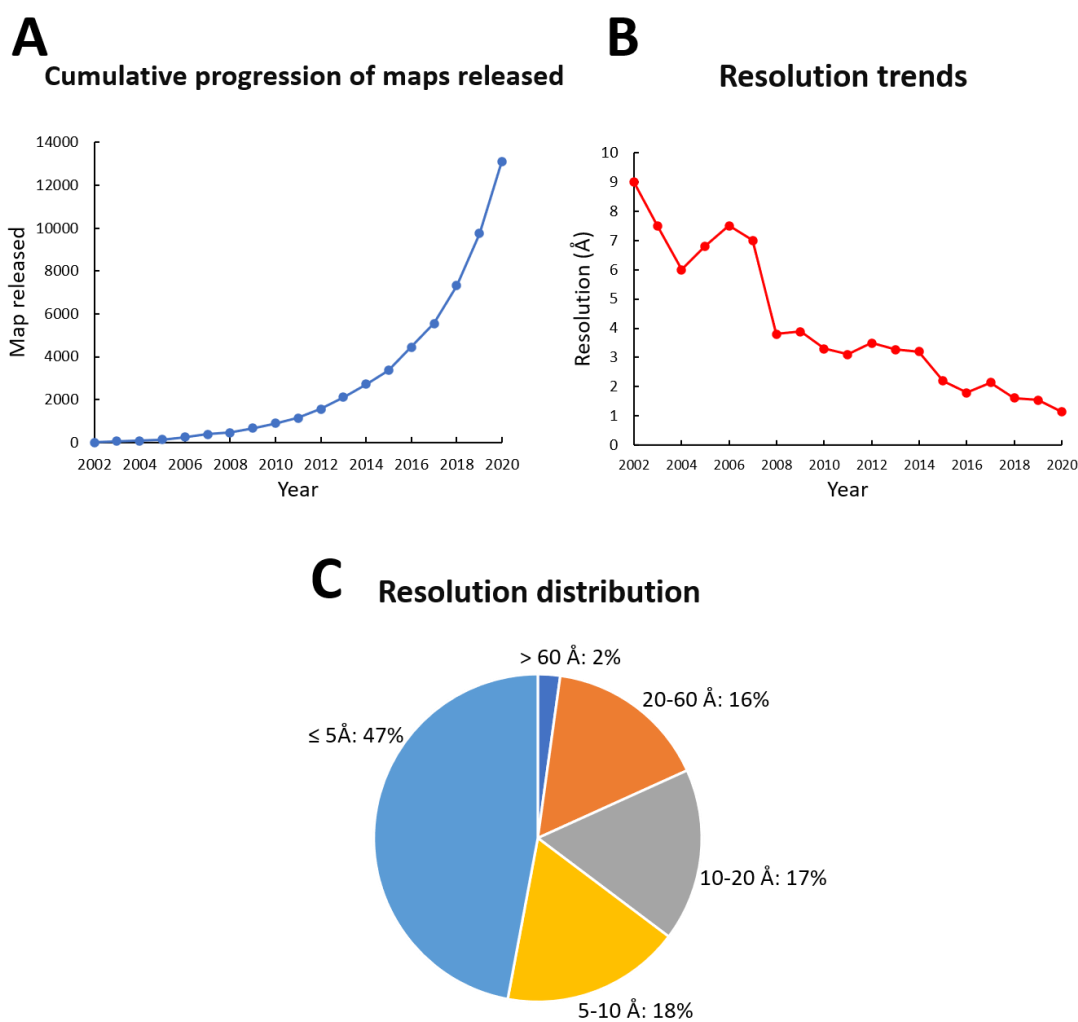


Figure I-3. Statistics of the EM maps in the EMDB

(A) Number of maps deposited in the EMDB. (B) Highest resolutions of the structures annually deposited in the EMDB. (C) Map resolution distribution in the EMDB.

(https://www.ebi.ac.uk/pdbe/emdb/statistics_main.html)

Due to the advantages of large mass and high symmetry, cryo-EM has been widely used to determine structures of icosahedral viral capsids. Most of the time cryo-EM only requires a considerably short time to obtain useful data of the viral structure. For example, the structure of the Zika virus is solved by cryo-EM within months of the outbreak (Kostyuchenko et al., 2016; Sirohi et al., 2016). To date, lots of icosahedral viruses are solved at higher than 3-Å resolution, such as Adeno-associated virus (1.9 Å)(Tan et al., 2018), human enterovirus D68 (2.2 Å)(Y. Liu et al., 2018), rhinovirus B14 (2.3 Å)(Dong et al., 2017), human picobirnavirus (2.7 Å)(Ortega-Esteban et al., 2020) and coxsackievirus A10 (2.7 Å)(Zhu et al., 2018). At this high resolution, side-chain densities should be clearly recognized and the protein backbone could be easily traced and built de novo.

1.2.1 Virus structures beyond icosahedral symmetry

However, icosahedral symmetry basically only exists in the viral capsid, while the gRNA or other packaged proteins usually do not follow the same symmetry. After being averaged by icosahedral symmetry reconstruction, the densities of asymmetric components will be obscured or erroneously averaged into artifacts. This symmetry-mismatch problem can easily be solved by the image processing with no symmetry applied in cryo-EM. For example, high-resolution asymmetric cryo-EM reconstruction of phage MS2 shows a capsid with one maturation protein embedded and a defined packaged gRNA (Dai et al., 2017) (Figure I-4). Through the asymmetry map, 16 high-affinity RNA stem-loops with one binding with the maturation protein are confirmed and

80% of the genome is traced. These high-affinity RNA stem-loops may serve as PSs. At the same time, the asymmetry EM map of the phage Q β , an MS2 related phage, is also obtained (Cui et al., 2017; Gorzelnik et al., 2016) (Figure I-4). Besides binding with a maturation protein, the gRNA of Q β also interacts with the additional internal CP dimer inside the Q β capsid.

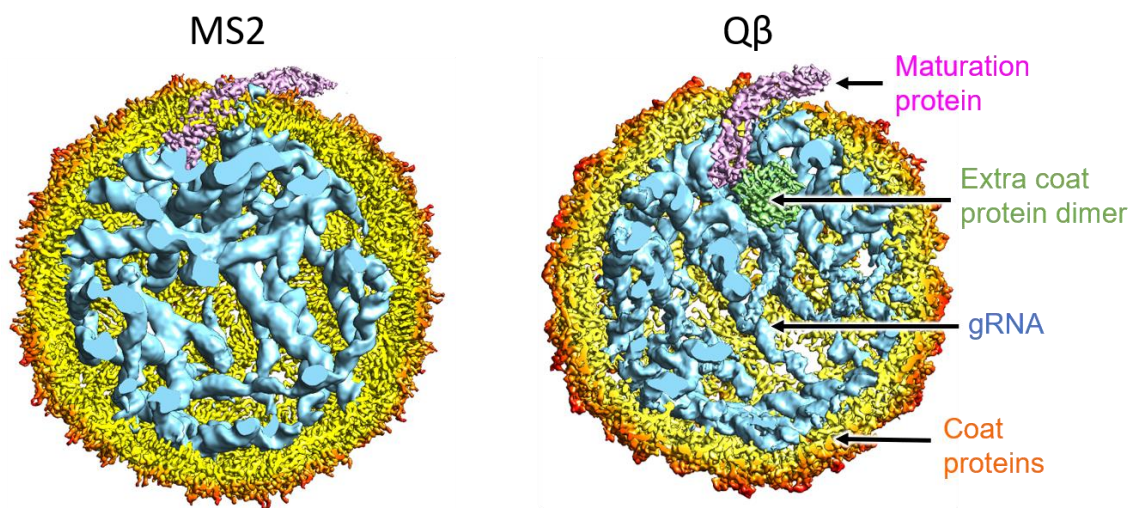


Figure I-4. Asymmetric reconstruction of MS2 and Q β

Cut-through view of MS2 and Q β asymmetric EM maps shows defined gRNAs with the maturation proteins. The capsids, maturation proteins, and presumed gRNAs are colored yellow, magenta, and light blue, respectively. The extra CP dimer inside the Q β capsid is colored green.

Another interesting question in structural virology is the interaction between the virus and its host. For studying this question, cryo-electron tomography (cryo-ET) is a particularly suitable technique. In cryo-ET, the target of interest is imaged multiple times at different angles (often at 1-2° intervals with +/- 60°) to provide different orientations that are combined to reconstruct a 3D structure. With the help of Liposomes, mini-cells,

or extracted host membranes, cryo-ET has been applied to study the entry of viruses, including influenza virus (Fontana, Cardone, Heymann, Winkler, & Steven, 2012), HIV (Sougrat et al., 2007), and bacteriophages (Hu, Margolin, Molineux, & Liu, 2013; Xu, Gui, Wang, & Xiang, 2016). On the other hand, the structure of the phage MS2 attached to its receptor, the F-pili, was recently solved by single-particle cryo-EM (Meng et al., 2019). With a similar idea as cryo-ET, the single-particle images of the MS2 and pili were taken at different tilted angles to deal with the preferred orientation problem from the long filamentous pili structures. By the near-atomic resolution EM map, the interactions at the MS2 and pili binding surface were determined and the slight orientation variation of the complex was also discovered.

1.3 Outline of this Thesis

This thesis will focus on realizing the ssRNA packaging of bacteriophage Q β and two small satellite plant viruses, STMV and SPMV (satellite panicum mosaic virus). Compared to that of human viruses, the packaging systems of these three viruses are relatively simpler but still very complex. To model the packaged gRNA of Q β , in chapter 2, we developed our in-house modeling method HNMMC, which was implemented in the MOSAICS-EM software to efficiently model the RNA into a middle-resolution map (~ 6 Å) with secondary structure restraint. In chapter 3, I used HNMMC simulation to build the gRNA model based on our asymmetry cryo-EM map of Q β . Then, in chapter 4, I applied cryo-EM on STMV and SPMV to obtain the

structural information of packaged gRNA. Finally, we end with a final remark and future direction for our work on the ssRNA packaging of viruses in chapter 5.

CHAPTER II ¹

HIERARCHICAL NATURAL MOVE MONTE CARLO REFINEMENTS OF FLEXIBLE RNA STRUCTURES WITH CRYO-EM DENSITIES

2.1 Introduction

Ribonucleic acids (RNAs) are important biological macromolecules that perform various functions inside the cell. In addition to acting as a carrier of genetic information (Crick, 1970), RNAs play active roles in cell survival and development, such as regulation of gene expression (Garst, Edwards, & Batey, 2011; Ha & Kim, 2014) and catalysis of protein synthesis (Cech, Zaug, & Grabowski, 1981; Guerrier-Takada, Gardiner, Marsh, Pace, & Altman, 1983; Schluenzen et al., 2000). Due to these essential roles, the mal-function of RNA is implicated in many diseases, including neurological diseases (Esteller, 2011; Fenoglio, Ridolfi, Galimberti, & Scarpini, 2013), cardiovascular diseases (Uchida & Dimmeler, 2015), and cancers (Huarte, 2015).

To perform their diverse functions, many RNAs need to precisely fold into their three-dimensional (3D) structures. The folding process of RNA is hierarchical (Woodson, 2010). The primary RNA sequence first forms secondary structures, such as helices, bulges, junctions, and loops, through local nucleotide interactions. These

¹ The basis for this chapter comes from “Hierarchical Natural Move Monte Carlo Refines Flexible RNA Structures into Cryo-EM Densities” by Jeng-Yih Chang, Zhicheng Cui, Kailu Yang, Jianhua Huang, Peter Minary and Junjie Zhang. *RNA*. 2020 Dec;26(12):1755-1766. Figures are republished with permission.

secondary structures then arrange into a tertiary structure, via long-range RNA interactions. While the secondary structures of RNA form within a short period (10^{-4} to 10^{-5} seconds), the tertiary structures of RNA are formed much slower (10^{-2} to 10^{-1} seconds), through a process which involves complex interactions between RNA secondary structures, as well as the interactions between RNAs and neighboring proteins (Crothers, Cole, Hilbers, & Shulman, 1974; Onoa & Tinoco, 2004). In addition, the secondary structures of RNA are generally more stable compared to the tertiary structures (D. H. Mathews, Sabina, Zuker, & Turner, 1999). Taking these facts into account, it is intuitive to model RNA structures in a hierarchical way. First, the secondary structures in RNA can be obtained by computational predictions, through methods such as comparative sequence analysis (Cannone et al., 2002; David H. Mathews, Moss, & Turner, 2010; Seetin & Mathews, 2012) and/or free-energy minimization (Eddy, 2004; D. H. Mathews & Turner, 2006). The accuracy of these predictions can be improved by incorporating the experimental mapping data of secondary structures, using chemical/enzymatic probing and RNA sequencing (Loughrey, Watters, Settle, & Lucks, 2014; Lu & Chang, 2016; Smola, Rice, Busan, Siegfried, & Weeks, 2015). Then the tertiary structures of RNA can be sampled by spatially arranging these RNA secondary structures.

Recent advances in cryo-electron microscopy (cryo-EM) have enabled the structural characterization of RNA molecules at unprecedented resolutions. One major advantage of using electrons to image RNA is that the phosphate backbone of RNA

strongly scatters electrons, giving a much higher contrast compared to proteins. Moreover, better imaging hardware (X. Li et al., 2013; McMullan, Faruqi, Clare, & Henderson, 2014; Ruskin, Yu, & Grigorieff, 2013) and more robust image processing software (Grigorieff, 2007; Moriya et al., 2017; Punjani et al., 2017; Scheres, 2012; G. Tang et al., 2007a) have enabled obtaining 3D density maps of ribosomal RNA (rRNA) to atomic (or near-atomic) resolutions (Fischer et al., 2015; Z. Liu et al., 2017; Shalev-Benami et al., 2016; X. Zhang et al., 2016). Many protein-RNA complexes, for example, spliceosomes and telomerases have also been determined to near-atomic or subnanometer resolutions (Jiang et al., 2015; Nguyen et al., 2015; Yan et al., 2015). New imaging strategies in cryo-EM, such as utilizing the direct-electron counting and phase contrast (Danev & Baumeister, 2016; Danev, Tegunov, & Baumeister, 2017; Khoshouei, Radjainia, Baumeister, & Danev, 2017), may facilitate the imaging of smaller RNA molecules (Baird, Ludtke, et al., 2010; K. Zhang et al., 2018). Furthermore, viral genomic RNAs (gRNAs), generally long and flexible in solution, can also be imaged and characterized by cryo-EM (Gopal, Zhou, Knobler, & Gelbart, 2012; H. Liu & Cheng, 2015; X. Zhang et al., 2015), with recent breakthroughs to reveal defined gRNA 3D folds inside the capsids of single-stranded RNA bacteriophages (Cui et al., 2017; Dai et al., 2017; Gorzelnik et al., 2016; Koning et al., 2016b; Meng et al., 2019; Zhong et al., 2016). Importantly, cryo-EM holds another advantage to study multiple states of a macromolecule in solution, allowing the mapping of an energy landscape in the conformational space (Dashti et al., 2014). Therefore, cryo-EM is gaining popularity for

imaging and generating 3D density maps of RNA at increasingly higher resolutions under different functional states.

Several computational methods have been developed to build atomic models of RNA into the cryo-EM maps. This process generally involves two steps: (1) building a starting model and (2) refining it into a density map. If the target RNA has a homologous sequence to a known RNA template structure, comparative modeling can be used to generate a starting model (Rother, Rother, Puton, & Bujnicki, 2011). In cases where a homologous template is unavailable but the target RNA is short (100-300 nucleotides), it is possible to predict the starting model *ab initio* (Cheng, Chou, & Das, 2015; Das, Karanicolas, & Baker, 2010). Otherwise, one can manually build fragments of the starting model then assemble them into the cryo-EM density before further refinements (Jossinet, Ludwig, & Westhof, 2010).

After building the starting model, refining the atomic coordinates of RNA into the cryo-EM map requires fitting as many atoms as possible into the density while keeping the proper stereochemistry of the model. Algorithms based on molecular dynamics calculations or simulated annealing optimizations are usually used in the model refinement (Liebschner et al., 2019; McGreevy, Teo, Singharoy, & Schulten, 2016). The accuracy of the refined model can be further improved by correcting errors in the RNA backbone (Chou, Sripakdeevong, Dibrov, Hermann, & Das, 2013). Most of these refinement tools need extensive computational resources and rely on high-quality initial

models. If the starting model has a significant structural variation compared to the target structure, it leads to two major problems: (1) inability to dock the starting model into the density with the correct orientation for further refinement; (2) prolonged time for the model refinement due to the inefficiency in sampling large conformational changes between the starting and target structure. Therefore, there is a need for a tool that can objectively and efficiently sample RNA conformations from a starting model to facilitate the modeling process.

Recently, we have developed a method called Hierarchical Natural Move Monte Carlo (HNMMC), which improves the conformational sampling efficiency in macromolecular modeling. Using Natural Move Monte Carlo with customizable hierarchical degrees of freedom (DOFs), we showed that the method can generate diverse canonical ensembles of macromolecular conformations (Sim, Levitt, & Minary, 2012). Implemented in the MOSAICS software package, HNMMC has been successfully used in several applications (Behrmann et al., 2015; Knapp, Demharter, Deane, & Minary, 2016; Krawczyk, Sim, Knapp, Deane, & Minary, 2016; Minary & Levitt, 2014). The combination of MOSAICS and EMAN2, a cryo-EM data processing software package (G. Tang et al., 2007a), resulted in a molecular refinement tool (named MOSAICS-EM), has been applied to the structure refinement of protein complexes from cryo-EM data (J. Zhang, Minary, & Levitt, 2012). Here we adapted MOSAICS-EM to refine RNA structures against cryo-EM maps. For example, a starting structural model of RNA can be obtained by comparative modeling, *ab initio* prediction, or manual

building, and before the final refinement, HNMMC in MOSAICS-EM can sample the possible structural variations of this initial RNA model within the density map, in a large scale, based on the customized secondary structural restraints, to maximize the coverage for a correct conformation.

In this chapter, we first established and validated HNMMC refinement with a known structure of tRNA. We then used it to automatically model the structure of the folding intermediate of the specific domain of ribonuclease P from *Bacillus subtilis*. In addition, we refined a structural model of an rRNA expansion segment, also known as the handle, in the ribosome of *Mycobacterium tuberculosis (Mtb)*, into its cryo-EM maps of 40 different states. We then applied this tool to build the atomic model of the gRNA fragment (2394-2604) from a single-stranded RNA virus, MS2, and finally expanded the same strategy to obtain two models of the complete gRNA with distinct conformations, which reveal multiple pathways for the viral assembly.

2.2 Results

2.2.1 Optimize HNMMC refinement using the simulated data of tRNA

The challenges in refining macromolecular structure using Markov Chain Monte Carlo (MC) simulations are usually the large number of DOFs. To overcome this problem, we group residues from a macromolecule into different levels of segments, representing different DOFs in the molecule. The grouped residues in each segment, termed natural move, translate and rotate collectively and can fluctuate individually

within the segment. Therefore, these natural moves are usually constructed from the rigid or semi-rigid natural organizations of a molecule, such as helices of an RNA. For example, the cloverleaf-like tRNA structure is modeled as four natural moves with each helix as one segment. The centers, allowing each helix to rotate and translate, were defined within the connecting residues (Figure II-1A). An energy function, which combines the molecular energy and the weighted EM energy, was used for the MC simulation with a modulated temperature profile (Figure II-1D).

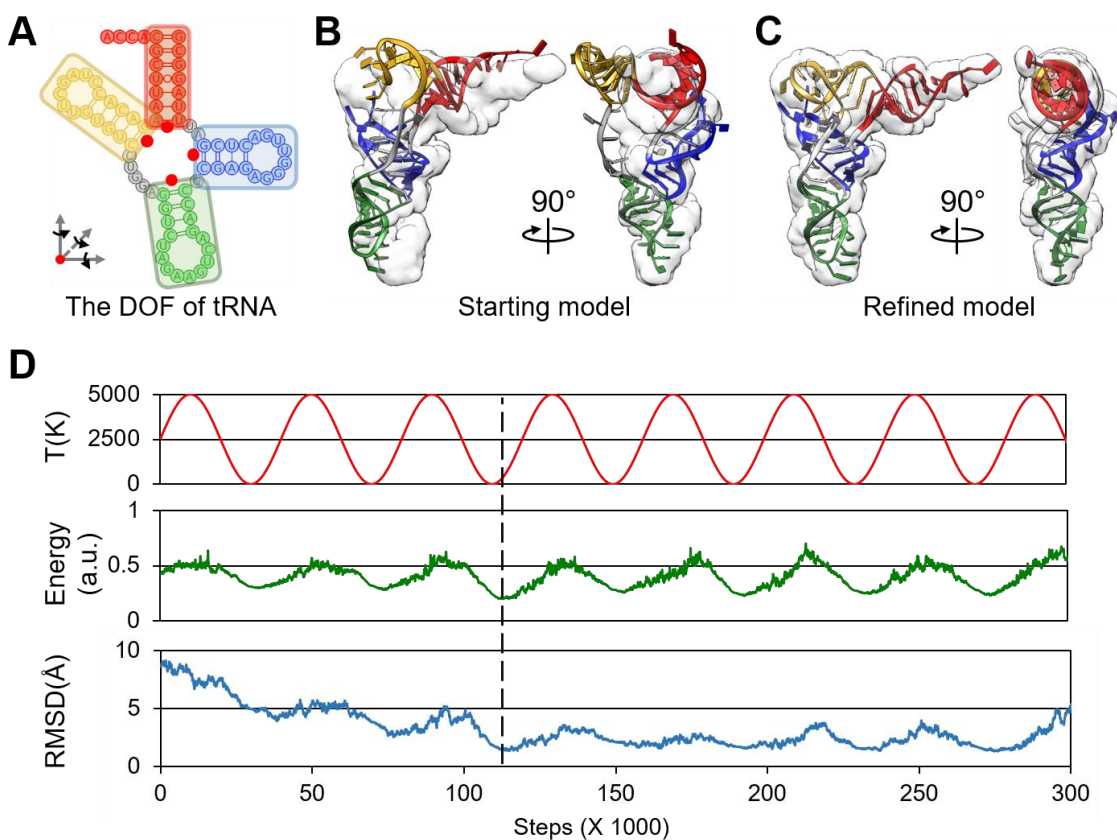


Figure II-1. HNMCMC refinement of a deformed tRNA

(A) The DOF of a tRNA. Each helical stem (red, blue, green, and yellow) of a tRNA is defined as a natural move (indicated in each box) that can translate and rotate independently, with each red dot as a center of rotation and translation. Two flexible unpaired loops are colored gray. (B) The starting tRNA model with 8.72-Å C4' RMSD to the target structure fitted in the target map. (C) The refined model with the lowest total energy (E_{total}) has a C4' RMSD of 1.37Å to the target structure. Each helical stem of the tRNA model is colored red, blue, green, and yellow, respectively, based on the DOF defined in Panel A. The transparent density is Gaussian-blurred from the target structure to 5-Å resolution. (D) The temperature, T (red), the total energy (green), and the C4' RMSD (blue) to the target structure are shown as a fluctuation against the refinement steps. The dashed line indicates the step when the structure with the lowest total energy is achieved.

To evaluate and optimize the protocol of HNMCMC refinement, different weights of the EM energy and different resolutions of the density map were used. Five target

maps of tRNA, Gaussian-blurred to resolutions from 5 Å to 30 Å (Table II-1), were generated from the crystal structure (PDB ID: 1EHZ) (Shi & Moore, 2000) by EMAN2 (G. Tang et al., 2007a). An initial model of tRNA was deformed from the crystal structure by MC simulation at a high temperature (300,000K), resulting in a C4' root-mean-square deviation (RMSD) with 8.27 Å (Figure II-1B), which was subsequently refined into the density maps.

Table II-1. The C4' RMSD of the tRNA models refined by different EM weights against simulated density maps at different resolutions

The red color indicates the refined results with the lowest RMSD.

		EM weight (ω)										
		0.001	0.01	0.1	1	10	50	100	200	500	1000	∞
Map resolution (Å)	5	2.69	6.00	4.30	3.30	2.09	1.90	1.92	1.37	2.03	1.92	3.24
	10	4.55	4.72	4.94	2.87	1.77	1.87	1.65	1.80	1.97	1.99	2.21
	15	6.16	7.03	4.56	3.22	2.09	2.39	2.25	1.90	2.09	2.55	2.46
	20	4.43	4.13	4.12	3.13	2.61	3.34	2.24	2.52	2.98	2.81	2.74
	30	3.44	4.17	4.59	2.91	2.97	3.43	3.74	2.54	2.48	3.61	2.56

The best-refined result was from the EM weight of 200 and the map at a 5-Å resolution, which confers the smallest C4' RMSD (1.37 Å) to the target structure (Figure II-1C, Table II-1). The synchronous fluctuation of the total energy and C4' RMSD with the annealing temperature indicates that the refinement is capable of escaping from local energy minima efficiently (Figure II-1D).

The refinement results of using different EM weights and map resolutions show that the higher resolution of the map yields a better-refined model. Increasing the EM weight can lead to a better-refined result when the map resolution is low. The refinement performs best when the EM weight is 200. Using this weight, all the simulated maps can guide the deformed model to the refined models with around 2-Å C4' RMSD to the target structure. The incorporation of the molecular energy prevents nonphysical conformations. The EM weight of 200 was also applied in the following HNMMC simulations of different molecules.

2.2.2 Refine the specificity domain (S-domain) of Bacillus subtilis ribonuclease P into the experimental cryo-EM map

The S-domain of the *Bacillus subtilis* ribonuclease P is a 154-nucleotide RNA that can bind pre-tRNA directly (Qin, Sosnick, & Pan, 2001). The folding behavior of the S-domain in solution has been previously characterized (Baird, Fang, Srividya, Pan, & Sosnick, 2007; Baird, Gong, et al., 2010; Baird, Westhof, Qin, Pan, & Sosnick, 2005). The conformation of a folding intermediate of the S-domain at high Mg^{2+} concentration was captured using single-particle cryo-EM (Baird, Ludtke, et al., 2010). Compared to the native state (PDB ID: 1NBS), this folding intermediate only forms secondary structures and short-range tertiary interactions, implying that the long-range tertiary interactions are formed in sequential steps during RNA folding (Baird, Ludtke, et al., 2010). In this example, we employed HNMMC to automatically refine the native

structure of the S-domain into its cryo-EM map of a folding intermediate at 15-Å resolution (EMD-5242).

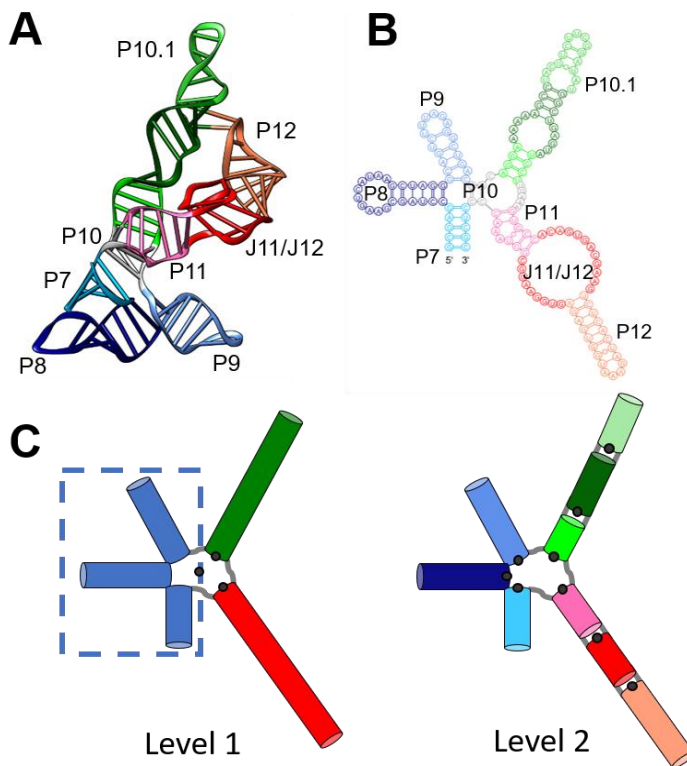


Figure II-2. The DOF of the ribonuclease P S-domain

(A) The crystal structure of the ribonuclease P S-domain. (B) The secondary structure of the S-domain. (C) The levels of DOF are defined based on the secondary structure. Level 1 separates the S-domain into three helical arms. Helices P7, P8, and P9 are combined into one group (the blue rectangle). Level 2 further separates Level 1 into more regions for a finer refinement. In Level 2, Helices P7, P8, and P9 are independent. Helices P11, P12, and Loop J11/J12 (the red arm in level 1) are divided into three regions. The long helix P10.1 (the green arm in Level 1) is also divided into three regions.

In the native state, the core of the S-domain has a T-loop motif (three-way junction). Helix P10.1 connects with two coaxially stacked short Helices P10 and P11

(Krasilnikov, Yang, Pan, & Mondragon, 2003) (Figure II-2A). P10 forms a four-way junction with Helices P7, P8, P9. P11 connects with P12 through two large internal loops, J11/12. Because it is a three-way junction, we can first define the DOFs of the S-domain as three natural moves: (1) P7/P8/P9, (2) P10.1, and (3) P11/J11/12/P12 (blue, green, and red, respectively in DOF 1, Figure II-3A). P10 and the loops at the junction are defined as flexible regions (grey in DOF 1, Figure II-3A, Figure II-2C). For finer adjustment of the model inside the density, we also set the second DOF defining each helix as one natural move and the large internal loops J11/12 as another natural move (DOF 2, Figure II-3A, Figure II-2C).

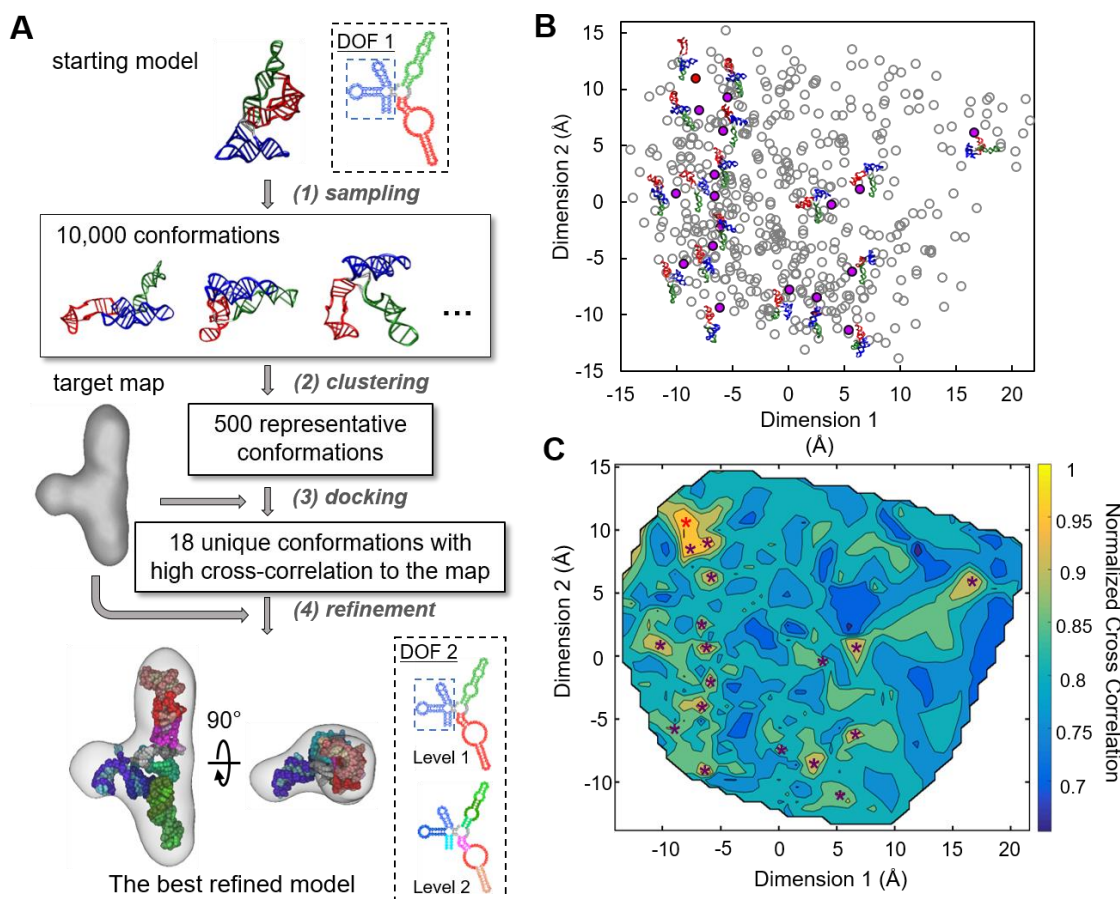


Figure II-3. HNMMC refinement of the native structure of the S-domain against the cryo-EM map of the folding intermediate

(A) The flow chart of the refinement protocol of the S-domain. DOF 1 is applied in the first step to generate 10,000 conformations without the restraint from the cryo-EM density. DOF 2 with two levels is applied in the last step for a detailed refinement. (B) The conformational difference between the 500 representative conformations from clustering the 10,000 conformations. The scatter plot shows the differences among the 500 models by multidimensional scaling computed using the matrix of the pairwise C4' RMSD among the 500 conformations. Each point represents one conformation. The 18 unique conformations with high cross-correlation to the target map are marked as solid purple or red dots with the red dot representing the conformation with the highest cross-correlation. (C) The heat map of the normalized cross-correlation of the 500 representative conformations to the target map. The 18 unique conformations with high cross-correlations are marked as stars. The red star at the top-left corner indicates the conformation with the highest cross-correlation.

To avoid bias from users, we customized the protocol, which automatically refines the native structure of the S-domain against the cryo-EM map of its folding intermediate (Figure II-3A). To accomplish that, one needs to first register the model and the map in roughly matched orientations. However, due to the large conformational difference between the initial model (native) and the cryo-EM map (folding intermediate), with the best possible fitting cross-correlation of only 0.43, such a direct registration is not reliable. Therefore, we used the following strategy: (1) apply a large-scale MC sampling (10,000 decoys) to broadly search the probable conformations without the EM energy; (2) cluster all 10,000 decoys into 500 models; (3) dock the 500 clustered models to the map by Situs (Wriggers, 2012); (4) pick the representative models with a unique conformation and a high docking score; (5) refine these representative models into the map locally using HNMMC in MOSAICS-EM.

To aid the selection of representative models, multidimensional scaling analysis by the R package (Team, 2015) was applied to visualize the conformational differences of the 500 clustered models (Figure II-3B). In addition, the docking scores (or normalized cross-correlation between the model and the cryo-EM map) of the 500 models were used to generate a heat map (Figure II-3C) on top of the plot of the multidimensional scaling. In total, 18 representative models (solid circles and stars in Figure II-3B and II-3C, respectively; Table II-2) were selected and refined into the cryo-EM map with the finer DOF, DOF 2. The refined model with the highest cross-correlation (0.88) to the map was chosen as the best-refined model which is in agreement with the model manually built

and validated previously (Baird, Ludtke, et al., 2010). The two long helical arms (P11/J11/12/P12 and P10.1) form a collinear arrangement in both our model and the previous model and the orientations of the short arm (P7/P8/P9) are also the same.

Table II-2. 18 representative models of the S-domain

The models are arranged through their cross-correlation score (from large to small) to the target map (transparent grey). The three helical arms are colored as Level 1 in Supplementary Figure 2.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18		

2.2.3 Model the large-scale conformational flexibility of the “handle” in *Mycobacterium tuberculosis* (*Mtb*) ribosomes.

We then expanded our application to refine the flexible region of rRNA inside the *Mtb* ribosome. Different from other bacterial ribosomes, the 23S rRNA in the large subunit (50S) of the *Mtb* ribosome has a 100 nucleotide-long helical expansion segment H54a, termed “handle” (Yang et al., 2017). The handle undergoes a large-scale conformational change when the small subunit (30S) associates with the 50S to form the *Mtb* ribosome (70S). In the 70S, the handle serves as an inter-subunit bridge of the 50S and the 30S and swings between the L1 stalk of the 50S and the mRNA exit site. Therefore, the handle may modulate the translation in the *Mtb* and affect the association of the 50S and 30S subunits (Yang et al., 2017).

To more quantitatively measure the flexibility of the handle in the *Mtb* 70S, HNMMC was applied to refine its model from the bent state in the 50S into the extended states in the 70S, captured by 40 classified cryo-EM maps (Yang et al., 2017). The initial model of the handle with its connected 3-way junction (rRNA fragment 1540-1659 and 1814-1828) was cut from the high-resolution structure of *Mtb* 50S (PDB ID: 5V7Q) (Yang et al., 2017). The target maps with 40 unique conformations of the handle were segmented from the 40 classified maps of the *Mtb* 70S (downloaded as a bundle through EMD-8648) (Figure II-4A) (Yang et al., 2017). The handle consists of several short helices and motifs, such as the sarcin-ricin motif and the tetra-loop (magenta, Figure II-

5). Seven levels of natural moves were defined to describe its bending and extending motions (Figure II-5A).

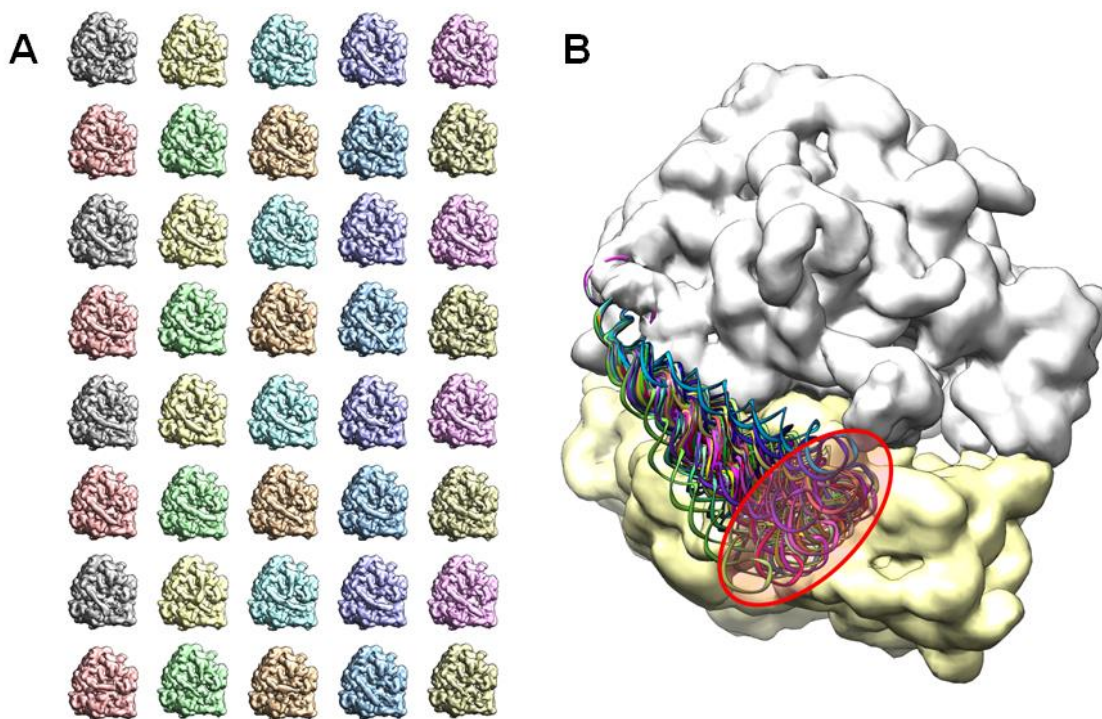


Figure II-4. Dynamics of the handle in the *Mtb* 70S ribosome depicted by HNMMC (A) 40 cryo-EM maps of the *Mtb* 70S ribosome with different conformations of the handle. (B) 40 models of the handle are refined by HNMMC into the classified density maps. The densities of the 50S and 30S subunits are colored white and yellow, respectively. The red oval outlines the range of motion for the tip of the handle which is around 60 Å in the short axis and 70 Å in the long axis.

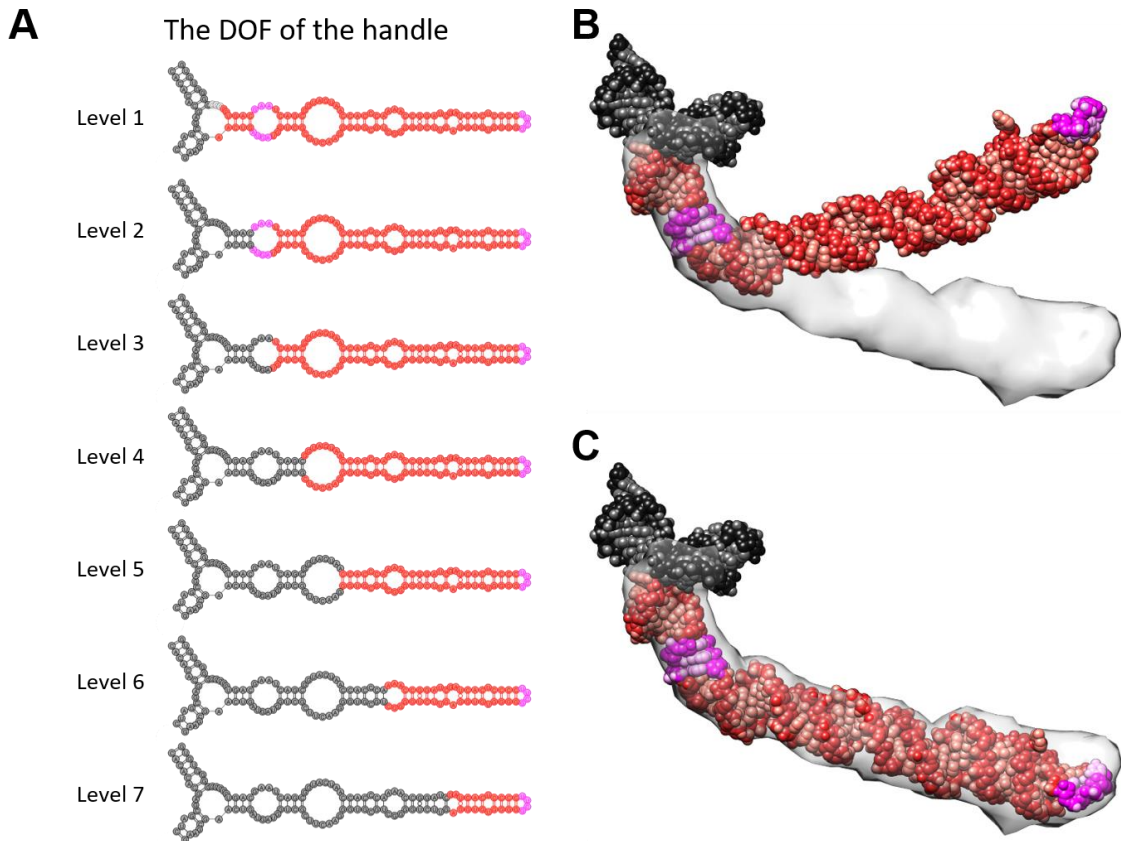


Figure II-5. HNMMC refinement of the handle in *Mtb* 50S rRNA

(A) Seven levels of the DOF to describe possible motions of the handle. For each level, the red fragment can move as a group while the black fragment is fixed. The magenta labels the sarcin-ricin motif (close to the three-way junction) and the tetraloop (at the tip of the handle). Combining these seven levels can effectively describe the bending of the handle at each junction. Note that each nucleotide is still allowed to fluctuate around its position within the red fragment for subtle conformational variations. (B) The initial model of the handle obtained from the *Mtb* 50S is docked against the cryo-EM density of the handle from the *Mtb* 70S. (C) The HNMMC refined model of the handle.

HNMMC refinements against the 40 maps improved the cross-correlations between the models of the handle and their corresponding maps from 0.2~0.4 to 0.6~0.7 (Table II-3). For the segment between the 3-way junction and the sarcin-ricin motif,

limited movement is observed when the entire handle extends from the conformation in the 50S to the one in the 70S. However, for the segments from the sarcin-ricin motif to the tip of the handle, it swings around 35° when forming the 70S from the 50S (Figure II-5B and II-5C). Moreover, the handle is also flexible in the 70S. Comparing these 40 refined models, the tip of the handle can travel around 60 \AA in the short axis and 70 \AA in the long axis (Figure II-4B). While the solvent exclusion volume and solvent-accessible area of a single conformation of the handle are around 71.9 nm^3 and 156.1 nm^2 , respectively, the 40 conformations of the handle lead to a much larger solvent exclusion volume and solvent-accessible area of 265.4 nm^3 and 301.6 nm^2 . Such increased solvent accessibility may help the handle efficiently interact with potential regulatory subunits or non-ribosomal factors (Z. Li et al., 2018; Mishra, Ahmed, Tyagi, Shi, & Bhushan, 2018; Polikanov, Blaha, & Steitz, 2012). Such a hypothesis is yet to be tested.

Table II-3. The refinements of the handle in 40 classified cryo-EM density maps of the *Mtb* 70S ribosome

Map number	Cross-Correlation Score (%)		Map number	Cross-Correlation Score (%)	
	Initial	Refined		Initial	Refined
1	39.32	63.67	21	29.33	68.98
2	29.11	62.57	22	25.34	67.80
3	21.43	69.50	23	24.70	71.44
4	24.65	68.84	24	36.99	62.31
5	31.03	65.56	25	25.72	69.77
6	32.03	70.83	26	26.36	67.88
7	27.08	67.28	27	29.10	69.20
8	29.48	68.94	28	29.55	69.34
9	30.51	65.28	29	28.36	67.53
10	37.21	59.92	30	30.72	68.29
11	28.06	67.49	31	27.11	62.97
12	34.02	59.88	32	29.63	64.47
13	30.37	66.48	33	27.79	70.69
14	42.35	58.69	34	25.74	70.76
15	28.19	70.35	35	30.19	70.28
16	25.65	72.33	36	25.81	62.91
17	20.50	69.97	37	42.48	57.59
18	24.99	61.21	38	30.05	64.49
19	27.11	66.35	39	27.45	60.81
20	34.02	59.88	40	30.65	63.75

2.2.4 Building atomic models of viral gRNA in the virion

Single-stranded (ss) RNA viruses use RNA as their genetic material and infect all domains of life, including animals, plants, and bacteria. Recently, it has been revealed by cryo-EM that the ssRNA bacteriophages, such as Q β and MS2, have a major 3D conformation of the genomic RNA (gRNA) among all viral particles (Cui et al., 2017; Dai et al., 2017; Gorzelnik et al., 2016; Koning et al., 2016b; Zhong et al., 2016). The gRNA presents stem-loops, termed the “packaging signals” (Dykeman, Stockley, & Twarock, 2013b; Rolfsson et al., 2016), which form stable interactions with the phage capsid proteins, potentially providing a roadmap for capsid assembly.

In a previous cryo-EM study of MS2 (Dai et al., 2017), a few gRNA stem-loops or packaging signals that directly interact with the viral capsid were solved to high resolutions. However, much of the gRNA densities were only resolved to intermediate resolutions or missing, due to the internal flexibility of the RNA within the capsid. Based on the cryo-EM map and the secondary structure prediction, approximately 80% of the gRNA backbone was manually traced, presenting long RNA helices connected by complex junctions. Although the backbone model shows the organization of the gRNA, the lack of atomic details in many regions prevents the understanding of the accurate RNA/RNA or RNA/protein interactions that can affect the gRNA packing and releasing. Such an atomic model of the gRNA was difficult to obtain due to the lower resolution and missing density throughout the cryo-EM map. Recently, a method has been developed in Rosetta to build atomic models of RNA by assembling fragments of the

RNA model into cryo-EM densities (Kappel et al., 2018) and applied to build 10 gRNA fragments inside the MS2 capsid. For each region, this modeling procedure (known as *De novo* Ribonucleoprotein modeling in Real-space through Assembly of Fragments Together with Electron density in Rosetta or DRRAFTER) started from at least one ideal RNA helix as an anchor point and was used to build the models, which connected the remaining helices with the ideal RNA helices, into their corresponding densities at intermediate resolutions.

Here, we first focused on the gRNA fragment 2394-2604, which was not modeled by the DRRAFTER (Kappel et al., 2018). The density of this region shows a 4-way junction motif. The initial model of gRNA fragment 2394-2604 was generated using *RNA Denovo* in Rosetta (Cheng et al., 2015; Das et al., 2010) based on its secondary structure (Figure II-6A). One stem-loop (corresponding to the gRNA fragment 2468-2481) has a high affinity with the viral capsid, therefore, was resolved to the high resolution previously (PDB ID: 5TC1) (Dai et al., 2017). The rest of the gRNA in this fragment does not directly interact with the capsid and is flexible, thus was resolved at a lower local resolution. Therefore, in our modeling, the high-resolution stem-loop (gRNA fragment 2468-2481) was used as a single fixed anchor while the remaining gRNAs were automatically fitted into the density through two runs of the refinement by four levels of natural moves (Figure II-6A), keeping the correct geometry of the stem-loops.

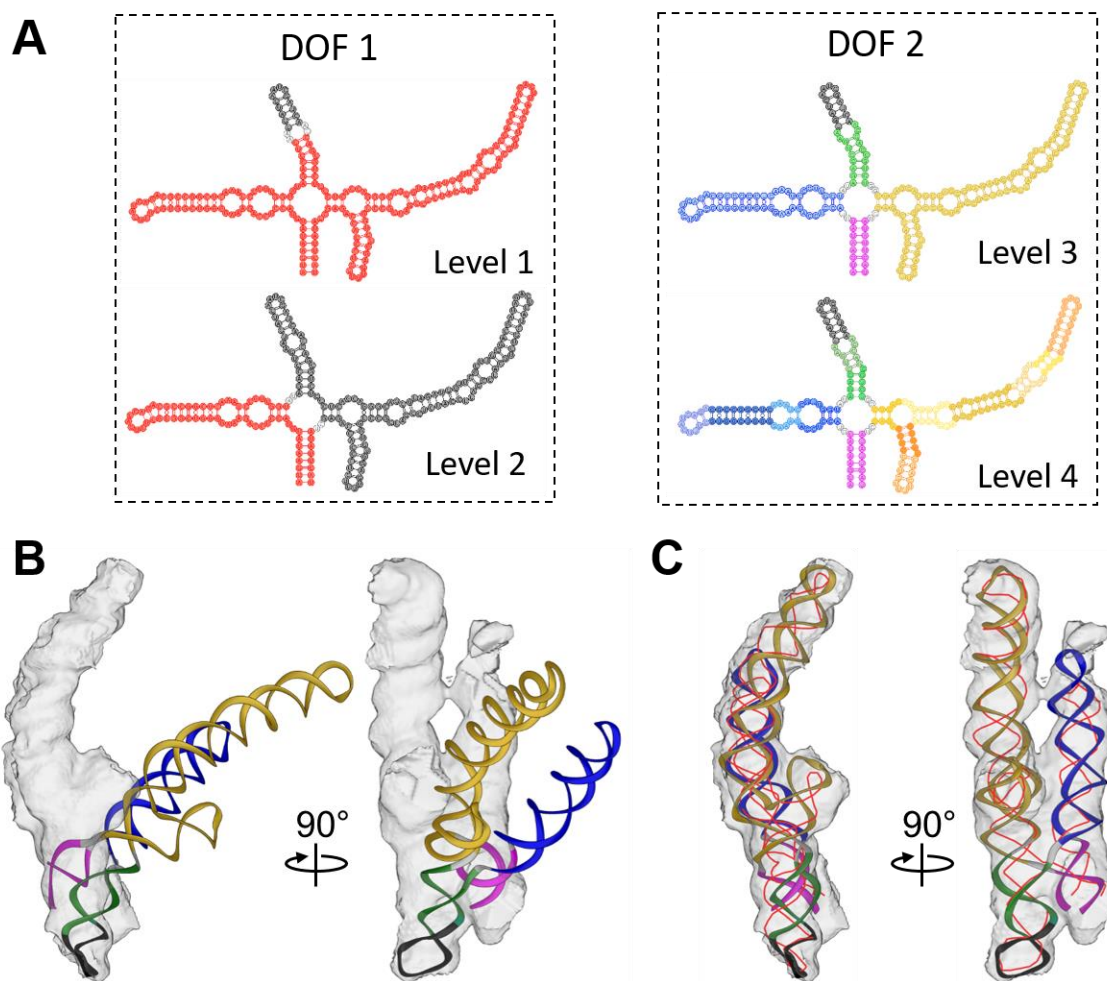


Figure II-6. HNMCMC refinement of the gRNA fragment 2394-2604 inside MS2
 (A) The definition of DOFs of the gRNA fragment 2394-2604. The red fragment in Level 1 and 2 of DOF 1 is the natural move that moves as a group. For Level 3 and 4 of DOF 2, the different colors represent different natural moves. The black region at all levels is fixed. In Level 3 and 4, the black region is the gRNA fragment 2468-2481 with a high-resolution EM structure. The gray color represents unpaired RNA loops that can also move. (B) The initial model of the gRNA fragment 2394-2604 and the corresponding cryo-EM density. The helical stems are colored as Level 3. (C) HNMCMC refined model of the gRNA fragment 2394-2604 against the target cryo-EM density. The thin red line is the backbone trace from the previously published result (Dai et al., 2017).

HNMMC refinement increased the cross-correlation of the model and the density from 0.18 to 0.73 (Figure II-6B and II-C). The refined model is mostly consistent with the manually traced backbone model (thin red line, Figure II-6C) (Dai et al., 2017), which forms a parallel coaxial stacking (cross) of four helical stems. However, compared to the backbone model, our second helical stem (blue in Level 3, Figure II-6A) is 15-Å longer (around eight nucleotides) at the tip. In the map of the entire gRNA, the density of this helix connects with another density, suggesting its interaction with another helical stem. As expected, in the connecting area, an additional density can be found for these eight nucleotides.

Using the same strategy above, in total, we built 17 fragments of the MS2 gRNA, which were connected to establish an atomic model of the entire gRNA with all the 3569 nucleotides. The model was then minimized using Phenix (Liebschner et al., 2019) and further refined against the cryo-EM map by 20 independent runs of HNMMC refinement with each RNA helix defined as one natural move (Figure II-7A). The 20 refined models have improved fitting to the map and maintained secondary structures in the flexible fragments with weak cryo-EM densities. The average C4' RMSD among the 20 refined models is 2.89 ± 0.24 Å and the average cross-correlation between the map and the model is 0.523 ± 0.043 . As expected, comparing these 20 models shows larger variations among the refined models in regions with weaker cryo-EM densities (Figure II-7B).

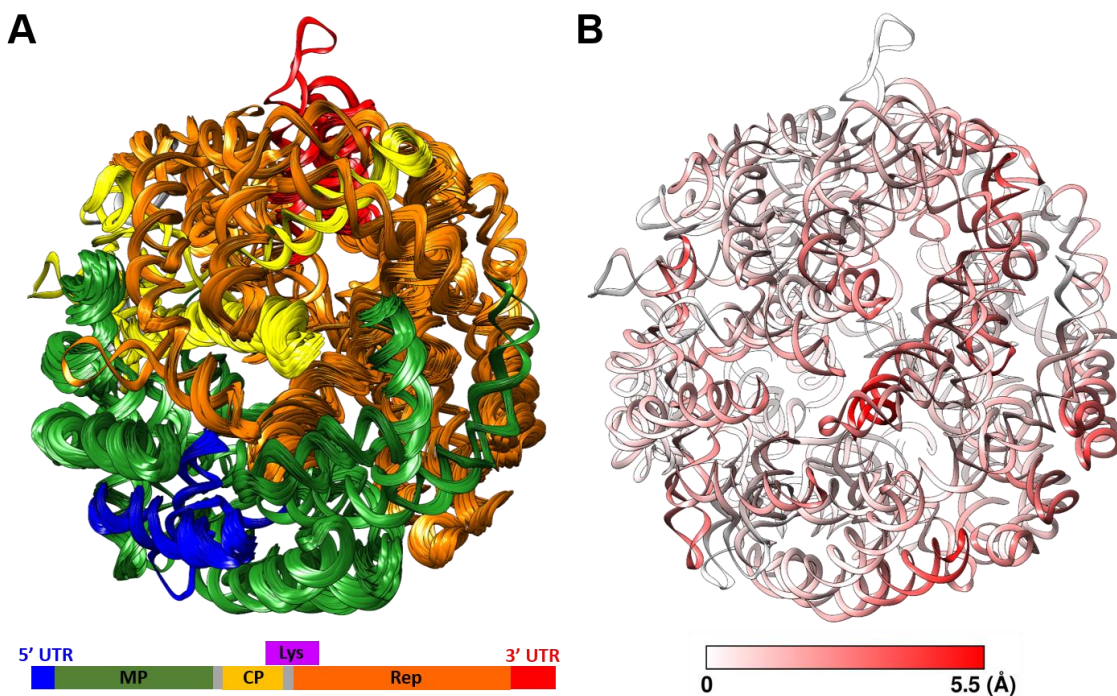


Figure II-7. HNMMC refined models of the complete MS2 gRNA

(A) The ensemble of 20 refined models colored according to the genes in the schematic diagram at the bottom (MP: maturation protein, CP: coat protein, Rep: replicase, Lys: lysis protein). The lysis gene is not colored on the ribbon models due to the overlap with the genes for the coat protein and the replicase. (B) The conformational variation among the 20 refined models at each nucleotide colored by the average C4' RMSD among the 20 models. The color scale from white to red denotes low to high RMSD.

To further understand the structural variations in the MS2 gRNA, a new cryo-EM dataset of MS2 virions are collected and processed. Focused classifications (Bai, Rajendra, Yang, Shi, & Scheres, 2015) around the region with weak RNA density identifies two dominant conformations (named C1 and C2) of the gRNA, existing in 45.2% and 54.8% of the total number of virions, respectively. After the 3D refinement, the final resolutions of the two cryo-EM maps are both 7.4 Å (Figure II-8).

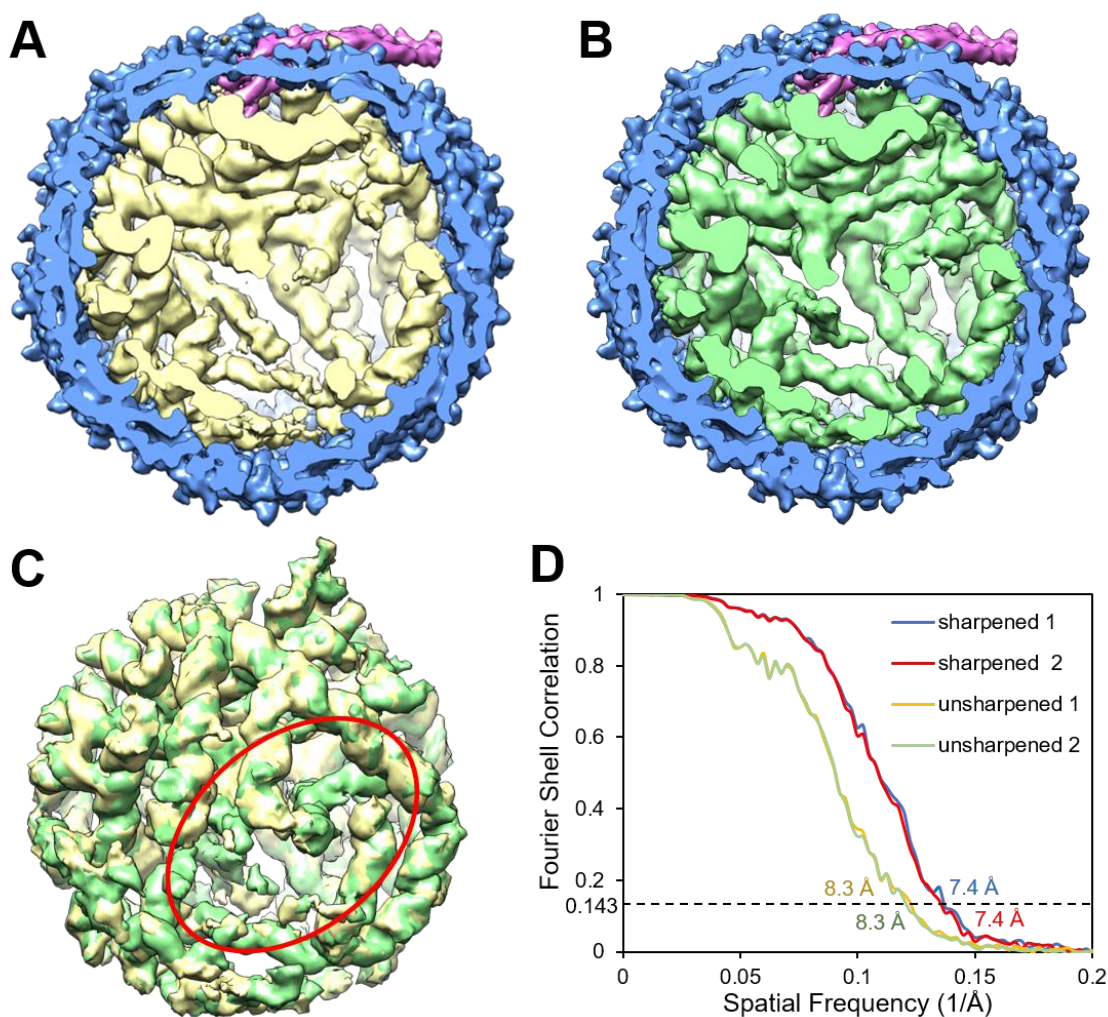


Figure II-8. Two cryo-EM maps of MS2 with different gRNA conformations
 (A) Cut-in view of the cryo-EM map showing gRNA Conformation 1. (B) Cut-in view of the cryo-EM map showing gRNA Conformation 2. (C) The comparison of two gRNA conformations. The region with the largest variation is marked by the red oval. (D) The Fourier Shell Correlation curves of the sharpened and unsharpened maps with two gRNA conformations. The curves of the sharpened maps are shown in blue and red, and the curves of the unsharpened maps are shown in yellow and green for Conformations 1 and 2, respectively. The gRNA densities (yellow and green for Conformations 1 and 2, respectively) in Panel A-C are from the unsharpened map for more completeness.

The atomic models of the gRNA of C1 and C2 are built and refined using HNMMC in MOSAICS-EM (Figure II-9 and II-10). The two models show differences in the flexible RNA fragments, especially of the gRNA fragment 614-879, 1643-1700, 1992-2125, and 2922-3087. The C4' RMSD of these four flexible fragments between two models are 32.58 Å, 19.26 Å, 24.23 Å, and 37.06 Å, respectively, while the C4' RMSD of the rest of the gRNAs is only 4.71 Å. In C1, two coat-protein dimers, D1 and D2 (transparent green and blue densities, Figure II-9) interact with the stem-loops of the gRNA fragment 1643-1700 (green ribbon, Figure II-9) and fragment 1992-2125 (blue ribbon, Figure II-9), respectively. However, in C2, D1 switches its interaction to a stem-loop from the gRNA fragment 2922-3087 (red ribbon, Figure II-9) and D2 interacts with the stem-loop of gRNA fragment 1643-1700 (green ribbon, Figure II-9). Moreover, in C2, due to the switch of gRNA fragment 2922-3087 (red ribbon, Figure II-9) to interact with D2, the interior space of the virion, which was originally occupied by this fragment in C1 is freed to accommodate another gRNA fragment 614-879 (yellow ribbon, Figure II-9). Such a rearrangement of the gRNA inside the capsid leads to a denser RNA packaging towards the maturation protein in C2 as compared to C1. Direct morphing between these two gRNA conformations shows collisions among themselves, indicating that the gRNAs are folded into such conformations during its condensation before the capsid is formed. This result is a piece of direct evidence that the gRNA of an RNA virus can be packed in multiple conformations through different RNA folding pathways.

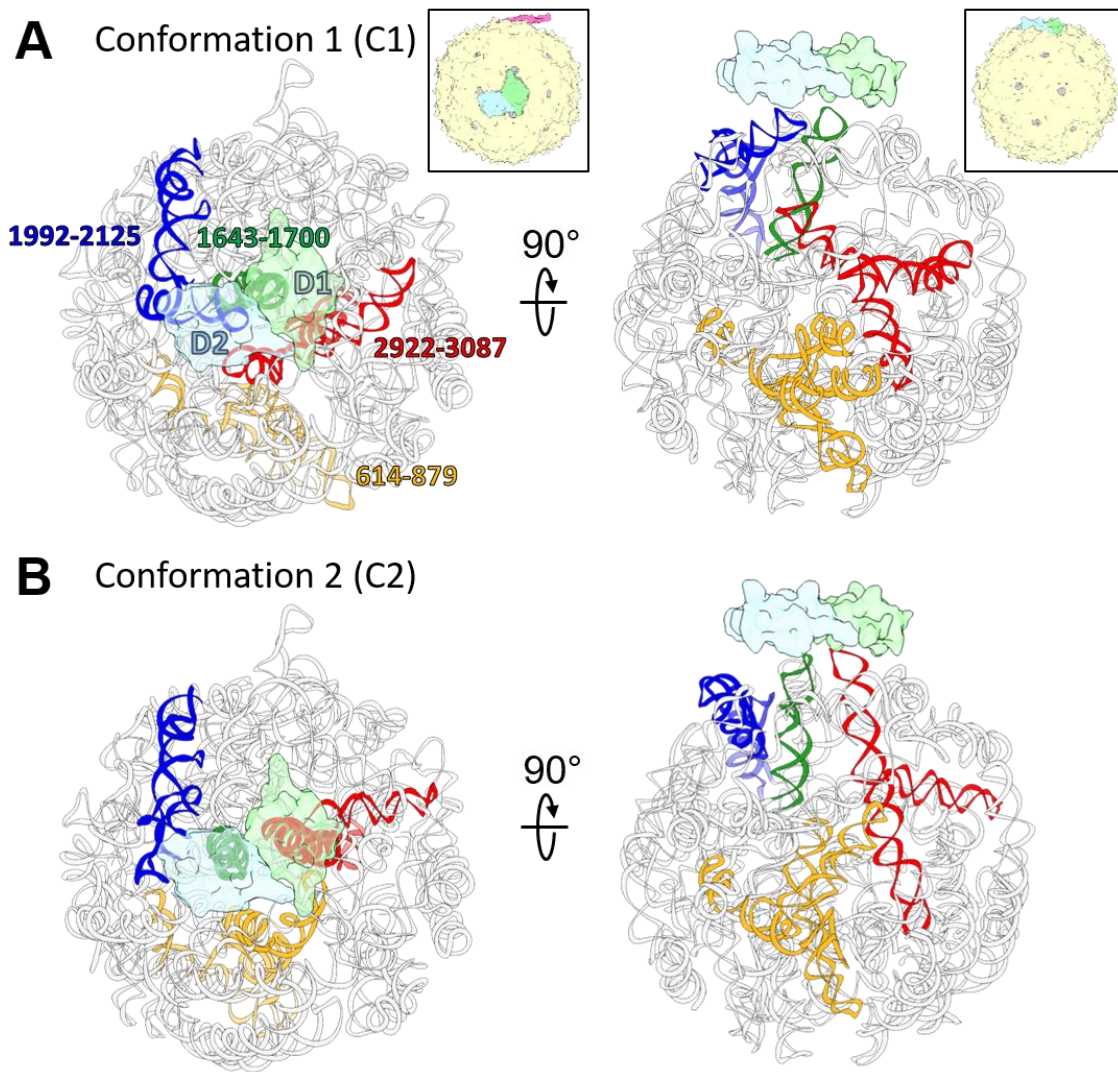


Figure II-9. Two conformations of the gRNA inside MS2

(A) and (B) Two dominant gRNA conformations of the MS2 (C1 and C2) are modeled based on the two cryo-EM maps. The gRNA fragments with large differences between the two models are colored yellow (gRNA fragment 614-879), green (gRNA fragment 1643-1700), blue (gRNA fragment 1992-2125), and red (gRNA fragment 2922-3087). Two copies of the coat protein dimers, which switch interacting stem-loop partners, are outlined by transparent green (Dimer 1: D1) and transparent blue (Dimer 2: D2). The insets of the capsids (magenta: maturation protein, green: D1, blue: D2, yellow: rest of the coat protein shell) show the viewing orientation for the corresponding columns and the location of the two dimers on the capsid.

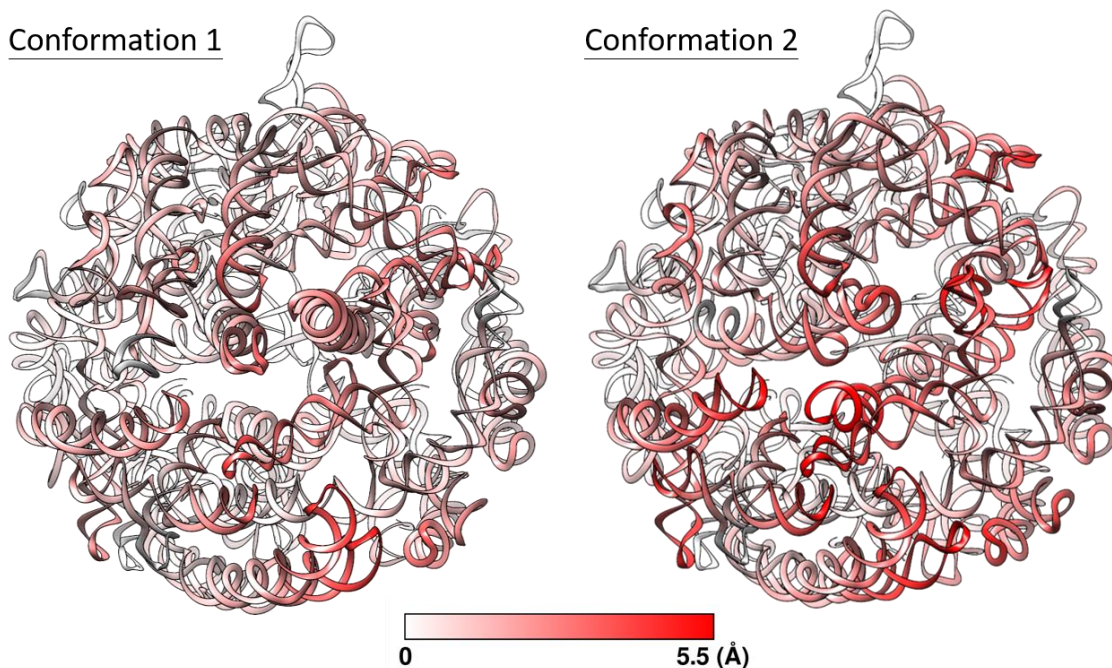


Figure II-10. Dynamics of MS2 gRNA at each nucleotide in the two conformations
 For each conformation, the average C4' RMSD of each nucleotide is calculated from 20 independently HNMMC refined models and labeled by colors from white (small RMSD) to red (large RMSD).

2.3 Discussion

In this study, we developed the molecular refinement tool, HNMMC, which combines MC simulation and customized hierarchical natural moves to efficiently refine RNA molecules into cryo-EM maps. Our energy term combines both EM energy and molecular energy to improve the fitting to the cryo-EM map while maintaining the proper stereochemistry of the model. With the simulated density map, HNMMC successfully refined a deformed tRNA model to the correct conformation. For RNA molecules with a large conformational change, HNMMC can perform large-scale conformational sampling and then locally refine the selected candidates to the target map. Following this strategy, the structure of the folding intermediate of the

ribonuclease P S-domain is automatically built from the structure of the native state. Moreover, HNMMC can be easily applied to model large RNA systems, such as ribosomes or gRNA of a virus. The customized hierarchical natural moves preserve the relatively rigid regions and allow more efficient traversing through the conformational-searching space to focus on major movements inside large molecules. For instance, two different conformations of the entire gRNA inside MS2 are built through fragments and refined into their cryo-EM maps using HNMMC.

The first step of running HNMMC is to define natural moves, which represent the potential motions of a molecule. For RNA, the natural moves can be defined using the information of the secondary structures or tertiary interactions (base-pairs, motifs, etc.) from the existing high-resolution structure or computational predictions. However, in some RNA, mispredictions of the secondary structures could occur (Lorenz et al., 2011), leading to inaccuracy in the defined DOFs. In HNMMC refinement, such wrongly defined DOFs can easily cause unrealistic geometries in the RNA tertiary structures and/or low cross-correlation between the density map and the refined model. Therefore, a bad refined model is a good indicator of potential errors in the secondary structures. On the other hand, the collective motion of residues within one natural move restrains these residues around the initial geometry. In that case, when an EM density is noisy, incomplete, or discontinuous, a restraint will not only keep the correct geometry of the grouped residues but also help anchor the entire region into the density without being biased by the artifacts in the density. For a map at a low resolution, the HNMMC

refinement may provide multiple final models with similar conformations. To select the best-refined model, the cross-correlation score between the model and the map is helpful, but a final visual inspection is still important to identify if the selected model has wrong placements of helices or minor conformational errors that are inconsistent with experimental evidence if available.

Finally, combining the MC optimization and the hierarchical natural moves reduces the computational time required for modeling. Compared to molecular dynamics-based flexible fitting (MDFF) (McGreevy et al., 2016), for example, HNMMC requires only 0.5 hours of CPU time as compared to 1.5 hours in MDFF to refine the model of the *Mtb* ribosomal handle into its EM map (Supplementary Fig. 6). Admittedly, HNMMC requires more memory (400 MB vs. 50 MB) and the speed of MDFF can be accelerated by GPU, which has not been utilized in HNMMC.

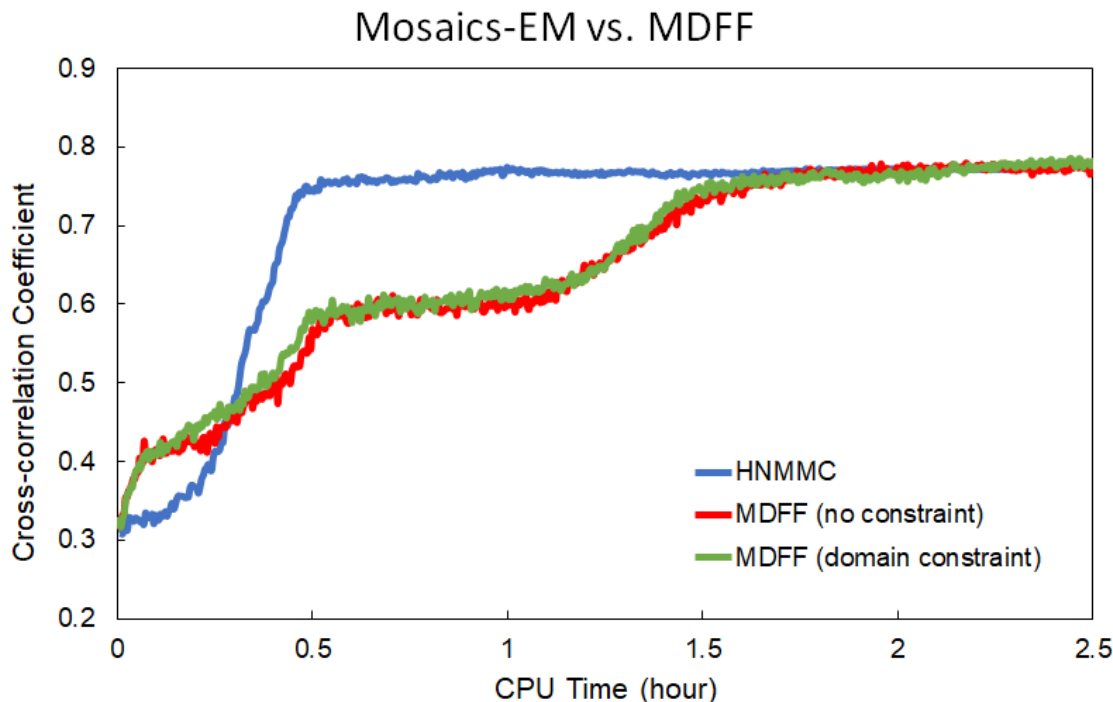


Figure II-11. The comparison between the model refinements in HNMMC and MDFF for the handle of the *Mtb* ribosome.

The domain restraint in the MDFF refinement was applied based on the secondary structural information used in HNMMC. The CPU time is calculated by the time usage on a single core.

2.4 Materials and methods

2.4.1 Multiscale Natural Move Representation of Macromolecules

The goal of the HNMMC refinement is to change the conformation of the starting model to maximize its fit against the target EM density. To exhaustively search the possible conformations of an RNA, the set of natural moves for the MC simulation can be constructed as the following: (1) Group nucleic acid residues into several segments of collective motions (natural moves) hierarchically based on their rigid or semi-rigid natural organizations, such as base-paired helical strands and specific motifs.

(2) Define the rotational and translational DOF of each segment. (3) Maintain the continuity between each segment by the recursive stochastic chain-closure algorithm (Minary & Levitt, 2010), ensuring the proper geometry of the connections and sufficient freedom in the flexible regions. (4) Move base-pairs and individual nucleotides in each segment. Moreover, natural moves can be combined into different levels to more efficiently represent global (fewer DOFs) and local movements (more DOFs) of an RNA, respectively. A common strategy is to use fewer DOFs at the beginning of the refinement to capture larger-scale movements of RNA domains and then switch to more DOFs at the later stage of the refinement to model subtle fluctuations within an RNA.

2.4.2 Temperature-modulated MC Simulation

For the energy function of the MC simulation, besides the EM energy (E_{EM}) defined by the negative cross-correlation between the target EM density and the density converted from the model, here, we also introduce molecular energy (E_{mol}) to prevent collisions within the model and improbable conformations. E_{mol} can be calculated by an empirical physics-based force field, Amber99-bs0 (parmbsc0) (Perez et al., 2007). Therefore, during each step of the MC simulation, the generated conformation is evaluated by a total energy function (E_{total}), which combines E_{EM} and E_{mol} . The total energy function is derived as $E_{total} = E_{mol} + \omega E_{EM}$. A weight (ω) at the EM energy is used to control the contribution arising from the experimental EM data. All the HNMMC refinements were performed with the simulated annealing optimization. A modulated temperature profile was applied to facilitate the exploration of different

conformations and help the structure efficiently escape from local energy minima (Figure II-1D). The protocol is implemented in the software package, MOSAICS-EM (J. Zhang et al., 2012).

2.4.3 Cryo-EM Data Collection for MS2

The RNA virus, MS2 particles were imaged under an FEI Tecnai F20 cryo-electron microscope with a field emission gun (Thermo Fisher Scientific) operated at 200 kV. Data was collected on a K2 Summit direct detection camera (Gatan) in the super-resolution mode. A nominal magnification of x29,000 was used, yielding a sub-pixel size of 0.625 Å. A 33-frame movie stack was recorded for each exposure, with a dose rate of 8 e⁻/pixel/s on the camera and 0.2 s per frame for a total exposure time of 6.6 s.

2.4.4 Cryo-EM Data Processing for MS2

The collected super-resolution movie stacks were processed as previously described (Meng et al., 2019). First, the movie stacks were binned by 2, then aligned and summed using Unblur (Grant & Grigorieff, 2015) with electron dose filtering. The defocus value of each summed micrograph was estimated using CTFFIND4 (Mindell & Grigorieff, 2003). The particles were picked semi-automatically by *e2boxer.py* in EMAN2 (G. Tang et al., 2007a). The particles were then downsampled by a factor of 2, yielding a pixel size of 2.5 Å, and refined by RELION (Scheres, 2012). To analyze the structural difference of the gRNA inside the capsid, focused classification in RELION

was used with a soft-edged mask around the flexible region of the gRNA. The orientation parameters for the 3D classification were directly taken from the consensus refinement. Four classes were requested in the 3D classification and yielded four maps representing 22.92%, 31.69%, 23.11%, and 22.28% of the total particles, respectively. Visual inspection grouped Classes 1, 4 together (45.2%) and Classes 2, 3 together (54.8%), due to the similar gRNA conformations, respectively, leading to two dominant gRNA conformations in MS2 virions. After refinement, we obtained two cryo-EM maps of MS2 with different gRNA conformations, both at 7.4-Å resolution.

2.5 Data availability

The MOSAICS-EM package and examples can be obtained from <https://cryoem.tamu.edu/software/>. The cryo-EM maps of MS2 are deposited into EM-Databank with accession ID: EMD-21659 (gRNA conformation 1) and EMD-21660 (gRNA conformation 2).

CHAPTER III
STRUCTURAL ASSEMBLY OF Q β VIRION AND ITS DIVERSE FORMS OF
VIRUS-LIKE PARTICLES

3.1 Introduction

Single-stranded RNA bacteriophages (ssRNA phages) infect a variety of Gram-negative bacteria (Dai et al., 2017; Gorzelnik et al., 2016; Klovin, Overbeek, van den Worm, Ackermann, & van Duin, 2002; K. Tars, Fridborg, Bundule, & Liljas, 2000; van Duin & Tsareva, 2006). The positive-sense RNA genome of ssRNA phages, from the 5' to the 3' end, encodes the maturation protein (Mat), the coat protein (CP), and the β -subunit of the RNA-dependent RNA replicase (Rep), with the lysis genes in variable locations of the genome for different ssRNA phages, often overlapping with other genes (Rumnieks & Tars, 2012). The genomic RNA (gRNA) is around 4,000 nucleotides in length and encapsidated in a near-icosahedral shell of CPs whose quasi-symmetry ($T=3$) is disrupted by the single Mat (Dai et al., 2017; Gorzelnik et al., 2016; Koning et al., 2016a). The Mat, embedded in the capsid, binds to both the gRNA and the host pili, which were recently resolved to near-atomic resolution for one model ssRNA phage, MS2, by single-particle cryo-electron microscopy (cryo-EM) (Meng et al., 2019).

Assembly of ssRNA phages is thought to initiate with the CPs binding to high-affinity stem-loops, either at the “operator”, a stem-loop at the start of the replicase for each phage, or other “packaging signals” spread throughout the genome (Borodavka et

al., 2012). Like many other viruses, the CPs of ssRNA phages can also assemble into virus-like particles (VLPs), which are not infectious (Zeltins, 2013). Many of these VLPs even adopt alternative particle sizes and shapes compared to the canonical $T=3$ icosahedron. For example, Q β has been reported to assemble into rod-like particles due to mutations of the FG loop in the CP (Cielens et al., 2000). Another mutation in the CP of MS2 can lead to $T=1$ icosahedrons (Asensio et al., 2016). Recently, it has also been reported overexpressing a CP dimer (two CPs tandemly linked) on a plasmid can lead to $T=4$ icosahedrons or mixed capsids in MS2 and PP7 (de Martin Garrido et al., 2020; Zhao, Kopylov, Potter, Carragher, & Finn, 2019); but almost all of these alternative sizes and shapes of VLPs were observed when the CP has been modified. Recently, it was discovered that overexpressing the CPs of ssRNA phages will also produce VLPs with $T=1$, $T=4$ or prolate symmetry (Lieknina et al., 2019; Kaspars Tars, 2020). However, it has not been observed that such a variety of VLPs can be produced from a normal infection by a wild-type ssRNA phage. The availability of different sizes and forms of VLPs is important for nanotechnological and biomedical applications, such as for batteries (Lee et al., 2009; Nam et al., 2006), vaccines (Bao et al., 2019; Frieze, Peabody, & Chackerian, 2016), and drug delivery (Karimi et al., 2016; Rohovie, Nagasawa, & Swartz, 2017). As a matter of fact, a basic understanding of the assembly mechanism of ssRNA phages is invaluable for expanding the applications of VLPs.

The canonical ssRNA phage Q β has a gRNA of 4,217 nucleotides with the conserved *mat-cp-rep* gene architecture (van Duin & Tsareva, 2006). The Mat of Q β

also serves as the lysis protein (Karnik & Billeter, 1983). The *cp* gene has a leaky stop codon that occasionally reads through to produce a minor CP (also known as A₁) that is required for infection (Hofstetter, Monstein, & Weissmann, 1974). The gRNA of Q β has been shown to have a defined 3D conformation inside the capsid (Figure III-1) (Cui et al., 2017; Gorzelnik et al., 2016). Surprisingly, inside the Q β capsid, there is a sequestered dimer of CPs interacting with an RNA stem-loop originated from a 5-way junction domain, which also presents a stem-loop to interact with the Mat (Cui et al., 2017). Such an RNA domain may form a local nucleation site for the capsid to condense around. However, the RNA sequence of this 5-way junction domain has not yet been identified. Identifying this sequence is essential for linking foreign RNA to the Mat to be delivered into host cells. Another observation is that only ~25% of the purified Q β particles show a defined gRNA conformation and have the Mat in the capsid. The other ~75% particles show no structured gRNA and were thought to be VLPs, as they lack the Mat and have perfect $T=3$ icosahedral symmetry (Cui et al., 2017). This is consistent with previous findings that only a fraction of particles are infectious (Leipold & Hofschneider, 1975; Watanabe & Watanabe, 1970).

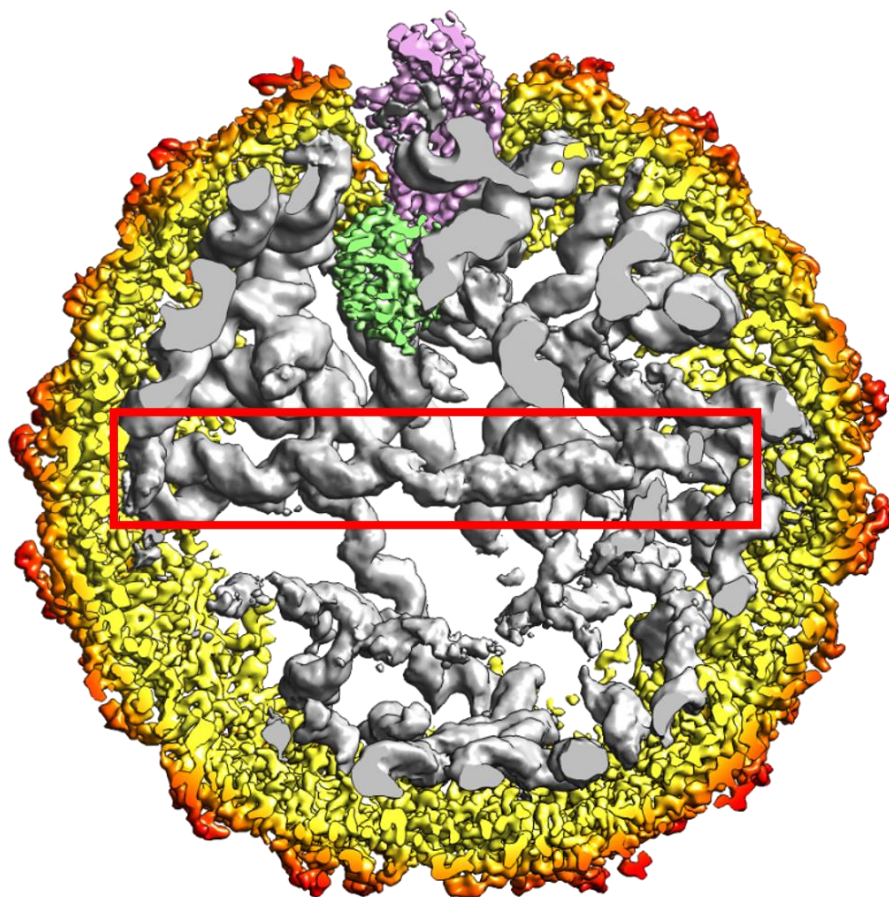


Figure III-1. Cut-through view of the asymmetry cryo-EM map of Q β
The capsid, Mat, gRNA, and internal CP dimer are colored yellow, magenta, grey, and green, respectively. The red box indicates the 200-Å long helical density horizontally spans the equator of the capsid.

In this chapter, we first establish the entire gRNA model of Q β using computational structural modeling based on its cryo-EM density map to not only reveal the RNA domain that interacts with the Mat but also all the RNA stem-loops that interact with the CPs for assembly. Such a model is consistent with our subsequent results from Electrophoretic Mobility Shift Assays (EMSA) and mutagenesis. We also model the long-range tertiary interactions within the gRNA, such as kissing loops, that may help

the RNA to compact during viral assembly. Using a new method of phage purification by PEG precipitation and gel-filtration chromatography, we have optimized a protocol to purify non-infectious particles and surprisingly, for the first time identified a variety of alternative forms of the VLPs directly from the wild-type phage infection. This is in contrast to the previously reported alternative forms purified after modifying the CP. In addition, we have observed forms that have not been observed before, such as the oblate and the D3 symmetrical small prolate forms, which show the versatility of the CP of Q β .

3.2 Results

3.2.1 The complete gRNA model inside the Q β virion

From our cryo-EM density map, we have built the complete atomic model of the Q β virion, including the entire 4,217 nucleotides gRNA in its dominant conformation in the capsid. The gRNA folds domain-by-domain with the 3' domain (the red ribbon model in Figure III-2) packaged close to the Mat and the 5' domain (the blue-ribbon model in Figure III-2) far away from the Mat. This orientation of the gRNA packaging is consistent with the electrophoretic mobility shift assay, in which the purified Mat has a much stronger affinity towards the 3' end of the gRNA compared to the 5' end (Figure III-3). The result that, in Q β , only the 3' end of the gRNA binds to Mat is in contrast to the previous claims that the Mat of ssRNA phages bind to both the 5' and 3' ends of the gRNA (Shiba & Suzuki, 1981), but in agreement with the recent observation in another ssRNA phage MS2 (Dai et al., 2017).

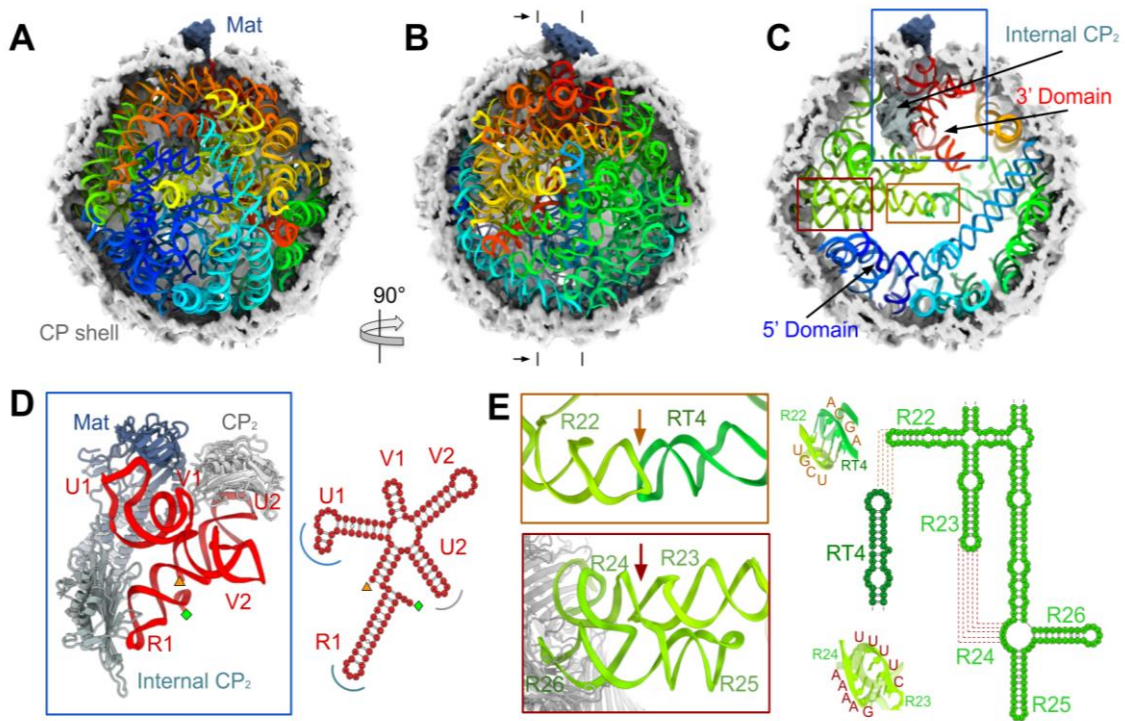


Figure III-2. The complete model of the dominant gRNA fold in Q β .

(A) Cut-open view of the Q β virion to display the gRNA, shown as ribbons and rainbow-colored from the 5' end (blue) to the 3' end (red). The CP shell and Mat are represented as density, colored light grey, and dark blue, respectively. (B) The view vertically rotated 90° with half of the CP shell removed. (C) Slice view between the marks in Panel B and viewed from the black arrows. The internal CP dimer (Internal CP₂), the 5' and 3' gRNA domains are light blue, blue, and red, respectively. (D) Left: zoom-in view within the blue box in Panel C showing the gRNA 5-way junction domain (red) interacting with the Mat (dark blue), an exposed CP dimer (CP₂, grey), and the internal CP dimer (Internal CP₂, light blue). The 5' and 3' ends of the 5-way junction domain are labeled as green diamond and orange triangle, respectively. Right: Secondary structure of the gRNA 5-way junction domain. The dark blue, grey, and light blue arcs denote the interactions with the Mat, exposed, and internal CP dimers, respectively. (E) Left: zoom-in views within the orange and red boxes in Panel C showing the two kissing loops that link helices R22, RT4, R23, and the bulge R24. The orange and red arrows indicate the locations of the kissing loops. Right: Models of the two kissing loops and their secondary structures showing the base pairing within the sequence.

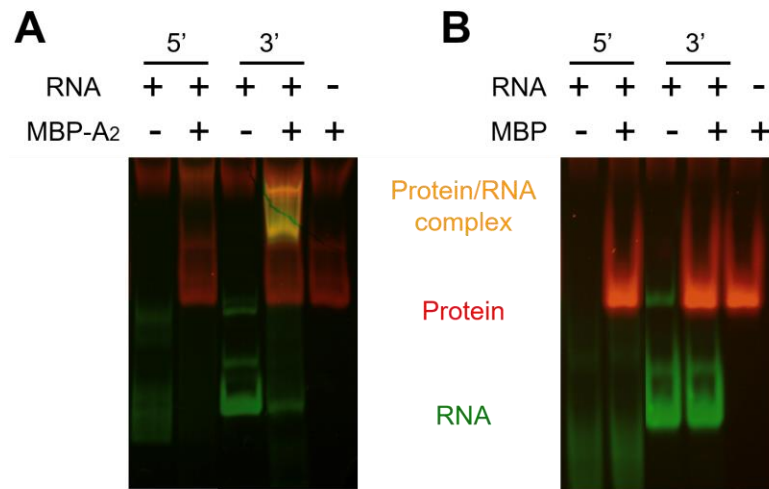


Figure III-3. The interaction of the Mat and the 3' end of Q β .

(A) Electrophoretic Mobility Shift Assay (EMSA) of the 5' and 3' UTR with the Maltose-binding protein (MBP) tagged A₂. The gRNA fragments of the 5' UTR (residues 1-60) and the 3' UTR (residues 4126-4217) were stained with SYBRTM Green (green). The protein was stained with SYPROTM Ruby (red). The co-localized band (yellow) of the RNA and protein at the lane of 3' represents the interaction of the 3' UTR and the Mat (B) EMSA of the 5' and 3' UTR with the MBP tag shows no interaction between the RNA and the MBP.

The RNA 5-way junction domain, which interacts with the Mat, internal CP dimer, and one exposed CP dimer, is determined to be the 3' untranslated region (UTR) of the gRNA (Figure III-2D, Figure III-4). The last stem-loop, U1, in the entire gRNA, specifically interacts with the Mat (Figure III-4A). The stem-loop U2 interacts with one exposed CP dimer. Unique to Q β is that the last stem-loop of the replicase gene, R1, specifically interacts with the internal CP dimer. The last six nucleotides of the entire gRNA, including the CCA tail, fold back to be part of stem-loop R1 by base-pairing with residues 4124-4126. Such a compact RNA 5-way junction domain has a biotechnological implication to serve as an adapter to the Mat for an RNA of interest to be delivered into the host cell. Given the fact that most ssRNA phages have a conserved

motif for the 3' UTR of the gRNA (Klovins et al., 2002), it is reasonable to think such a linkage between the 3' UTR and the Mat is conserved in ssRNA phages, leading to a common mechanism for gRNA packaging and delivery. The specific RNA structures within the 5-way junction domain are required for Q β infectivity, as mutations within either the U1 or R1 stem-loops don't affect the production of phage-like particles but abolish infectivity (Figure III-5), primarily due to their requirements for the incorporation of the Mat into the capsid. Consistently, Western blots of the Mat shows that, for these mutants, the Mat is still expressed in the host but failed to be incorporated into the phage capsids. (Figure III-5).

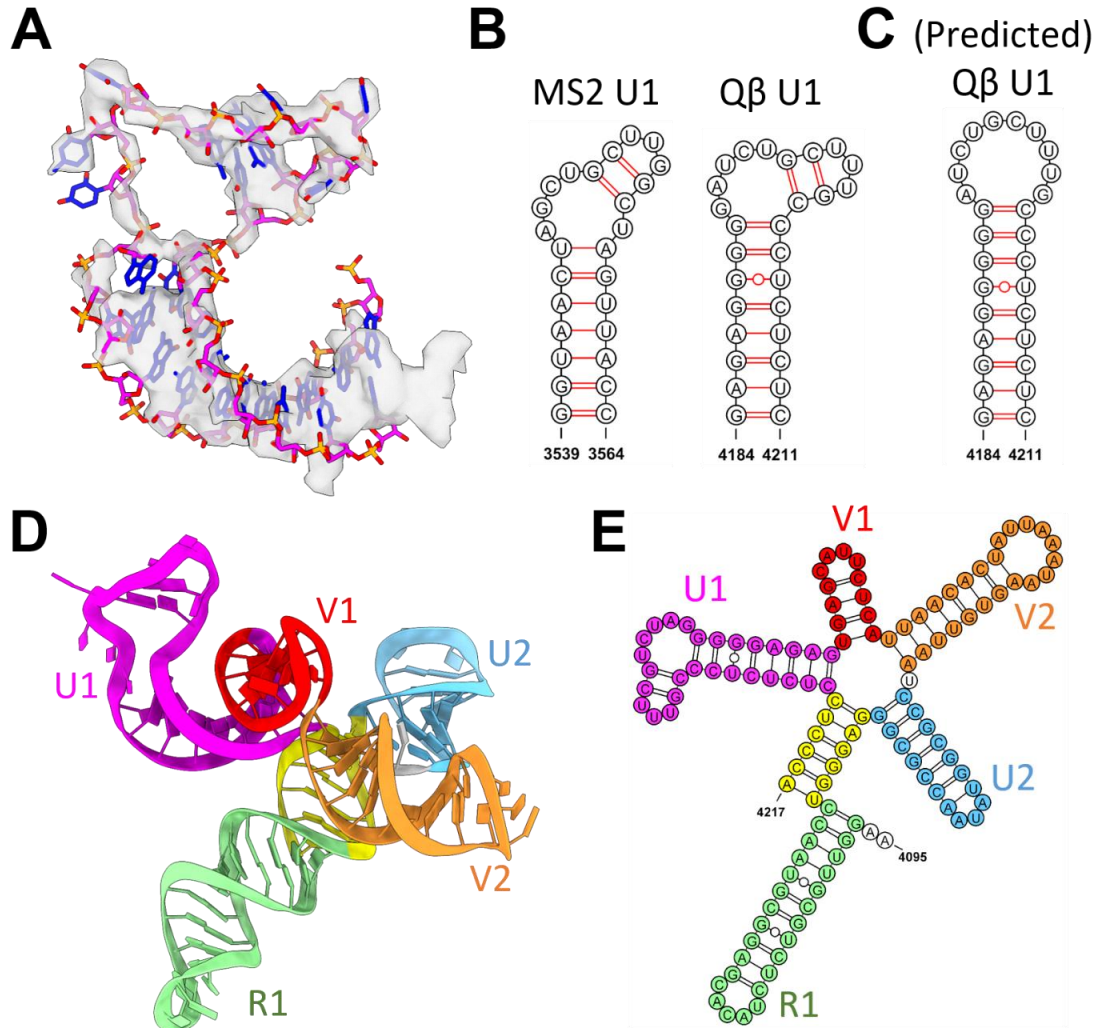


Figure III-4. The 3' UTR of the Q β gRNA

(A) The model of the U1 in the Q β 3' UTR with its cryo-EM density. (B) The secondary structures of the U1 in the MS2 and Q β 3' UTR are observed in the experimental data. (C) The theoretical predicted secondary structure of the U1 in the Q β 3' UTR. (D) The model of the 3' UTR in the Q β . The sequences of R1, U2, V2, U1, and the connecting helix are colored green, blue, orange, red, magenta, and yellow, respectively. (E) The secondary structure of the 3' UTR in the Q β . Each helical stem is colored as Panel D.

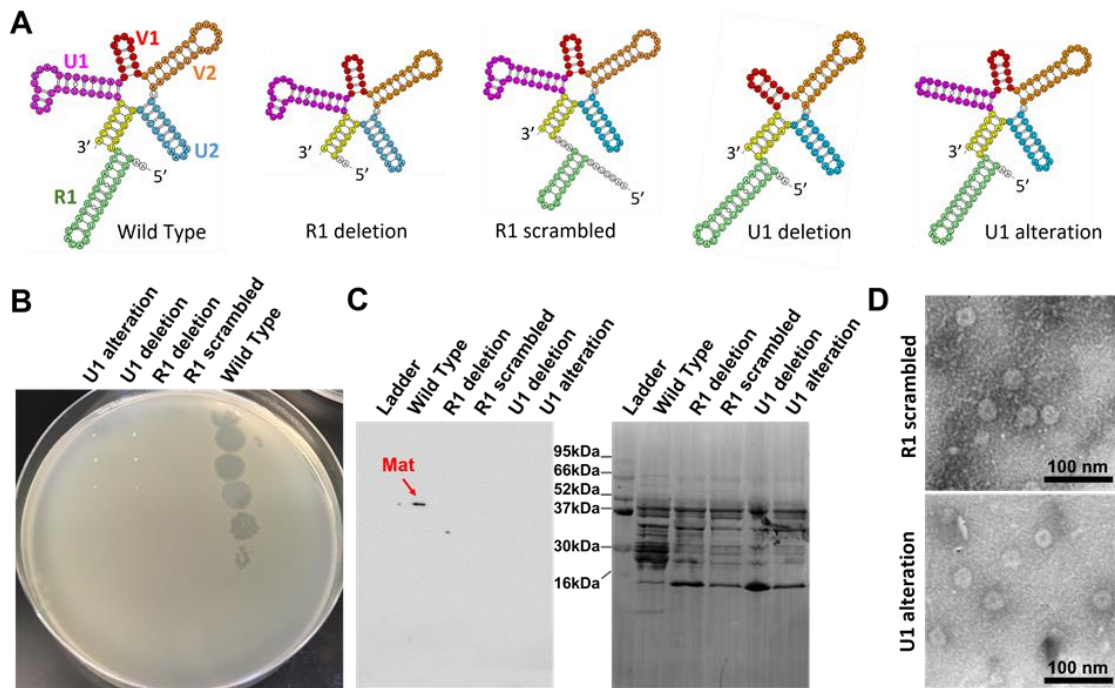


Figure III-5. Specific RNA stem-loops at the 3' end of the gRNA required for infectious particle formation

(A) Based on the 3' end of the wild type gRNA, four mutants were engineered on a cDNA plasmid to validate the model of the gRNA. The R1 loop, which binds the internal CP dimer, was either deleted (R1 deletion, second from left in Panel A) or re-engineered to maintain coding capacity but with an alternative RNA structure (R1 scrambled, middle in Panel A). For U1, which does not code for any proteins and binds to the Mat, two mutations were engineered, such that the entire stem-loop was deleted or the loop was altered (U1 deletion and U1 alteration, second from the right and on the right, respectively). (B) Titers from overnight cultures of bacteria show that the mutants are all non-infectious, while *E. coli* with the wild-type plasmid produces infectious particles. (C) A Western Blot using an antibody against Mat (A₂) shows that CsCl purified particles produced from these mutants do not incorporate the A₂. Wild-type and mutants were produced from basal level expression off of a plasmid in female *E. coli* (F⁻), to prevent normal infectious replication. The left panel is the Western Blot of the mutants and wild-type, while the right panel is the same membrane, stained after the blot to show total protein. Basal level expression produces low amounts of phage with host proteins seen in the stained membrane. (D) Two of the mutants, R1 scrambled and U1 alteration, were purified and the negative-stain EM showed that particles were formed.

One striking feature of the gRNA organization in Q β is a \sim 200-Å long helical structure that horizontally spans the equator of the capsid when the Mat is defined as the north pole (Figure III-1) (Gorzelnik et al., 2016). In the complete model of the gRNA, it was determined that such a long helical structure is in fact formed by shorter RNA stem-loops through two kissing loops (indicated by orange and red boxes in Figure III-2C, zoomed-in Figure III-2E). One kissing loop is formed between residues “2844-UCGU-2847” of stem-loop R22 in the rep gene and residues “1876-ACGA-1879” of stem-loop RT4 in the gene for the read-through domain (Weiner & Weber, 1971). The other is a branched kissing loop (D. Liu et al., 2020) formed between residues “2811-UUUUC-2815” of stem-loop R23 and residues “2749-GAAAA-2753” of bulge R24 in the rep gene. Sequence alignment shows the nucleotides supporting the formation of these two kissing loops are conserved in Q β -like phages (Figure III-6). In addition, the bulge R24 encodes the amino acid residue Arg133 and Lys134 in the β -subunit of Q β replicase. These two amino acids are important for Q β replicase to recognize the gRNA template during the initiation of the replication (Gytz et al., 2015). The sequence conservation for these amino acids has been coupled into the structural morphology of the virus particles, as mutations in bulge R24 may disrupt the formation of the kissing-loop, leading to instability of the gRNA during the packaging of the phage virion. This long-distance interaction could cause selective pressure against mutations in this region.

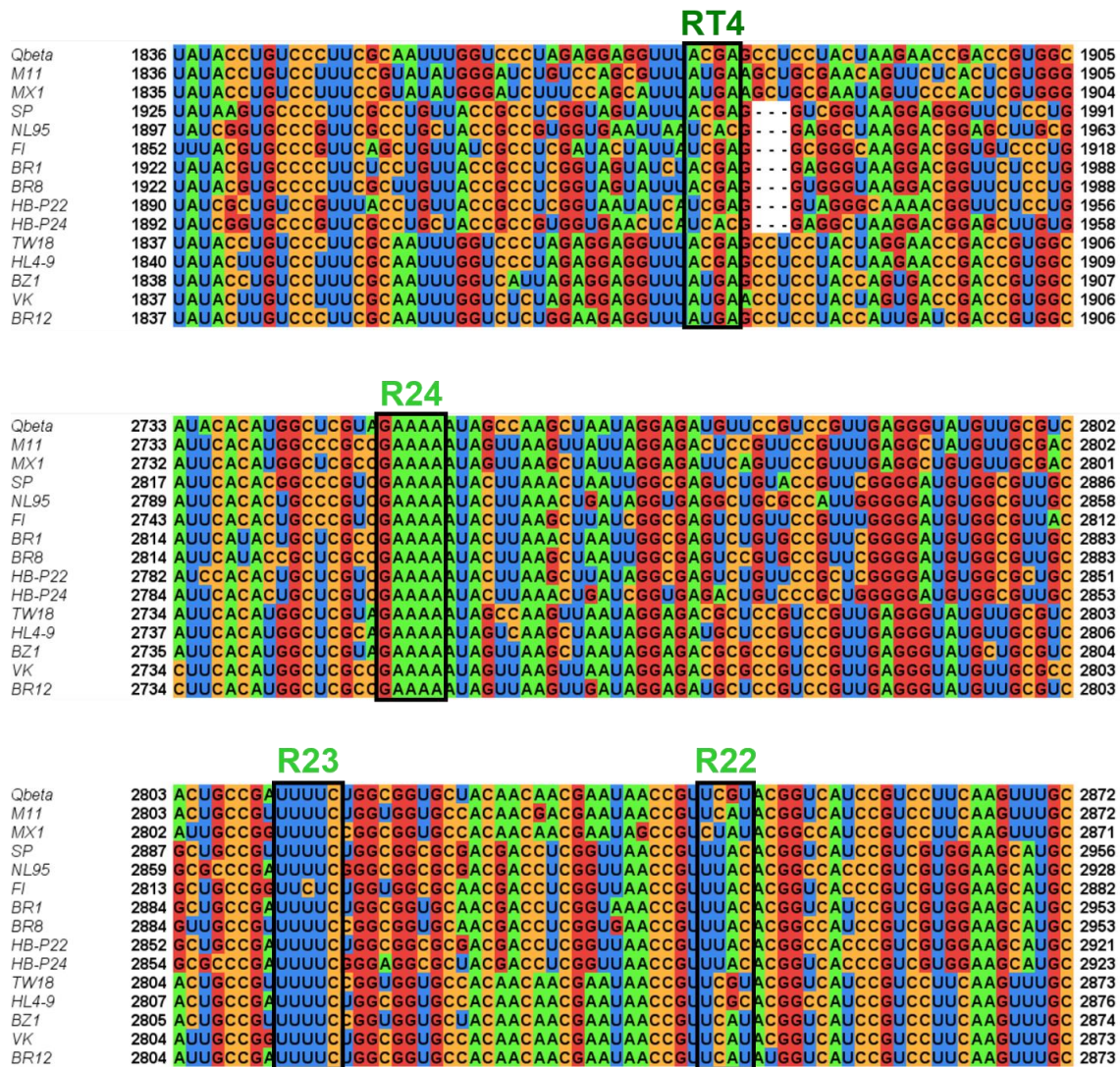


Figure III-6. Conservation of the two kissing-loops in the Qβ-like phages
 The sequences are aligned by Clustal Omega (Madeira et al., 2019) and represented by Jalview (Waterhouse, Procter, Martin, Clamp, & Barton, 2009). R22, R23, and R24 are at the *replicase* gene. RT4 is at the *read-through* gene.

3.2.2 Operator-like RNA stem-loops dominantly support the pentamers of the CPs

The capsid for an infectious ssRNA phage virion has a near icosahedral capsid with a triangulation number of 3 ($T=3$), which consists of twelve pentamers and twenty hexamers of the CPs (Figure III-7A). One CP dimer, at the two-fold axis on the shared

edge of two hexamers, is replaced by the Mat, resulting in a total of 89 CP dimers on the capsid. For Q β , an extra internal CP dimer interacts with stem-loop R1 at the 3' end of gRNA, still making a total of 90 CP dimers. Such an icosahedron can be unwrapped onto a hexagonal lattice on each thick edge sits a CP dimer (Figure III-7B). Out of the 77 stem-loops within the Q β gRNA, 59 of them have a distance of less than 5Å to the protein component, including the CP shell, the Mat, and the internal CP dimer. 33 of these stem-loops (termed operator-like stem-loops) interact with CP dimers in a fashion similar to the actual translational operator, wherein the backbones of the RNA stem-loops follow a similar trace relative to the CP-dimer compared to the actual operator (Table III-1). They are indicated as circles on the hexagonal lattice (Figure III-7C). The locations of these operator-like stem-loops are throughout the entire gRNA sequence (labeled in bold black font in Figure III-7D) and are proposed to be directly involved in the assembly of the viral capsid (Stockley et al., 2013). Interestingly, some of the stem-loops interact with the CP dimers in a different fashion from the operator-like RNA stem-loop (with the backbones of these stem-loops folding in an alternative manner) and are classified as two types. One type of the stem-loops points toward the center of the CP dimers (termed CP-sandwiched stem-loops, Table III-2), and the other type is anchored on just one of the CP in a CP dimer (referred to as CP-anchored stem-loops, Table III-3). While the CP-sandwiched stem-loops may have relatively strong affinities to the CPs and recruit CP dimers to facilitate the capsid assembly, the CP-anchored stem-loops may have lower affinities to the CPs but can still contribute to the stability of the viral capsid (Figure III-8). The locations of these CP-sandwiched and CP-anchored stem-loops in the

entire gRNA sequence are labeled in normal black and grey, respectively in Figure III-7D.

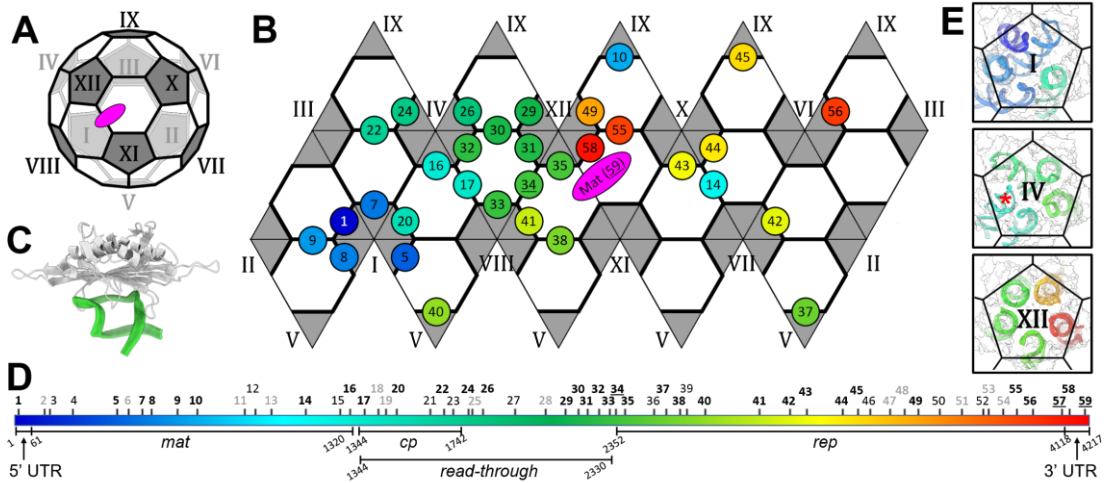


Figure III-7. Operator-like RNA stem-loops in the gRNA that interact with the capsid shell

(A) The near icosahedral cage of Q β with the Mat labeled as a magenta oval. The twelve pentamers are dark grey and labeled from I to XII. (B) The cage is unwrapped to display the capsid as a 2D map, with each thick line representing a CP dimer. The asymmetric A/B dimer is between a grey pentamer and a white hexamer. The symmetric C/C dimer is between two white hexamers. The rainbow-colored circles label the operator-like stem-loops from the 5' (blue) to 3' (red) ends of the gRNA. (C) A model showing an operator-like RNA stem-loop interacting with a CP dimer. (D) The sequence of Q β gRNA is rainbow-colored. The UTRs, the genes of *mat*, *cp*, *rep*, and the *read-through* are labeled. The 59 stem-loops, which have a distance of less than 5Å to the capsid proteins are numbered and with the positions labeled on the sequence with operator-like, CP-sandwiched, and CP-anchored stem-loops labeled bold black, black, and grey, respectively. The actual translational operator and the stem-loop that interacts with the internal CP dimer are labeled by the underlined numbers 34 and 57, respectively. The underlined 59 labels the stem-loop that interacts with the Mat. (E) The operator-like RNA stem-loops (rainbow-colored) interact with the CPs in Pentamers I, IV, and XII. The red star labels one CP-sandwiched RNA stem-loop (number 15 in Pentamer IV).

Table III-1. Operator-like RNA stem-loops

Number 34 is the operator. Number 57 is the stem-loop interacting with the internal CP dimer. Each stem-loop-CP-dimer complex is shown in two orthogonal views (left and right).














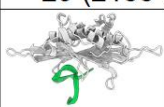

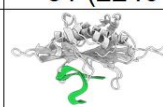


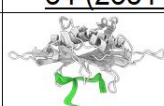

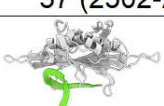



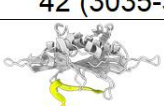
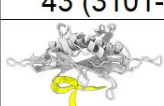







1 (6-19)	5 (393-406)	7 (492-505)	8 (528-541)
			
9 (628-640)	10 (710-722)	14 (1134-1147)	16 (1308-1320)
			
17 (1363-1375)	20 (1490-1502)	22 (1677-1689)	24 (1773-1784)
			
26 (1829-1842)	29 (2155-2167)	30 (2206-2218)	31 (2240-2252)
			
32 (2287-2299)	33 (2329-2341)	34 (2351-2364)	35 (2386-2399)
			
37 (2502-2516)	38 (2603-2616)	40 (2702-2714)	41 (2915-2927)
			
42 (3035-3047)	43 (3101-3114)	44 (3242-3255)	45 (3303-3316)
			
49 (3539-3551)	55 (3926-3938)	56 (3710-3726)	57 (4104-4416)
			
58 (4132-4144)			
			

Table III-2. CP-sandwiched RNA stem-loops.

They have cross-correlation scores between 0.30 and 0.55 to the density of the operator. Each stem-loop-CP-dimer complex is shown in two orthogonal views (left and right).

3 (128-140)		4 (220-232)		12 (937-949)		15 (1267-1280)	
21 (1622-1634)		23 (1711-1723)		27 (1954-1966)		36 (2503-2515)	
39 (2629-2641)		46 (3349-3362)		50 (3623-3636)		52 (3794-3807)	

Table III-3. CP-anchored RNA stem-loops.

They have cross-correlation scores smaller than 0.30 to the density of the operator. Each stem-loop-CP-dimer complex is shown in two orthogonal views (left and right).

2 (106-118)		6 (436-449)		11 (893-905)		13 (998-1101)	
18 (1418-1431)		19 (1447-1460)		25 (1788-1800)		28 (2706-2088)	
47 (3429-3442)		48 (3485-3498)		51 (3712-3724)		53 (3815-3828)	
54 (3873-3886)							

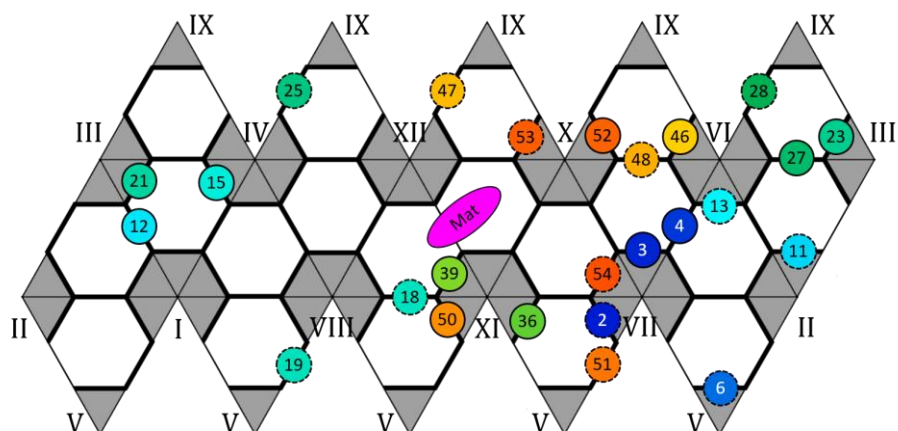


Figure III-8. CP-sandwiched and CP-anchored RNA stem-loops

Similar to Figure III-6B, the RNA stem-loops are labeled with circles on the unwrapped capsid with each thick line representing a coat protein dimer. The solid and dash circles represent the CP-sandwiched and CP-anchored stem-loops, respectively (Table III-2 and III-3). The circles are colored rainbow by their related sequence as Figure III-6D from the 5' to 3' end.

In the previous icosahedral structure of ssRNA phages MS2 and Q β , the CP dimers are classified as A/B dimers (located on the edge between a pentamer and a hexamer) and C/C dimers (located on the edge between two hexamers) (Golmohammadi, Fridborg, Bundule, Valegard, & Liljas, 1996; Golmohammadi, Valegard, Fridborg, & Liljas, 1993). Notably, most of the operator-like stem-loops interact with the A/B dimers, with only 8 out of the 33 interacting with the C/C dimer and 1 interacting with the internal CP dimer. Surprisingly, in Q β , the actual translational operator stem-loop interacts with a C/C dimer on the capsid (labeled as 34 in Figure III-7B). In MS2, it was previously reported that binding of the 19-nt long RNA fragment containing the translational operator will change the C/C dimer into A/B dimer in solution (Stockley et al., 2007). However, in the context of the entire gRNA, such a

conformational change induced by the operator-binding may also be constrained by the fold of the gRNA and the location of this stem-loop in the gRNA. Still, the operator-like RNA stem-loops preferentially bind to the pentamers on the capsid with eleven out of the twelve pentamers having at least one operator-like sequence bound. Strikingly, Pentamers I, IV, and XII have at least four constituting A/B dimers bound with an operator-like RNA stem-loop (Figure III-7E). That is almost half of the operator-like stem-loops within the whole gRNA clustered inside this local patch of the capsid. These stem-loops are mostly located within the 5' half of the gRNA. As the gRNA is synthesized from the 5' end to the 3' end, in the event of co-replicative genome packaging (Saxena & Lomonosoff, 2014), it is intuitive that a local patch of the capsid would form first on the 5' half of the gRNA.

3.2.3 Q β virions are being made along with VLP of different forms

Recently, $T=4$ capsids have been observed when overexpressing the covalently-linked CP dimers in ssRNA phages PP7 and MS2 (de Martin Garrido et al., 2020; Zhao et al., 2019). Here, we have observed under normal wild-type phage infection, $T=4$ capsids are also made, albeit a small percentage (0.5%) of the overall number of particles (Figure III-9A, Figure III-10, Table III-4). Surprisingly, we have also observed about 1.7% of the VLPs show a “walnut-like” shape (Figure III-10). A 3D reconstruction of these particles reveals that they are a prolate form with $T=3$ and $Q=4$, while T (or T_{end}) and Q (or T_{mid}) describe the triangulation number of the two caps and the elongated faces, respectively. The $T=4$ and prolate VLPs have an internal volume of around 8,750

nm^3 and $6,900 \text{ nm}^3$, respectively, compared to the $5,300 \text{ nm}^3$ of the $T=3$ VLPs. That is 65% and 30% more internal space compared to the $T=3$ capsids. More interestingly, we have observed smaller particles compared to the $T=3$ VLPs. In order to analyze the structures of these smaller particles, we improved our phage purification protocol by using gel-filtration chromatography instead of the density gradient (Figure III-11). Although this method does not increase the number of $T=4$ or prolate particles, to our surprise, almost a quarter of the VLPs are of the smaller sizes now. Further structural characterization of these smaller particles revealed two previously unreported forms of the VLPs. One is the oblate form with $T=3$ and $Q=2$; the other is a small prolate form elongated along the 3-fold axis from the two $T=1$ caps and the elongated faces consist of two triangulation numbers, $Q_1=2$ and $Q_2=3$ (Figure III-12). However, the $T=1$ icosahedral VLPs are not observed in our dataset. Figure III-8A shows the five forms of the VLPs that are purified from a normal infection of the wild-type $Q\beta$, each containing 240, 210, 180, 150, and 132 copies of the CPs for the $T=4$, prolate, $T=3$, oblate, and small prolate forms, respectively (Figure III-9B). While all of the five forms contain twelve copies of the pentamers, they consist of 30, 25, 20, 15, and twelve copies of the hexamers for the $T=4$, prolate, $T=3$, oblate, and small prolate forms, respectively. Strikingly, purified by gel-filtration chromatography, the oblate and small prolate forms are around 21% and 3%, respectively. Such a large number of smaller particles may have been left out due to their having a different density as compared to the regular $T=3$ capsids and ignored in the previous density-gradient based purification methods.

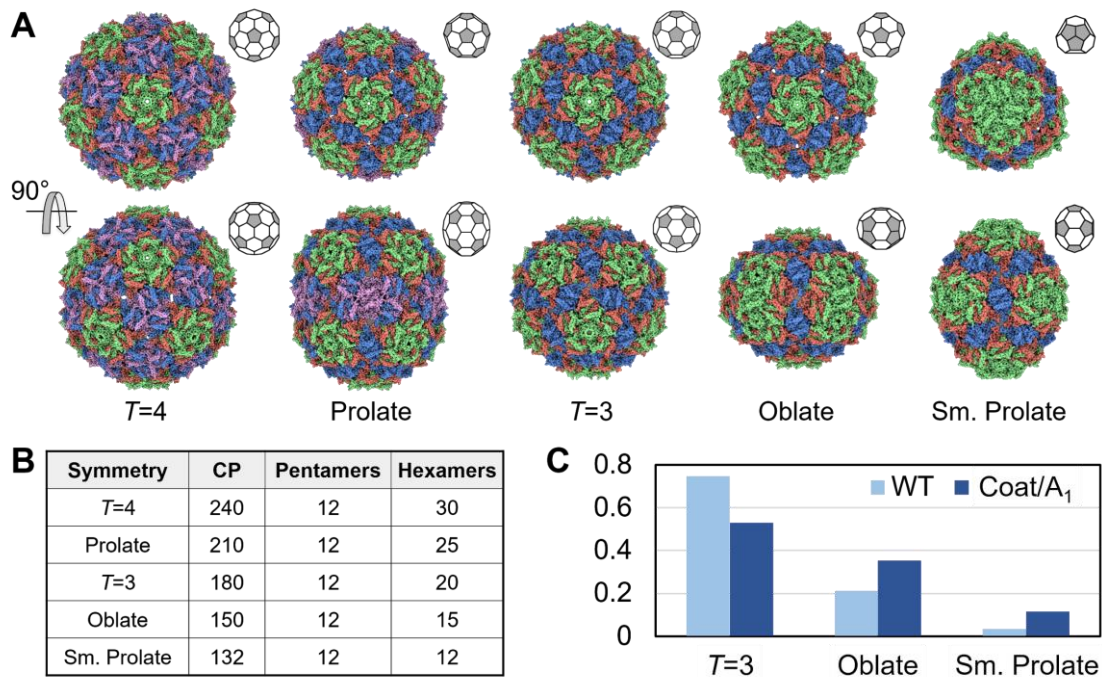


Figure III-9. Structures of the Q β VLPs

(A) Top: cryo-EM structures of the Q β VLPs of $T=4$, prolate, $T=3$, oblate, and small prolate symmetries from left to right. The corresponding symmetry cages are drawn with pentamers and hexamers colored grey and white, respectively. Bottom: the same structure horizontally rotated 90°. (B) The table shows the number of CPs, pentamers, and hexamers for each form of VLP. (C) The percentages of the $T=3$, oblate and small prolate VLPs produced by the wild-type Q β infection (light blue) and over-expressing the *cp/A₁* gene (deep blue).

Table III-4. Percentages of different Q β VLPs via different purification methods

	$T=3$	$T=4$	Prolate	Oblate	Small Prolate
Wild Type (CsCl gradient)	97.23%	0.45%	1.72%	0.60%	< 0.01%
Wild Type (gel-filtration)	74.67%	< 0.01%	0.97%	21.09%	3.27%
<i>cp/A₁</i> (gel-filtration)	52.85%	< 0.01%	0.29%	35.28%	11.58%

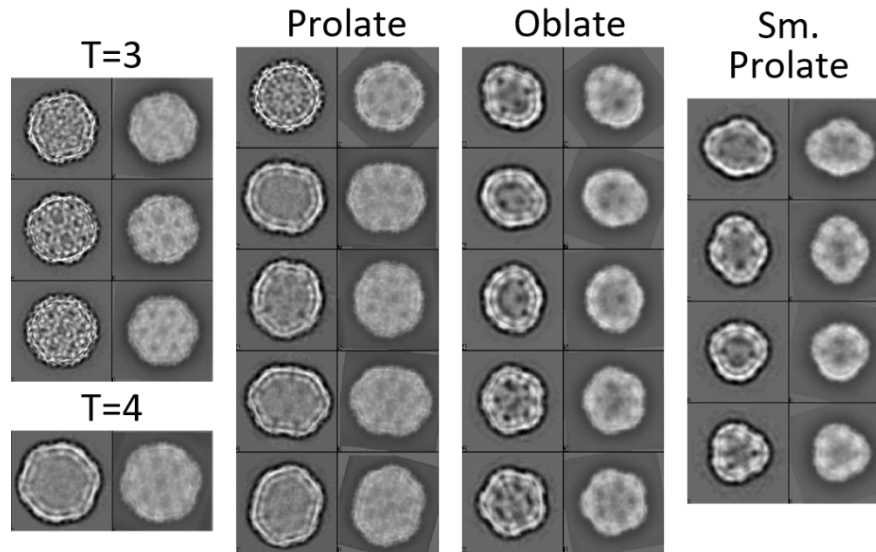


Figure III-10. The 2D class averages of Q β VLPs

Refinement results against class averages of each VLP conformation. The first column of each conformation shows the projections of the refined map with different orientations. The second column shows the 2D experimental class averages similar to the first column.

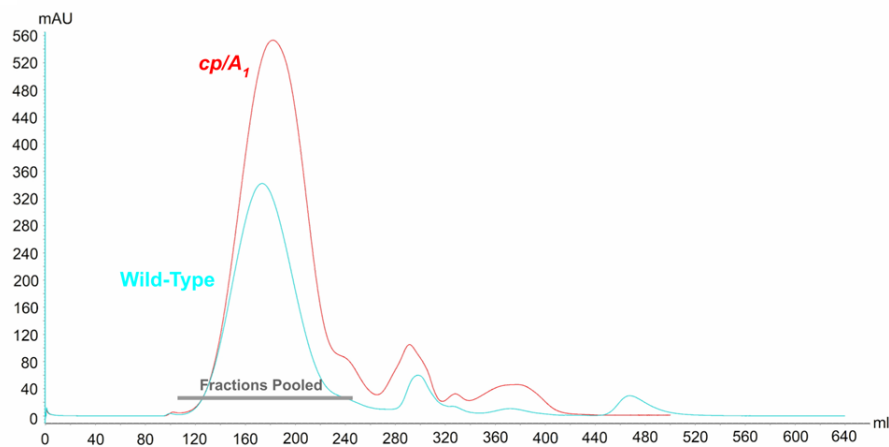


Figure III-11. The purification of ssRNA phages and VLPs using gel-filtration chromatography.

Wild-type produced phage from an infection (cyan) or the *cp/A₁* overexpressed (red) purified over an S500 column.

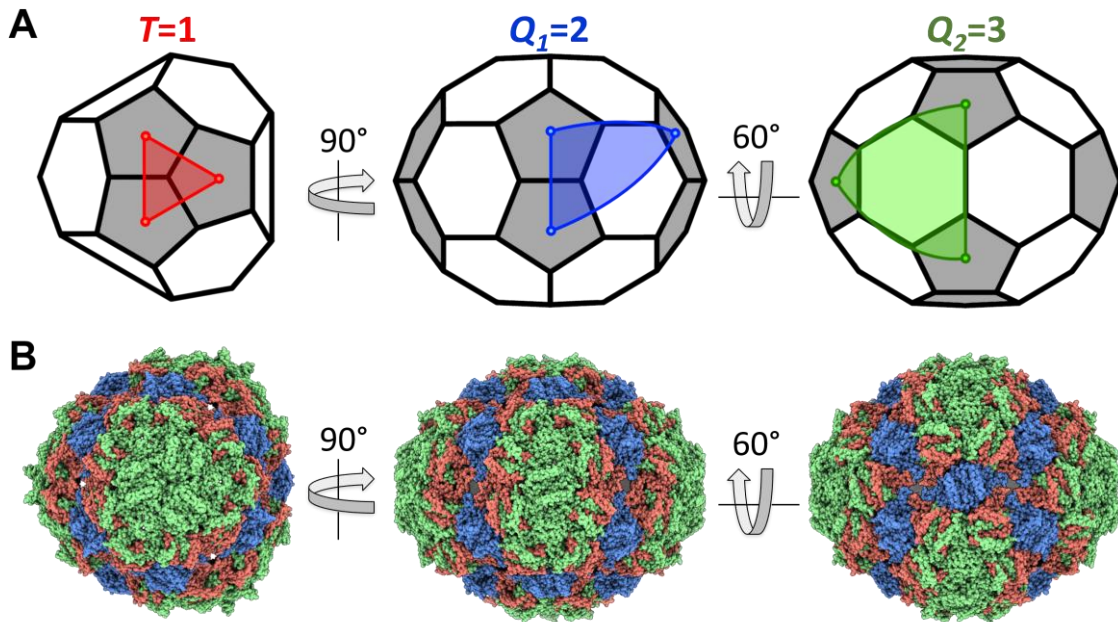


Figure III-12. The triangulation numbers of the small prolate

(A) The cage of the small prolate with the triangulation number of each face. Pentamers are colored grey. The facet triangles of $T=1$, $Q_1=2$, and $Q_2=3$ are colored red, blue, and green, respectively. (B) Views of the small-prolate model corresponding to the orientation of Panel A.

When over-expressing the gene containing only the *cp* and *A₁* (*read-through*) fraction of the Q β genome from a plasmid, the percentage of the oblate and small prolate forms increase to 35% and 12%, respectively (Figure III-9C). The large increase in the percentage of these smaller particles may be due to the presence of the large amount of the shorter RNA (expressing just the *cp* and *A₁*) inside the cell which may also be encapsulated by smaller particles of the oblate and small prolate forms. Additional densities were observed when performing asymmetric reconstructions of classified

particles with the oblate and small prolate forms, however, due to the lower resolution and potential heterogeneity of the RNA within the capsid, it is not feasible to model the RNA inside these VLPs (Figure III-13).

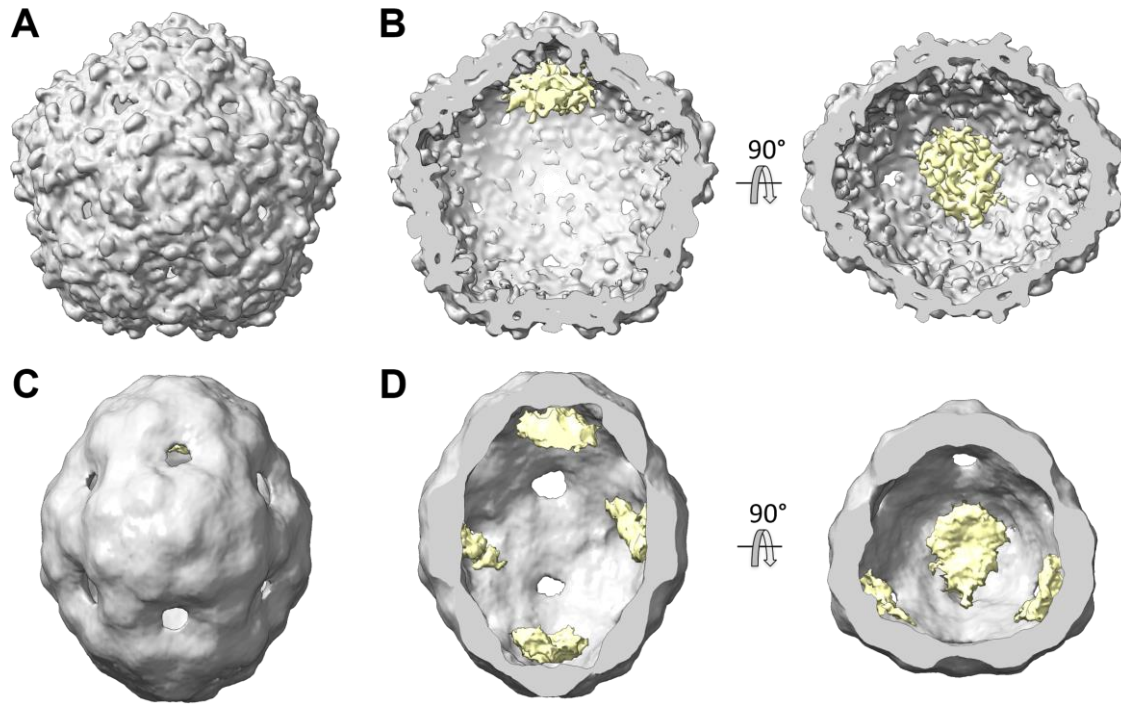


Figure III-13. The extra densities inside the oblate and small prolate VLPs
(A) An asymmetric cryo-EM reconstruction of the oblate VLP. (B) Cut-through view of the asymmetric oblate map. (C) An asymmetric cryo-EM reconstruction of the small prolate VLP. (D) Cut-through view of the asymmetric small prolate map. The density of capsid is color grey. The extra density inside the capsid is colored yellow.

3.2.4 The plasticity in the hexamers supports the formation of different VLPs

Apparently, all these different forms of the VLPs confer different curvatures in the capsids. To investigate the underlying structures that support these different morphologies given only one type of the CPs, we analyzed the assembly and structural

plasticity in the CPs. Notably, all the CP dimers for these different forms of the VLPs exhibit small structural variations mostly localized in the flexible loop regions (Figure III-14). The same holds true for all of the pentamers in every form (Figure III-15). Surprisingly, large structural variations exist within the hexamers in these different forms. Based on its relative locations on the capsid and the symmetry of the encompassing VLP, hexamers can have different arch angles and can be classified into different types (Figure III-16A) based on three angles, θ , ϕ , and ψ , defined by the arch angles of the three opposite CP subunit pairs (Figure III-16B). For the $T=3$ and $T=4$ VLPs, both are formed by only one type of hexamer. For the oblate and small prolate VLPs, there are two types of hexamers in each form. For the prolate VLP, there are three types of hexamers. This would give rise to a total of nine types of hexamers (colored differently in Figure III-16A and III-16C). While the ϕ and ψ angles fluctuate between 150 and 160° for different forms of the VLPs, the θ angle can change dramatically from 130° to 180°. Therefore, we classify the nine hexamers into four different groups based on the values of θ (Figure III-16D-G). The “normal arched” type, which comes from the prolate, oblate, small prolate, and regular T-3 capsids, has a θ angle of 155°. The “over arched” type, which provides more curvature with an θ angle of 130°, includes hexamers from the oblate and small prolate VLPs. The $T=4$ capsid has one type of the hexamer that has an θ angle of 170°, defined as a “less arched” type. Interestingly, the hexamer located on the equator of the prolate capsid has an arch angle of 180°, which is defined as a “straight” type. Such large plasticity within the hexamers of the Q β CPs enables the assembly of different forms of the capsids. Previously it has been reported that some Q β

CP mutants can form rod-like particles (Cielens et al., 2000), which should be elongated using tubes of the straight type hexamers described here.

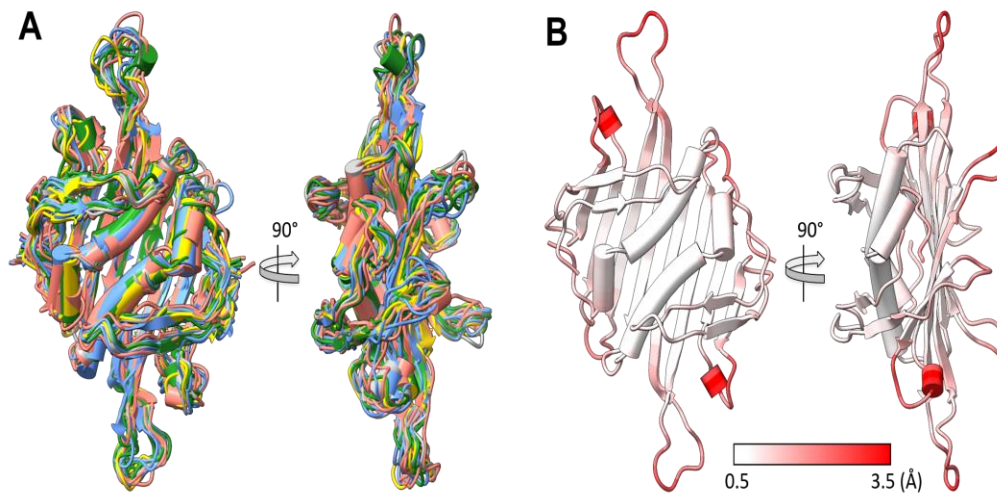


Figure III-14. The comparison of CP dimers in different forms of Q β VLPs. (A) Aligned CP dimers from different forms of Q β VLPs. The CP dimers in the forms of $T=3$, $T=4$, prolate, oblate, and small prolate are colored grey, yellow, green, blue, and red, respectively. (B) The flexibility of a CP dimer. The color scale from white to red shows the root-mean-square deviation of C α of each residue in a CP dimer calculated using dimers in different forms of VLPs.

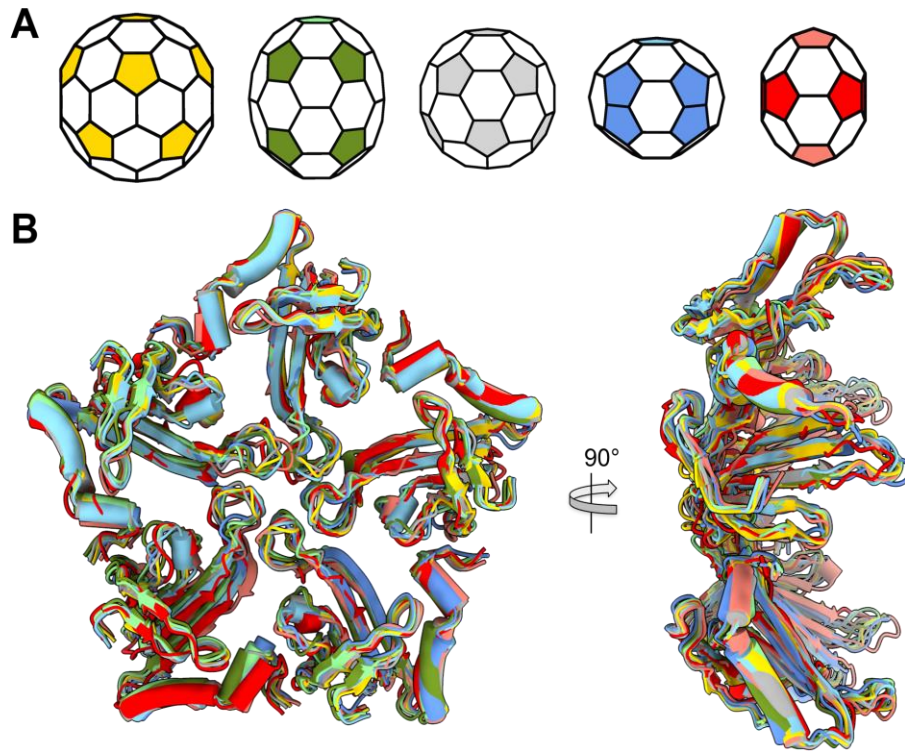


Figure III-15. The comparison of pentamers in different forms of Q β VLPs. (A) The cages of different forms of VLPs with the pentamers colored. (B) Structure overlay of the pentamers in each form against the pentamer in the $T=3$ VLP. Each pentamer is colored by its corresponding color in Panel A.

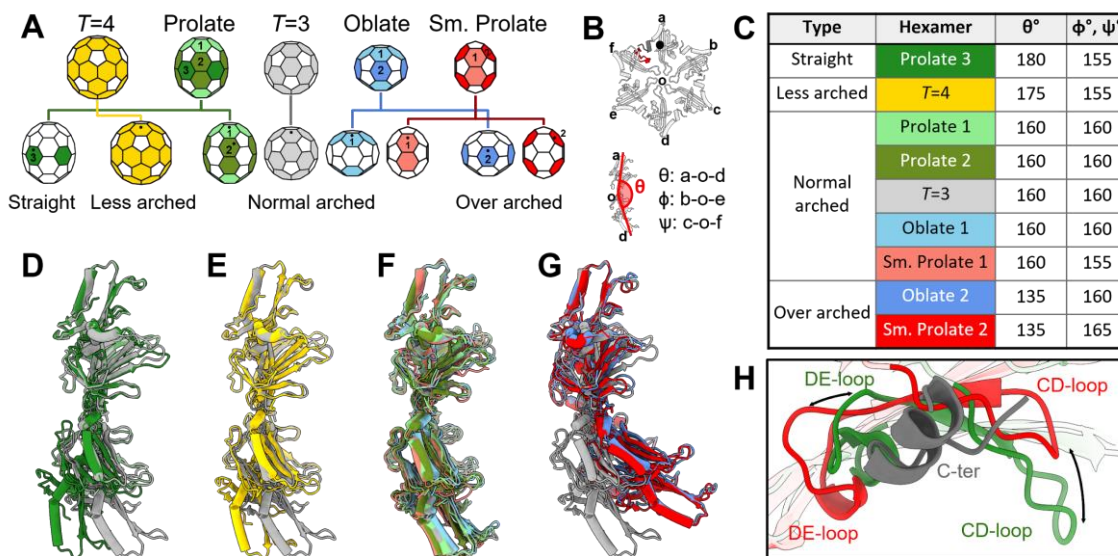


Figure III-16. Plasticity in the hexamers of different Q β VLPs

(A) Different forms of the VLPs with the pentamers in white and hexamers colored. The conformations of the hexamers are classified to be straight, less arched, normal arched, and over arched. (B) The model of a hexamer with the six subunits labeled from a to f. The curvature of the hexamer can be defined by three arch angles θ (between subunits a and d), ϕ (between subunits b and e), and ψ (between subunits c and f). (C) The table showing the values of θ , ϕ , and ψ for each form of the VLPs. ϕ and ψ have the same value for each VLP form due to the symmetry. The names of the hexamers are highlighted with the corresponding color in Panel A. (D-G) Structural overlay of the hexamers in each type against the regular $T=3$ hexamer. The models are aligned based on subunit a of the hexamers as labeled by a black dot in Panels A and B. (H) The relative motion between the C-terminus (dark grey) of a subunit and the CD- and DE-loops of its neighbor. The black arrows indicate the movements of the CD- and DE-loops between the over arched (red) and straight (green) hexamer conformations. The location of the C-terminus, the CD- and DE-loops are labeled in Panel B.

To further investigate how plasticity within the hexamers is allowed, we zoomed into the interface between two neighboring CP monomers within one hexamer. Such an interface is formed between the FG-loops of neighboring subunits and between the C-terminus of one CP and the CD- and DE-loops of its neighbor (Figure III-17). While the neighboring FG-loops are linked together through disulfide bonds (Golmohammadi et

al., 1996), which are always maintained among all these different types of hexamers, the interactions do change between the C-terminus of one CP and the CD- and DE-loops of its neighbor (Figure III-16H). If we align the CP **a** in a hexamer (Figure III-16B), from the over arched type to the straight type of the hexamer, the CD-loop and DE-loop of CP **f** move 14Å away from and 6Å towards the C-terminus of CP **a**, respectively. Such a relative rocking of the CD- and DE-loops around the C-terminus of the neighbor, provides the flexibility within the hexamers and allows subunits to tilt up or curve down relative to each other to achieve different curvatures of the hexamers, facilitating the assembly for the different forms of the capsids.

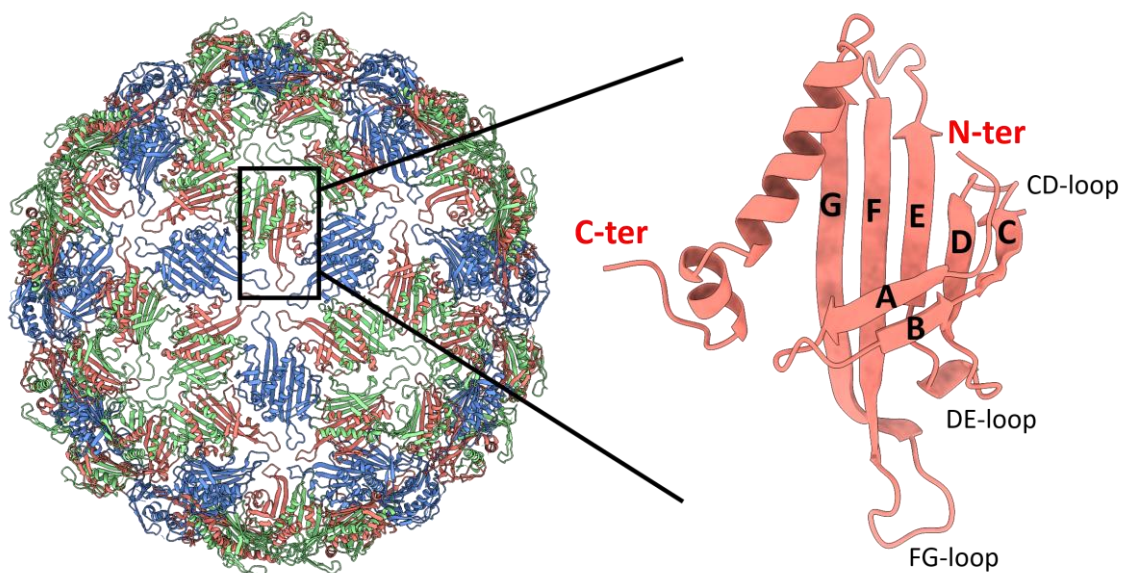


Figure III-17. The structure of Q β CP monomer

A Q β CP monomer (right) is extracted from the capsid (left). The β -strands are labeled from A to G. The CD-loop, DE-loop, FG-loop, N- and C-termini are also labeled.

3.3 Discussions

In this paper, we have presented the complete model of the Q β gRNA in its capsid, which provides insight into the assembly process of the wild-type Q β . The folding of gRNA is consistent with the fact that the secondary structures start to form in the nascent gRNA during replication, which is further folded into local domains. Such a co-replicative folding of the gRNA is beneficial to the phage since it prevents the nascent gRNA from annealing back to the template RNA, which has been shown to hinder RNA replication (Tomita, Ichihashi, & Yomo, 2015). The fact that the gRNA folds domain-by-domain inside the virion into a dominant 3D conformation is consistent with the gradual folding of the nascent gRNA and the sequential binding of the capsid proteins, which significantly reduce the complexity of the RNA-folding trajectories. The large pool of the translated CP dimers in the vicinity of the gRNA will bind to the operator-like RNA stem-loops in these partially folded RNA domains to stabilize the folded domains. If the gRNA were to fold and bind the capsid proteins after it has been fully synthesized, it may fold into more heterogeneous conformations. It has been shown that in solution, the binding of the 19-nt long operator RNA would change the conformation of the C/C dimer to an A/B dimer, which presents an asymmetric conformation in the FG-loops that favors the formation of a pentamer (Stockley et al., 2007). This is consistent with our observation that most of the operator-like RNA stem-loops are under A/B dimers of the CPs in the wild-type Q β . Indeed, the high concentrations of the operator-like RNA stem-loops, present in adjacent regions of the Q β gRNA sequence, may favor the formation of these pentamers, which are embedded

into a near icosahedral $T=3$ capsid by the free FG-loops of the other subunit in the dimers, within the pentamer, to form hexamers.

The locations of these operator-like RNA stem-loops in the entire gRNA also suggest an order for the RNA packaging and phage capsid formation in Q β (Figure III-18). In packaging a nascent gRNA that is still coming out of the replicase, the first few operator-like stem-loops at the 5' region of the gRNA, which are synthesized at the beginning, first bind CP dimers to coordinate the formation of Pentamer I. Other pentamers and hexamers are then formed, with the last stem-loop in the gRNA, U1, binding to the Mat to be incorporated into the capsid. Intuitively, such a gradual folding and packaging of the gRNA may confine the flexibility of the gRNA and help such a ~4,000-nt long RNA fold into a defined 3D structure. Such a proposed model for RNA-mediated capsid assembly is consistent with our previous observation that the capsid of the wild-type Q β deviates slightly from an ideal icosahedral cage (Cui et al., 2017; Gorzelnik et al., 2016), particularly by exhibiting a crack in the capsid around Pentamer X, where the β -region of the Mat points. Such an imperfection of the capsid suggests this patch of the CP shell is formed at the end of the capsid assembly when gRNA is folded and packaged with very little room for the adjustment of the capsid to completely close up.

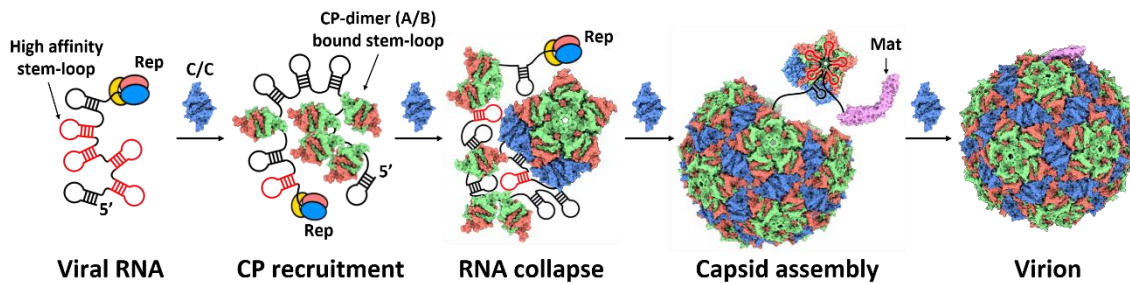


Figure III-18. Proposed model for the co-replicative assembly of the Q β virion
 Proposed assembly model for the Q β virion along with gRNA replication. The gRNA synthesized by the replicase (Rep) presents several high-affinity stem-loops for CP dimers, which upon binding to CP dimers change their conformations from C/C to A/B. The protein-protein interactions of nascent CP dimers facilitate RNA collapse and form the first pentamer, then other pentamers and hexamers. Almost towards the end of the assembly (after stem-loop U1 has been synthesized), the Mat binds to the 3' UTR and is incorporated into the capsid to form a complete virion.

Compared with a related phage MS2, in which almost all the particles show a defined 3D RNA structure (Dai et al., 2017), Q β only has 25% of the particles showing a defined 3D structure of the encapsidated RNA (Cui et al., 2017). One striking difference between the two phages is the charge in the inner surface of the capsids. For Q β , the CP contains a highly positively-charged EF-loop facing the inside of the capsid, which can interact with the negatively charged RNA backbone. Such an electrostatic interaction can facilitate the recruitment of the CP dimers onto the viral gRNA, but on the other hand, since the electrostatic interaction is non-specific, this may coordinate the CP dimers onto non-specific locations on the RNA backbone, leading to heterogeneous packaging of the RNA for each phage particle or packaging of non-viral RNAs. Another striking feature of the Q β CP is the presence of a small percentage of the minor capsid protein A₁ due to the readthrough of the leaky stop codon at the end of the *cp* gene (Hofstetter et al., 1974). Although where A₁ is integrated into the phage capsid is not

known, however, since the extension for the extra domain of A_1 is from the C-terminus of the CP and the C-terminus is involved in the subunit interface, the linker that connects the CP to the A_1 domain may affect the capsid assembly, potentially lead to more flexibility in the hexamers.

While there are always twelve copies of the pentamers in different forms of the $Q\beta$ VLPs, the numbers of the hexamers do vary. Interestingly, the geometries of a few forms of these capsids are related. For example, from the oblate capsids, one can convert it to a $T=3$ capsid by separating the two end-shells from the equator and adding five more hexamers where the two half-shells rejoin (Figure III-19). The same procedure can repeat to generate the prolate capsid from the $T=3$ capsid. One can even repeatedly add hexamers in a multiplication of five in the middle of the prolate to obtain the rod-like VLP particles. Such forms of the rod-like VLPs will need a significant number of hexamers relative to the pentamers, which is achieved by the mutations in the FG-loop (Cielens et al., 2000).

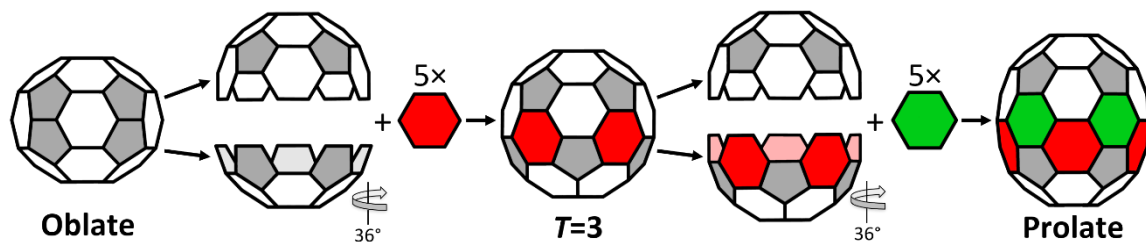


Figure III-19. The geometrical relationship between the oblate, $T=3$, and prolate VLPs

The geometrical relationship between the oblate, $T=3$, and prolate VLPs. The conversion from the oblate to $T=3$ VLPs can be achieved through the division of the oblate VLP,

along the 5-fold axis, into two hemispheres, with one hemisphere rotated 36° relative to the other, followed by the insertion of five hexamers. The elongation from the $T=3$ to prolate VLP follows the same protocol with an addition of five more hexamers.

Besides the 5-fold non-icosahedral capsids, Q β can also form a small prolate VLP elongated along the 3-fold axis which is a new conformation of the virus capsid. In the elongated face, different from the 5-fold capsids having the same triangulation number, the 3-fold small prolate consists of two different triangulation numbers, $Q_1=2$ and $Q_2=3$ (Figure III-11). Moreover, the triangulation number of the two caps of the small prolate is $T=1$, which means that Q β has the potential to form a $T=1$ icosahedral VLP. However, we did not find it in this study.

One possible reason for Q β CPs to form different forms of capsids is to have the capability to enlarge the size of the capsid for RNA recombination in evolution. RNA recombination plays an important role in RNA viruses. That allows an RNA virus to increase genetic drift to adapt new functions, for example, binding to new hosts, evolving lysis proteins to lyse the new hosts, evolving new lysis proteins for their current host, etc. So, there may be some advantages in making larger VLPs to package a larger recombinant RNA. The two smaller capsids, oblate and small prolate, might be the accompanying products in the evolution.

For enlarging a capsid, there are two classical mechanisms: (1) increasing the triangulation number, like the $T=3$ to the $T=4$ icosahedral VLP, and (2) conversion of an isometric to prolate head, like the $T=3$ icosahedral to the prolate VLP. In the first case,

hexamers are added across all faces of the capsid, while in the second, hexamers are added across the five-fold axis. These two mechanisms can be easily achieved through simple mutations and have been observed in different viruses, such as Q β (Fiedler et al., 2012), Sindbis virus (Ferreira, Hernandez, Horton, & Brown, 2003), and Rous sarcoma virus (Butan et al., 2010). Moreover, single point mutations in T4 phage can convert the capsid from prolate to isometric or generate giant heads (Doermann, Eiserling, & Boehner, 1973). Therefore, the evolutionary barriers for altering capsid volume through these two classical mechanisms seem to be low.

However, it also has a possibility that these non-canonical VLPs are just byproducts of the packaging strategy of Q β . Q β has evolved to do packaging in a simple way. Q β efficiently packages its RNA while regulating the replicase. These non-canonical VLPs may be produced in controlling the replicase expression. Evolutionarily, the non-canonical VLPs are mostly dead ends because the smaller ones may not have enough space to package the entire genome and the bigger ones may not have the maturation protein packaged.

It has not been studied how the CPs structurally protect the RNA of different lengths. Our results that over-expressing the 986-nt *cp/A₁* gene leads to more percentages of the oblate (from 21% to 35%) and the small prolate (from 3% to 12%) forms, suggesting the RNA length may have an effect on the forms of the VLPs. As we have seen in the wild-type Q β , the entire wild-type gRNA forms operator-like stem-loops to

coordinate the pentamers for the assembly as well as the long helical structure joined by kissing-loops that set the diameter of the capsid, but how RNAs get packaged into VLPs is still not clear. However, the diversity of the forms of VLPs may allow manipulations of the ssRNA phage CPs to package RNA of different lengths for various biomedical or biotechnological applications.

3.4 Materials and Methods

3.4.1 Image processing of the Q β gRNA

To obtain a better gRNA density, we combined the images of the Q β particles with and without MurA (the lysis target) bound, collected from the previous study (Cui et al., 2017) with different pixel sizes. All the images are scaled to a pixel size of 1.25 Å. After the 2D cleaning and the 3D classification using RELION (Scheres, 2012), we obtained 86,081 particles with a defined gRNA structure. The unsupervised asymmetric refinement yielded a final map at 6.1-Å resolution before sharpening and 4.6-Å resolution after sharpening, while the gRNA density was less noisy and more complete in the unsharpened map, which was used for the modeling of the RNA.

3.4.2 Modeling of wild-type gRNA

To locate the starting point for tracing the entire gRNA structure of Q β , we first focused on the density attaching to the Mat and the internal CP dimer. This density shows a clear 5-way junction geometry and is unique in the map of Q β gRNA. To find the sequence of this gRNA fragment, we used two methods: One is to predict the

secondary structure of the gRNA by the phylogenetic information (Beekwilder, 1996); the other is to predict local secondary structures using the minimum-free-energy method with sliding-windows of 100~120 nucleotides (Reuter & Mathews, 2010). Both of these methods show that only the 3' UTR can form a 5-way junction with compatible helical lengths for each encompassing stem-loop. Coincidentally, the 3' UTR of a related ssRNA phage, MS2, also interacts with its Mat (Dai et al., 2017). More generally, the folds of the 3' UTRs in all ssRNA phages are very similar, despite the differences in their hosts and gRNA sequences (Klovins et al., 2002). Therefore, the high-resolution structure of this region from the MS2 was used as a reference to model the structure of the 3' UTR of the Q β . Based on the density, the structure of the Mat-bound gRNA stem-loop of the Q β , U1, is similar to the U1 of MS2, which could not be computationally predicted correctly. Compared to the predicted secondary structure, one G-C pair in the U1 of Q β needs to be separated to form a bulge with two new G-C pairs formed at the tip (Figure III-3). The modified U1 was refined using Rosetta RNA Denovo (Cheng et al., 2015; Das et al., 2010) into the density at 4.4-Å resolution (EMD-8708) (Cui et al., 2017) and then was used to build the entire 5-way junction of 3' UTR based on the secondary structure by Rosetta RNA. The final model of the 5-way junction fits well with the density after the HNMMC (J. Zhang et al., 2012) and MDFF (Singharoy et al., 2016) refinements.

Starting from the structure of the 5-way junction, the entire gRNA structure of the Q β can be traced using its secondary structure from the phylogenetic prediction

(Beekwilder, 1996). Combining the computational tools of RNA modeling and EM density fitting, including Rosetta RNA Denovo, MOSAICS-EM (J. Zhang et al., 2012), and Situs (Wriggers, 2012), we built the atomic models of the Q β gRNA fragments based on the corresponding segmented densities of the unsharpened map. These segments were then combined to produce a final model of the complete Q β gRNA, which was refined using the real-space refinement in PHENIX (Adams et al., 2010) with the secondary-structure restraints.

3.4.3 Purification of Q β and VLPs using gel-filtration

In order to purify the non-infectious VLPs, either from over-expressing the *cp/A₁* or basal-level expression of the U1 and R1 mutants, another purification protocol was established. Single colonies of BL21(DE3) in the case of *cp/A₁* (pET28) or DH10B for wild-type Q β and the U1 or R1 mutants (pBRT7QB) were used to inoculate starter overnight bacterial cultures with appropriate antibiotics. The starter cultures were used to inoculate 6 x 500 ml cultures with 1:4 aeration grown at 37°C. The *cp/A₁* cultures were induced with 0.5 mM IPTG once the absorbance was at OD₆₀₀ = ~0.6, then grown for ~4 hrs. The cells were then spun down, resuspended in Q β buffer, and lysed using a French Press. The lysate was clarified by spinning at 25,000 x *g* and the supernatant was saved. The supernatant had PEG6000 and NaCl added to 10% w/v and 500 mM, respectively. The samples were stored at 4°C, then spun down at 8,000 x *g* for 30 min with the supernatant decanted and the pellet was resuspended in 125 mM NaH₂PO₄, pH=8.0. The resuspended pellet was mixed with an equal volume of chloroform and

vortexed. The suspension was spun down at 40,000 x g and the aqueous layer was removed and extracted with chloroform, in the same manner, four more times. After the PEG was extracted with chloroform in this manner, the samples were concentrated using a 10 kDa MWCO centrifugal concentrator until the volume was ~3-4 ml. The samples were loaded onto a HiPrep 26/60 Sephacryl S500 column, run with 125 mM NaH₂PO₄. Between different samples, the column was washed with 0.5 N NaOH and then buffer. Purifying wild-type and the U1/R1 mutants were the same as purifying the *cp/A₁* VLPs, except that they were not overexpressed. Instead, the phage/mutants were produced from basal-level transcription from pBRT7QB. Cells were removed by spinning down at 4,000 x g, then the supernatant was filtered and PEG-precipitated as above.

3.4.4 Purification of Q β and VLPs using CsCl gradient

The purification using CsCl gradient is almost the same as using gel-filtration. Only the last step of going through the gel-filtration column is replaced by the CsCl gradient. The sample after the PEG-precipitation step and chloroform extraction, CsCl is added to 0.550 g/ml to make the density of the resulting solution 1.38 ± 0.01 . The solution is divided into tubes and spun in the centrifuge for 24 hours at 37,000 RPM (~105,000 x g) at 2°C. The tubes are then carefully removed. Usually, two bands are visible in the tube, with the virus band located lower. The virus band of each tube can be removed with a syringe and pooled. The combined material is placed in two or three tubes and centrifuged for the second centrifugation. Upon completion of the second run, the virus band in each tube is collected by punching a small hole in the bottom of the

tube and allowing the material to drop out of the bottom. The virus isolated from each tube is combined and dialyzed against three changes of buffer.

3.4.5 Electrophoretic mobility shift assay

The RNA fragments were *in-vitro* transcribed using HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). The linear DNA templates for *in-vitro* transcription were amplified by PCR from the Q β cDNA. The binding reaction contains 1 μ M of the RNA fragment, 10 μ M of purified MBP tagged Mat (MBP-A₂) or 10 μ M of purified MBP (6xHis-MBP), 2 U/ μ l of SUPERase•In™ RNase Inhibitor, and the binding buffer (20 mM Tris-HCl pH 7.6, 150 mM NaCl). The reactions were incubated for 20 mins at 37°C. After the incubation, the 1X EMSA loading dye was added to the reaction. The final volume of 10 μ l of the reaction was loaded into 5% Native-TBE polyacrylamide gel and separated at 200 mV for 30 mins at 4°C. The gel was then stained with SYBR™ Green and subsequently stained with SYPRO™ Ruby using the Electrophoretic Mobility Shift Assay (EMSA) Kit (ThermoFisher).

3.4.6 Cryo-EM specimen preparation and data collection

The Cryo-EM grids were prepared by applying 3 μ L of the sample solution to a C-flat 1.2/1.3 400-mesh Holey Carbon Grid (Protochips) at 22°C with 100% relative humidity and vitrified using a Vitrobot Mark III (FEI Company). Movie stacks of the gel-filtration purified Q β particles (wild-type and the *cp/A₁* overexpression) were collected on an FEI Tecnai F20 cryo-electron microscope operated at 200 kV. Data were recorded using

SerialEM (Mastrorade, 2005) on a K2 Summit direct detection camera (Gatan) at a nominal magnification of 29,000 x g, yielding a subpixel size of 1.25 Å. Each movie stack has 33 frames with 0.2 s per frame for a total exposure time of 6.6 s, yielding a total dose of $\sim 36 e^- / \text{Å}^2$.

3.4.7 Cryo-EM data processing

The movie stacks of the gel-filtration purified particles were aligned and filtered by MotionCor2 (Zheng et al., 2017) (1.25 Å/pixel). The images of the CsCl-purified particles are from the previous collection (Cui et al., 2017) (2,370 micrographs, 1.216 Å/pixel). All the Q β particles were picked by Gautomatch from Dr. Kai Zhang and then downsampled by two times. The particles were separated by their 2D classified images ($T=4$, oblate, $T=3$, prolate, and small prolate, respectively). In total, for the CsCl purification, we have 1,138 $T=4$ particles, 4,316 prolate particles, and 1,489 oblate particles; for gel-filtration purification, combining wild-type Q β and over-expressed *cp/A1*, we have 1,635 prolate particles, 38,125 oblate particles, and 6,842 small prolate particles. Due to the small numbers of the $T=4$ and prolate VLPs, we combine the corresponding particles from the new data from the TF20 cryo-EM with a Gatan K2 camera with our previously collected data from the JEM3200FSC cryo-EM with a Gatan K2 camera. The new data are rescaled to the pixel size of 2.432 Å, which is 2-time downsampled compared to the original image. Similarly, the merged particles of the oblate and small prolate VLPs were rescaled to 2.5 Å, which is 2-time downsampled compared to the original images. For the symmetry, we applied icosahedral for $T=4$ particles, D5 for

prolate and oblate particles, and D3 for small prolate particles during the refinements. The refinements yielded the maps of $T=4$, prolate, oblate, and small prolate particles at 5.6-Å, 6.1-Å, 6.2-Å, and 8.9-Å resolutions, respectively (Figure III-20). Refinements of the particles without downscaling did not improve the final resolutions of the maps.

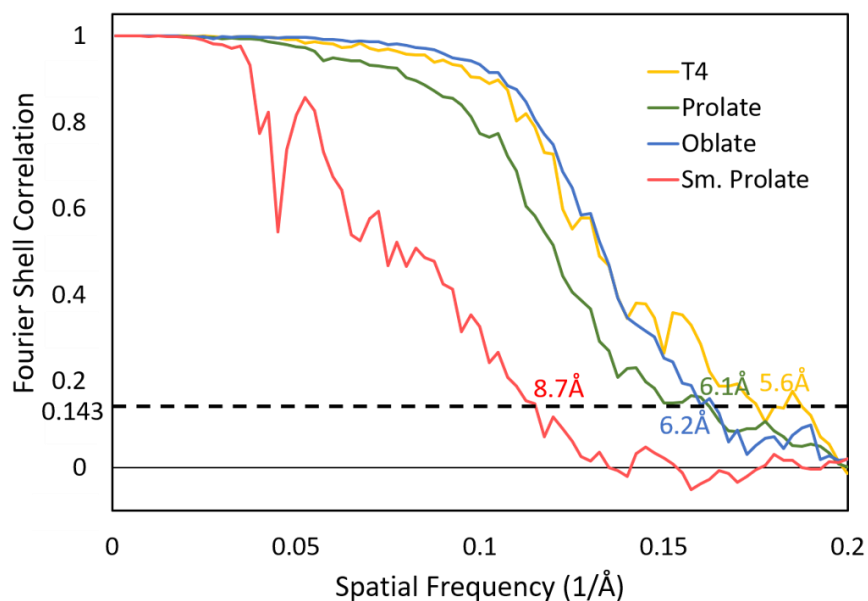


Figure III-20. The Fourier Shell Correlation (FSC) curves

The FSC curves for the single-particle cryo-EM maps of the Q β VLP with $T=4$, prolate, oblate, and small prolate conformations are shown in yellow, green, blue, and red, respectively.

3.4.8 Modeling of VLP capsids

To build the atomic models for the Q β capsids of different forms, the structure of the CP from an icosahedral refinement (PDB ID: 5KIP) was first rigidly fit and subsequently refined into each VLP capsid. The refinements were performed by using RosettaCM (Wang et al., 2016) with icosahedral symmetry for the $T=4$ particle, D5

symmetry for the prolate and oblate particles, and D3 symmetry for the small prolate particle.

CHAPTER IV

STRUCTURAL ASSEMBLY OF SATELLITE PLANT VIRUSES

4.1 Introduction

Satellite viruses are a group of subviral molecules that require a helper virus for replication and movement (Scholthof, Jones, & Jackson, 1999; Simon, Roossinck, & Havelda, 2004). Satellite panicum mosaic virus (SPMV) and satellite tobacco mosaic virus (STMV) are two well-known satellite viruses in plants that need their helper viruses, typically panicum mosaic virus (PMV) and tobacco mosaic virus (TMV) respectively, to replicate and systemically spread in plants (Turina, Desvoyes, & Scholthof, 2000). These two viruses are both positive-sense single-stranded RNA (ssRNA) viruses containing 824 bases for SPMV and 1,508 bases for STMV. Interestingly, these genomic RNAs (gRNAs) encoding their coat proteins (CPs) both have low sequence similarity with the gRNAs of their helper viruses (Mirkov, Mathews, Du Plessis, & Dodds, 1989; Turina, Maruoka, Monis, Jackson, & Scholthof, 1998).

Both SPMV and STMV are $T=1$ icosahedral viruses with a diameter of 17 nm (Ban & McPherson, 1995; Larson et al., 1998). Each capsid of 60 identical CPs encloses and protects its gRNA. The crystal structure of STMV reveals that 30 RNA double-helical segments, each about nine base pairs, are closely attached to the CP dimers at the 2-fold axes inside the capsid (Larson et al., 1998). However, the linkers between these helical segments cannot be clearly visualized. For SPMV, the gRNA does

not have clear helical segments associated with the CP dimers in the crystal structure (Makino, Day, Larson, & McPherson, 2006), but the gel mobility shift assays show that its CP has a high affinity for binding its gRNA (Desvoyes & Scholthof, 2000).

With the constraint on the 30 gRNA helical segments of STMV, Larson and McPherson proposed that the gRNA of STMV locally forms a series of short-range stem-loops that serve as package signals and the CPs bind to these stem-loops when they emerge from the replication complex (Larson & McPherson, 2001). This hypothesis is supported by the results of atomic force microscopy (AFM) experiments (Kuznetsov, Dowell, Gavira, Ng, & McPherson, 2010). Based on the assumption, combining the data from chemical probing, Schroeder et al. produced an ensemble of models for the secondary structure (Schroeder, Stone, Bleckley, Gibbons, & Mathews, 2011) Each model has 30 gRNA stem-loops of only local base pairing while the long-range base pairs are absent in all these models. Using these models of secondary structure, Zeng et al built an all-atom model of STMV including every amino acid and nucleotide (Zeng, Larson, Heitsch, McPherson, & Harvey, 2012). However, the *in vitro* result of the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis is not consistent with these models. The *in vitro* transcribed STMV gRNA contains long-range base pairing (Athavale et al., 2013). Besides, Larman et al revealed that the STMV gRNA adopts distinct conformations in the packaged and unpackaged state (Larman, Dethoff, & Weeks, 2017). Therefore, the organization of the STMV gRNA inside the capsid is still an open question.

Like Q β (Cui et al., 2017; Gorzelnik et al., 2016) and MS2 (Dai et al., 2017), here, the single-particle cryo-EM analyses were applied on SPMV and STMV to obtain the gRNA structures of these two small plant viruses. We successfully reconstructed similar icosahedral structures with symmetry applied in both viruses. However, defined gRNA structures cannot be visualized with asymmetry reconstruction. Then, we used the focused classification which excludes the EM single of the capsid to focus on classifying the gRNA only. Similar features can be revealed in the classified gRNA densities, but a defined gRNA structure with clear grooves still cannot be visualized.

4.2 Results

4.2.1 Icosahedral Structure of SPMV and STMV

With icosahedral symmetry applied, the particle images of SPMV and STMV yield the cryo-EM density maps of 3.8-Å and 5-Å resolution, respectively (Figure IV-1). For SPMV, as the crystal structure, the capsid is built by 60 copies of the CP. Each CP is formed by an eight-stranded β -barrel core with a “jelly roll” topology. Sidechains for the coat protein are clearly seen and the structure is consistent with the previously solved crystal structure (PDB ID: 1STM) (Ban & McPherson, 1995). The density of residue 1 to 16 of the SPMV CP is also missing in the EM map. The missing fragment is Arginine-rich and may be embedded in the gRNA for the genome packaging and structural stability. Besides, the residues Tyr₂₄, Lys₁₀₁, Thr₁₀₃, and Arg₁₄₈ at the inner surface have their side chains facing the genome (Figure IV-2A). These residues may

play an important role in interacting with the gRNA in SPMV. For STMV, although the resolution of its icosahedral EM map is lower while side chains are not clear, the structures of α -helices and β -barrels of the CP are clearly resolved as the crystal structure (PDB ID: 1A34). The flexible N-terminal fragment wraps around the RNA density, but the first 12 residues cannot be visualized in the EM map and are also Arginine-rich. (Figure IV-2B). The missing fragments at a similar position and with similar residue composition imply that this region may serve an identical function in SPMV and STMV.

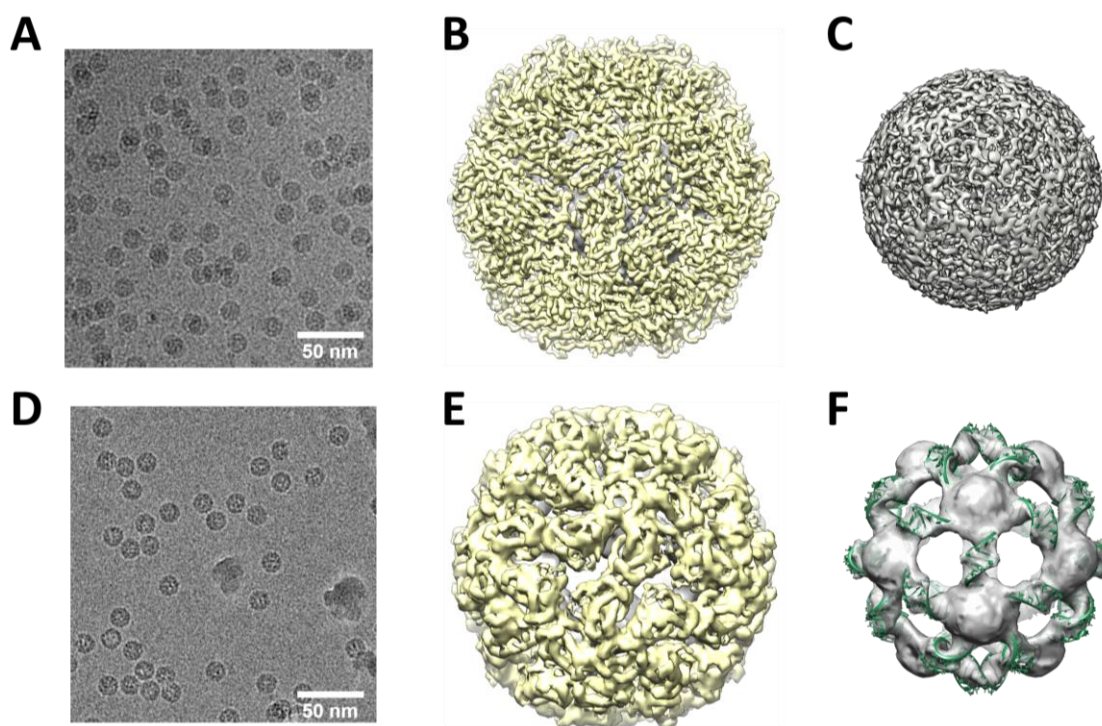


Figure IV-1. The structures of SPMV and STMV

(A) The single-particle cryo-EM image of SPMV. (B) The EM density map of SPMV. (C) The SPMV gRNA density packaged inside the capsid. (D) The single-particle cryo-EM image of STMV. (E) The EM density map of STMV. (F) The STMV gRNA density packaged inside the capsid. Models of 30 gRNA helical fragments attached to the capsid

are shown in green. The densities of the capsid and the RNA are colored yellow and grey, respectively.

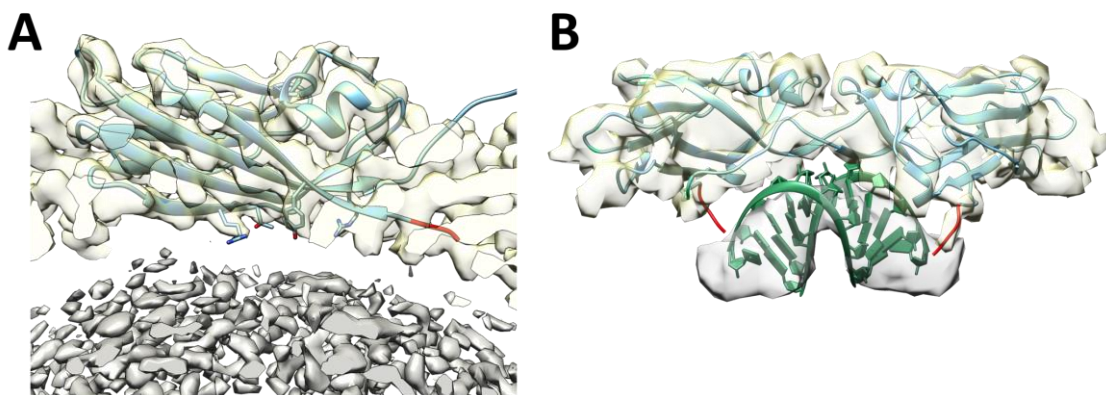


Figure IV-2. The interaction of the CP and the RNA in SPMV and STMV

(A) The SPMV CP (blue) with its EM density. The densities of the capsid and the RNA are colored transparent yellow and grey, respectively. The side chains showing at the inner surface from left to right are Lys₁₀₁, Thr₁₀₃, Tyr₂₄, and Arg₁₄₈. The red fragment is the N-terminus with the first 16 residues missing. (B) The STMV CP dimer (blue) wrapped around the RNA helical fragment (green). The densities of the capsid and the RNA are colored transparent yellow and transparent grey, respectively. The red fragment is the N-terminus with the first 12 residues missing.

In both viruses, gRNA densities are also found to attach the capsids. For STMV, like crystal structure, the density of around nine base-paired gRNA helix interacts with two coat proteins at every 2-fold axis (Figure IV-1F). 30 helical densities connecting with strong densities at 5-fold apexes build the entire gRNA densities of STMV under icosahedral symmetry applied. For SPMV, the gRNA densities only form a spherical shell instead of having clear helical densities like STMV. The missing density of the RNA helix implies that compared to it of STMV, the gRNA of SPMV may have weaker interactions with the capsid. Measurements of the density distribution of the virus EM

maps also show that the gRNA has a tighter binding with the capsid protein in STMV (Figure IV-3).

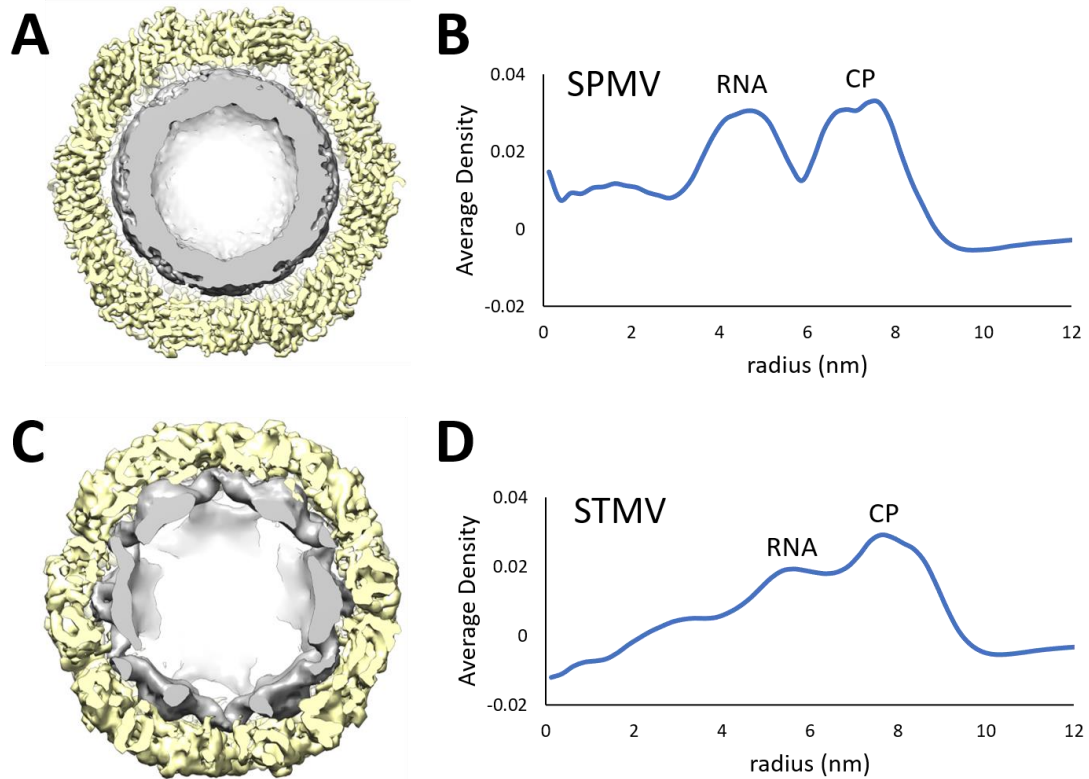


Figure IV-3. The composition of the SPMV and STMV EM maps

(A) Cut-open view of the SPMV EM map. (B) The density distribution of the SPMV EM map. (C) Cut-open view of the STMV EM map. (D) The density distribution of the STMV EM map. The densities of the capsid and the RNA are separated based on the crystal structure of the capsid. The capsid density is colored yellow and the RNA density is colored grey.

4.2.2 The asymmetric structure of SPMV and STMV

In both Q β and MS2, the EM maps reconstructed with icosahedral symmetry and without symmetry applied represent different gRNA densities inside the capsids. There is an apparent symmetry mismatch between the capsid and the gRNA. Therefore, I

reconstructed the EM maps of SPMV and STMV without symmetry applied. The resolutions of SPMV and STMV asymmetry maps are 7.0 and 8.2 Å, respectively (Figure IV-4). For the SPMV, the densities of the capsid and the gRNA are similar to them reconstructed with icosahedral symmetry applied. The gRNA density is still a spherical shell and a defined helical structure cannot be visualized. For STMV, compared to the icosahedral symmetry map, besides the relatively weak RNA density attached to the capsid, the asymmetry map has additional RNA density at the center of STMV, but the density does not show a clearly defined RNA long helix or stem-loop. These unstructured RNA densities mean that the asymmetry reconstructions are not that successful in both STMV and SPMV.

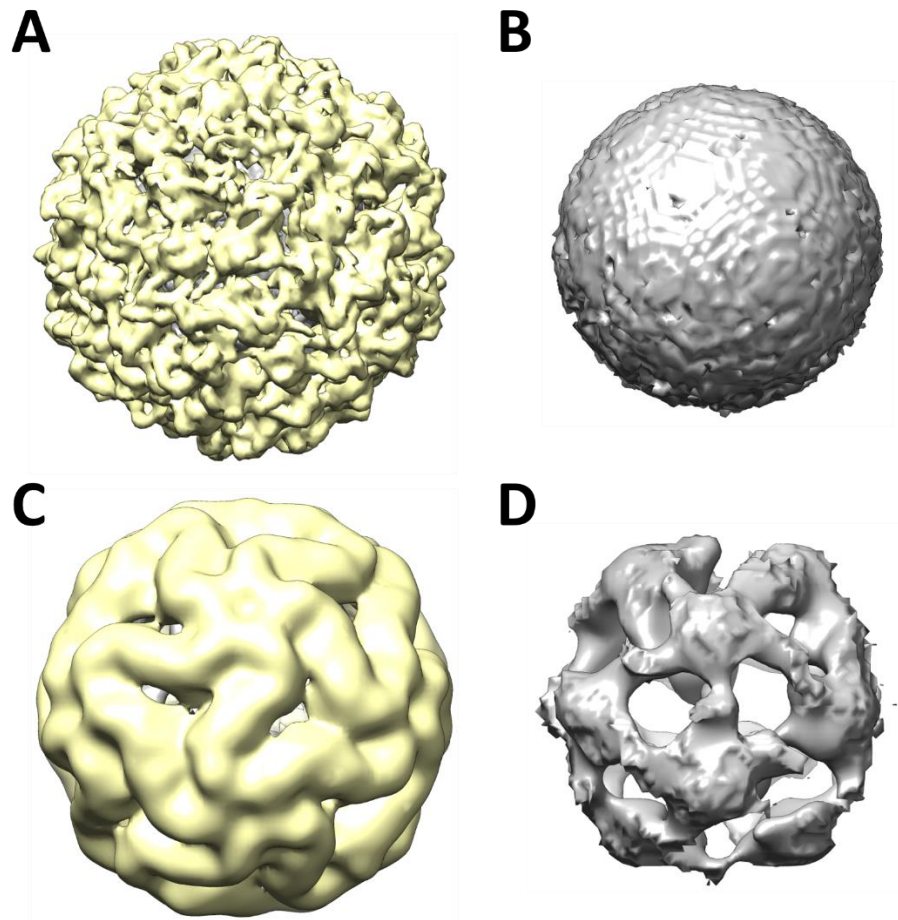


Figure IV-4. Asymmetric reconstruction of SPMV and STMV

(A) The asymmetric map of SPMV. (B) The SPMV gRNA density by asymmetric reconstruction. (C) The asymmetric map of STMV. (D) The STMV gRNA density by asymmetric reconstruction. The densities of the capsid and the RNA are colored yellow and grey, respectively.

4.2.3 Focused Classification of the gRNA inside SPMV and STMV

The icosahedral-like EM density under the asymmetry reconstruction brings one possible problem that the capsid of a virion may provide a strong signal and guide the EM map reconstruction. Therefore, I apply the focused classification (Bai et al., 2015) which removes the signal of the capsid to specifically separate the conformations of only

gRNA (Figure IV-5). To obtain the gRNA only images, the projection of the capsid is generated using the asymmetry reconstructed map and then is subtracted from every experimental particle image. The orientation of the projection of each particle image is also obtained from the consensus reconstruction.

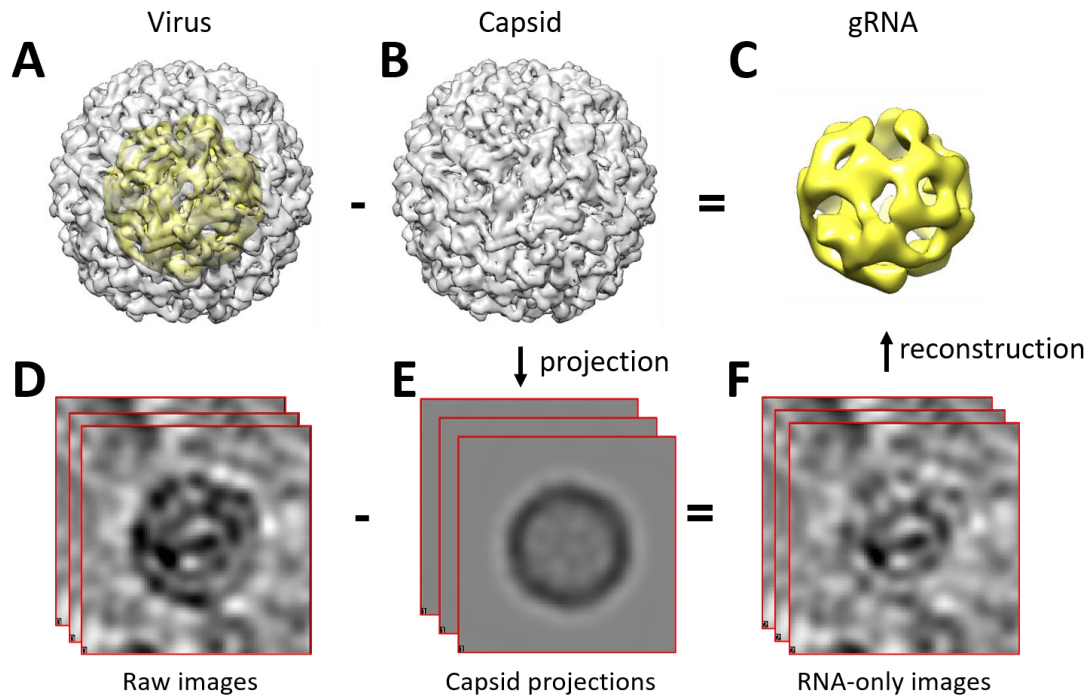


Figure IV-5. Focused classification with capsid-signal subtraction

(A) A 3D reconstructed map of a virus. The densities of the capsid and the RNA are colored transparent grey and yellow, respectively. (B) The capsid density that would like to ignore in focused classification. (C) The RNA density which would like to focus classification on. (D) Experimental particle images assumed to be 2D projections of the entire virus in panel A with different projection orientations. (E) 2D projections of the capsid density using the orientations from the consensus reconstruction. (F) The projection of the RNA only after capsid-signal subtraction.

Based on the number of the particles in each dataset, these subtracted particle images containing only the gRNA signal were classified into six and four classes on

SPMV and STMV, respectively (Figure IV-6 and IV-7). Unfortunately, none of these classified maps have a clear RNA helical structure in both datasets, which means that a defined gRNA structure still cannot be found in both viruses. I also did refinement of these classified RNA-only images, but I still cannot find any clear RNA helical structure. However, although the density of the helical structure is not clear, in some classes, some similar features can be found. These features imply that the gRNA of each virus may be packaged using a similar way in some regions.

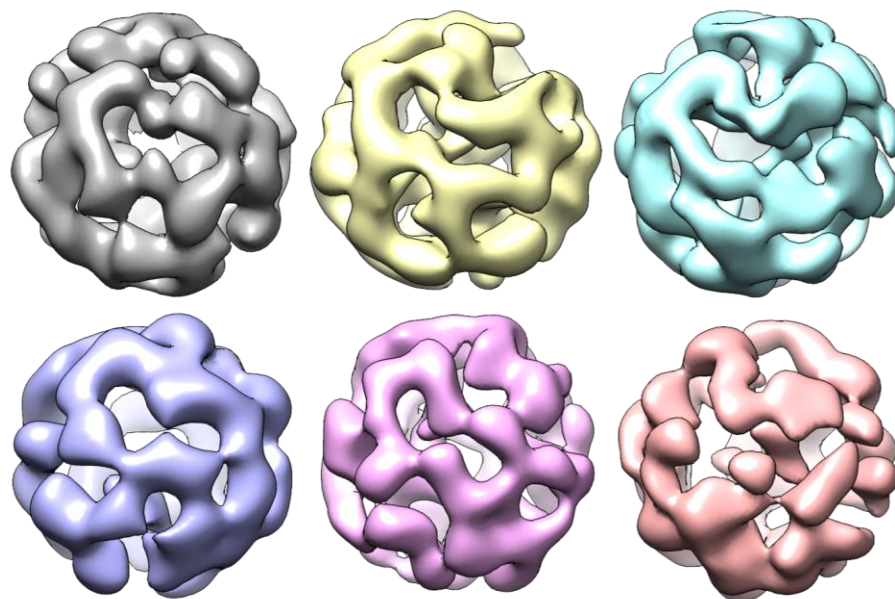


Figure IV-6. Focused classification of the SPMV gRNA

The RNA-only particle images of SPMV were separated into six classes. The orientation of each classified map is selected to represent a similar view of these maps which may contain similar features.

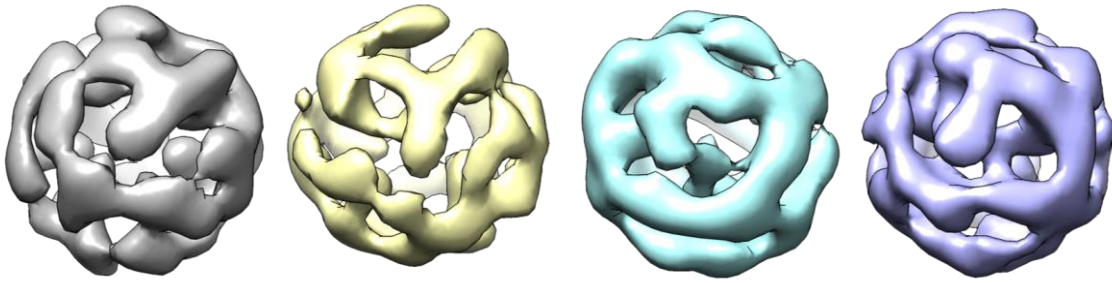


Figure IV-7. Focused classification of the STMV gRNA

The RNA-only images of STMV were separated into four classes. The orientation of each classified map is selected to represent a similar view of these maps which may contain similar features.

4.3 Discussion

Using single-particle cryo-EM, the gRNA densities of SPMV and STMV inside the capsids can be visualized. With the icosahedral symmetry applied, STMV shows the gRNA densities of 30 helical fragments which are around half of the genome attached inside the capsid. Several models have been proposed based on the constraint of these 30 helices which yield that all these proposed models only can form short-range RNA base pairs (Schroeder et al., 2011; Zeng et al., 2012). Similar icosahedral ordered RNA densities are also found in other viruses, like Bean Pod Mottle Virus (BPMV) and Pariacoto Virus (PaV) (Chen et al., 1989; L. Tang et al., 2001). However, without symmetry, the defined gRNA structures inside Q β and MS2 show a very different organization to their icosahedral reconstructed densities. Both the gRNA structures have long RNA helical stems and also long-range base pairs. Therefore, it has a large possibility that the gRNA density of STMV from the icosahedral EM map is a symmetry average conformation. Moreover, the SHAPE analysis of the *in vitro* transcribed STMV

gRNA which is not consistent with the icosahedral density suggests that the STMV gRNA contains long-range base pairing (Athavale et al., 2013). Additional work is required to identify the real gRNA structure after packaging inside STMV.

Different from Q β and MS2, a defined gRNA structure cannot be obtained in both SPMV and STMV with the asymmetry reconstruction. It denotes that the structure of the gRNA of SPMV and STMV may be variant. However, it is hard to imagine that the gRNA has a dynamic structure after it is packed while the volume inside the capsid is relatively small and restrains the gRNA movement. So SPMV and STMV may have several ways to enclose their genome into the capsid and generate multiple conformations of the gRNA. For Q β , only 20% of particles have a defined gRNA conformation inside the capsid (Cui et al., 2017). For SPMV and STMV, the percentages may be even lower. Acquiring more data may solve the problem of multiple conformations. The larger dataset can be separated into more classes which may have a defined structure of gRNA.

One major difference between the ssRNA bacteriophage and the ssRNA satellite plant virus is the pathway of the infection. To infect a host, the ssRNA phage binds to the host pilus using its maturation protein (Mat). The Mat bound with the gRNA is then released from the capsid and brought into the host through the retractile force of the host pilus (Krahn, O'Callaghan, & Paranchych, 1972; Meng et al., 2019). For the satellite plant virus, the infection becomes relatively simple. The virus enters the host through

endocytosis and uncoats to release its gRNA (Navarro, Sanchez-Navarro, & Pallas, 2019). Compared to ssRNA phage, the infection pathway of the satellite virus is relatively passive. The ssRNA phage may require a well-organized gRNA to help the gRNA release from the capsid more efficiently when the host pili retract, but the satellite virus may not need it. Therefore, the satellite virus can pack its gRNA more flexibly resulting in a defined gRNA structure hard to be obtained.

4.4 Materials and Methods

4.4.1 Grid Preparation and Cryo-EM Data Collection

SPMV and STMV were obtained from Dr. Karen-Beth G. Scholthof in the Department of Plant Pathology and Microbiology at Texas A&M University and Dr. Kevin Weeks in the Department of Chemistry at the University of North Carolina at Chapel Hill, respectively. The cryo-EM grids were prepared by applying 3 μ L of the sample solution to a C-flat 2/1 400-mesh Holey Carbon Grid (Protochips) at 22 °C with 100% relative humidity and vitrified using a Vitrobot Mark III (FEI Company).

2565 movie stacks of SPMV were collected by two cryo-electron microscopes. 1031 stacks were recorded under a JEM3200FSC with a field emission gun (JEOL) operated at 300 kV. A nominal magnification of 30,000 \times was used, yielding a subpixel size of 0.608 Å. The other 1534 movie stacks were imaged under an FEI Tecnai F20 cryo-electron microscope with a field emission gun (FEI Company) operated at 200 kV. A nominal magnification of 29,000 \times was used, yielding a subpixel size of 0.625 Å. Both

data were collected on a K2 Summit direct detection camera (Gatan) in the super-resolution electron counting mode (X. Li et al., 2013). Each movie stack has 33 frames with 0.2 s per frame for a total exposure time of 6.6 s.

For STMV, 2365 movie stacks were collected under an FEI Tecnai F20 cryo-electron microscope (TF20, FEI Company) at 200 kV. Data were imaged on a K2 Summit direct detection camera (Gatan) at a nominal magnification of 29,000x in the super-resolution electron counting mode, yielding a subpixel size of 0.625 Å. Each movie stack has 33 frames with 0.2 s per frame for a total exposure time of 6.6 s.

4.4.2 Cryo-EM image processing

The collected super-resolution movie stacks were first binned by 2, then aligned and filtered using MotionCor2 (Zheng et al., 2017). The defocus value of each aligned micrograph was determined using Gctf (K. Zhang, 2016). The particles were semi-automatically picked by e2boxer.py in EMAN2 (G. Tang et al., 2007a). The micrographs of SPMV collected on the JEM3200FSC were downsampled to match the pixel size (1.25 Å) of those collected on the TF20. The 2D and 3D classifications and 3D refinement of the image data were performed in RELION (Scheres, 2012).

After 2D cleaning and 3D classification, 164,825 and 122,784 good particle images were chosen for SPMV and STMV, respectively. The icosahedral symmetry 3D refinements yield the EM maps of SPMV at 3.8-Å resolution and of STMV at 5-Å

resolution. Without symmetry applied, the resolution of the refined map is 7.0 Å for SPMV and 8.2 Å for STMV.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

How an RNA is packaged during virion assembly is not a simple story. To understand the mechanism of the packaging, having a structure is very helpful. This work combines cryo-EM and computational simulation to build atomic models of packaged viral gRNA. In chapter 2, we developed the simulation method, HNMMC, to efficiently model an RNA based on its EM density and applied the method to different kinds of RNA including the viral gRNA from the phage MS2. We then *ab initio* built the gRNA structure from another phage Q β and proposed an assembly model of Q β in chapter 3. In chapter 4, the EM results of the small plant viruses show that the viral RNA may be packaged in multiple conformations.

The HNMMC simulation implemented in MOSAICS-EM is a great tool for refining an RNA into the EM density at a middle resolution (5-10 Å). Even if the density has some missing regions, the customized degree of freedom defined by domains or secondary structures can restrain the conformation, so the structure at the missing region cannot be disrupted. Moreover, the computation time of refinement is less in MOSAICS-EM. However, MOSAICS-EM has one weakness that memory usage is not efficient. For short RNA fragments (200-400 nts), the memory requirement is a few hundred megabytes (MB), but the memory requirement increases to tens of gigabytes (GB) when the RNA length reaches 4,000 nts. The large memory requirement limits the application

of MOSAICS-EM. For example, the full-length MS2 and Q β gRNA can be only refined on a workstation or a cluster with big memory. If one virus has a larger genome size than MS2 or Q β , MOSAICS-EM will have problems simulating it. Therefore, some modification has to be done in MOSAICS-EM to optimize memory usage.

The Q β gRNA model built by HNMMC simulation shows 33 “operator-like” RNA stem-loops. It shows that the stem-loops of the Q β genome has more than 40% (33/77) are like PSs. Q β uses a large number of high-affinity PSs to attract its CP for assembly. These PSs may also provide a strong restriction for the Q β assembly and affect the formation of a defined RNA structure. Also, the binding affinity of CP to PS for phages is in a low nanomolar regime (2.5-0.4 nM) (Witherell, Gott, & Uhlenbeck, 1991; Witherell, Wu, & Uhlenbeck, 1990), which is 10 to 100-fold higher than the PS-CP interaction for CCMV (Bruinsma et al., 2016) or HIV-1 (Comas-Garcia et al., 2017). For such strong PS-CP interaction and many PSs in phages, one explanation could be the nature of the host. Because phages infect bacteria, replication and assembly of phages do not occur in specific intracellular compartments (de Castro, Volonte, & Risco, 2013), which means that the phage CP has to identify the viral RNA in an environment with a high concentration of cellular RNAs. However, except for the operator which has been largely studied (Rumnieks & Tars, 2014), the binding of other PSs with CP still needs to be evaluated, and also the biological function of these PSs in the assembly process is very interesting.

In MS2, alternative RNA-CP interactions are resulting in two different dominant packaged gRNA structures. The same secondary structure in the two structures shows that the secondary structure of gRNA should be very stable, but the tertiary structure can be variant. However, multiple gRNA conformations cannot be found in Q β . Following the same image processing procedure as that applied in MS2, the focused classification only gave us one defined Q β gRNA structure. Compared to the MS2 gRNA, the Q β gRNA has more long-range tertiary interactions including the long structure that spans the equator of the capsid. This long helical structure is formed by two RNA kissing-loop interactions which may be important to stabilize the entire packaged gRNA structure. The mutation experiments of the bases at the kissing-loops can be applied to test the importance of these specific tertiary interactions in the Q β assembly. More long-range interactions may also mean that the packaging of Q β gRNA is harder. It may explain why only 30% Q β particles have a defined structure of the gRNA with a maturation protein-bound (Cui et al., 2017), while the other 70% particles that should have the same RNA have variant or probably wrong packaged RNA conformations.

From the EM images, we found that the Q β CPs can form different capsid morphologies, small prolate, oblate, $T=3$, prolate, and $T=4$. Interestingly, all the EM maps of these VLPs with different capsid conformations have RNA densities attached to the capsid (Figure V-1). Because of the specific interaction between the viral PS and CP, presumably, most of these packaged RNAs inside the capsids should be viral gRNA even if the capsid conformations are different. However, considering the internal volume

of each VLP (Table V-1), although the two smaller VLPs, small prolate and oblate, both have space to accommodate the full-length Q β gRNA, the packaging of gRNA in these two VLPs should be very difficult. The full-length gRNA in both VLPs will be very compact resulting in a large electrostatic repulsion, especially for the small-prolate VLP whose internal volume is almost the same as the volume of the entire genome. It has a large possibility that the RNAs inside these two VLPs are degraded viral gRNAs. However, host RNAs can also be packed inside these VLPs because, at the interior-surface of the capsid, the Q β CP has a positively charged loop, SRNRK, that can neutralize the negatively charged RNA. These different forms of VLPs should be separated and analyzed to know what kinds of RNA are packaged in them.

Table V-1. The internal volumes of different VLPs

The volume of Q β genome: $\sim 3,000 \text{ nm}^3$

Conformation	Volume (nm³)
Small prolate	3,400
Oblate	3,800
<i>T</i> =3	5,300
Prolate	6,850
<i>T</i> =4	8,750

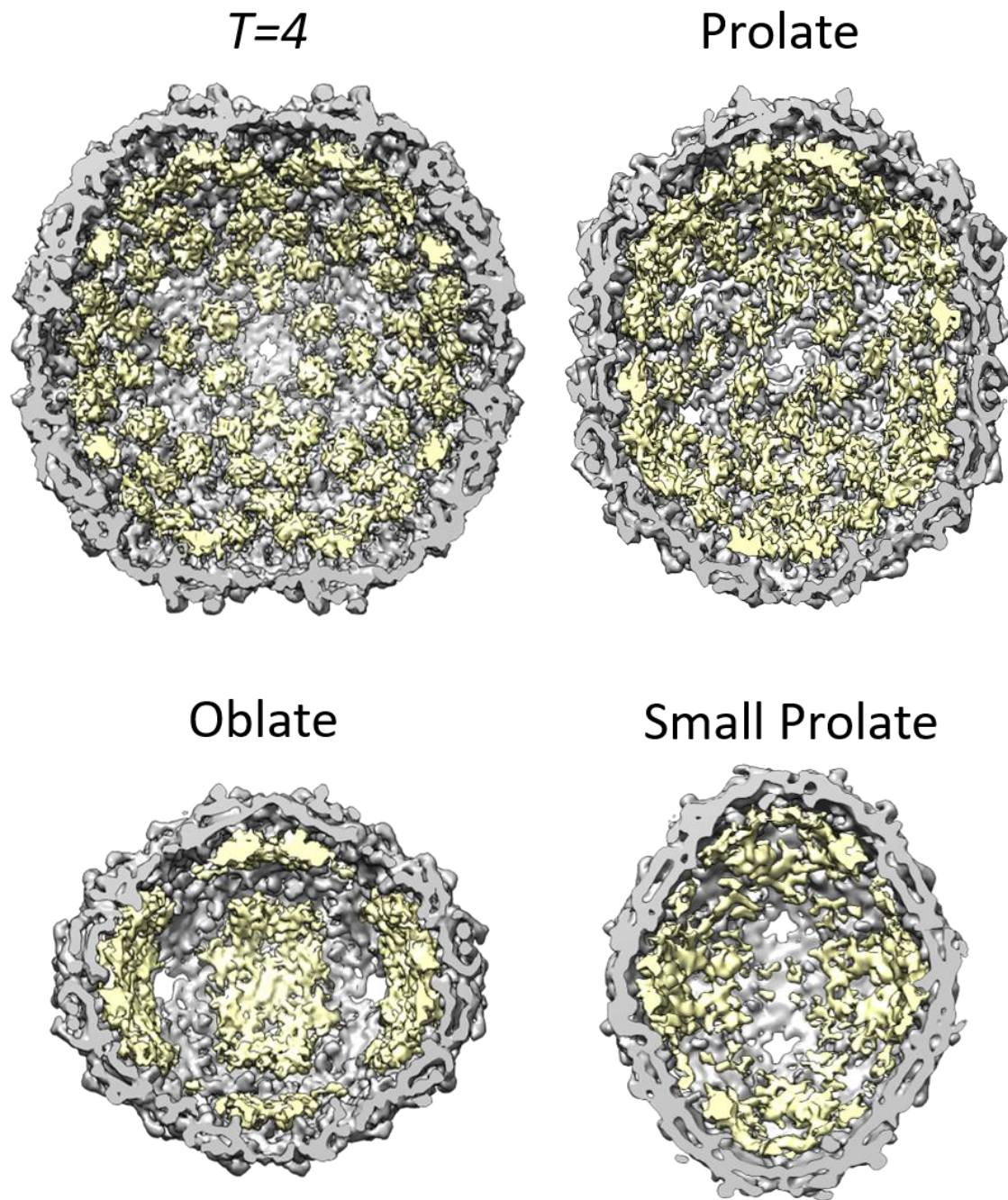


Figure V-1. The RNA densities inside different forms of VLPs
Cut-through views of the EM maps of different VLPs with symmetry applied. The density of capsid is color grey. The extra density inside the capsid is colored yellow.

In addition, the RNA length seems to be related to the formation of different VLPs. When we use only gene *cp/A1* to over-express CP and its read-through protein A1, the formation of oblate and small-prolate particles significantly increases. On *cp/A1*, there are 10 possible PSs to trigger the assembly of VLPs. Moreover, the length of *cp/A1* is only around 1,000 nts which is only one-fourth of full-length RNA. Therefore, the smaller capsid can easily encapsidate the *cp/A1* and the electrostatic repulsion is not severe. This result also shows the restriction of the physical size of the capsid in the RNA packaging, but the clear limitation of the RNA length in each Q β VLP requires further evaluation.

Unfortunately, we cannot obtain a defined gRNA structure in both SPMV and STMV, although some similar features can be found between the classified RNA densities. This result shows that the structure of packaged RNA inside these two small viruses is variant, so more data need to be acquired. For Q β , only 30% of particles have a defined gRNA structure (Cui et al., 2017). For SPMV and STMV, because they do not have to locate the maturation protein, they may have multiple defined gRNA structures but each of them takes only a small percentage.

Besides obtaining a gRNA model for understanding the packaging, the works in this thesis also demonstrate the power of incorporating cryo-EM and molecule modeling in structural virology. The procedures and strategies we used in the thesis can be easily applied to other viruses, such as the phages AP205, PP7, and ϕ Cb5. With more and more

gRNA structures determined, the mechanism of RNA packaging will be more and more clear. The knowledge can help to build virus-based cellular delivery systems for different sizes of RNA cargos in the future.

REFERENCES

- Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G., Rayment, I., Rossmann, M. G., . . . Tsukihara, T. (1980). Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature*, *286*(5768), 33-39. doi:10.1038/286033a0
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, *66*(Pt 2), 213-221. doi:10.1107/S0907444909052925
- Adrian, M., Dubochet, J., Lepault, J., & McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, *308*(5954), 32-36. doi:10.1038/308032a0
- Aksyuk, A. A., & Rossmann, M. G. (2011). Bacteriophage assembly. *Viruses*, *3*(3), 172-203. doi:10.3390/v3030172
- Asensio, M. A., Morella, N. M., Jakobson, C. M., Hartman, E. C., Glasgow, J. E., Sankaran, B., . . . Tullman-Ercek, D. (2016). A Selection for Assembly Reveals That a Single Amino Acid Mutant of the Bacteriophage MS2 Coat Protein Forms a Smaller Virus-like Particle. *Nano Lett.*, *16*(9), 5944-5950. doi:10.1021/acs.nanolett.6b02948
- Athavale, S. S., Gossett, J. J., Bowman, J. C., Hud, N. V., Williams, L. D., & Harvey, S. C. (2013). In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One*, *8*(1), e54384. doi:10.1371/journal.pone.0054384
- Bai, X. C., Rajendra, E., Yang, G., Shi, Y., & Scheres, S. H. (2015). Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife*, *4*. doi:10.7554/eLife.11182
- Baird, N. J., Fang, X. W., Srividya, N., Pan, T., & Sosnick, T. R. (2007). Folding of a universal ribozyme: the ribonuclease P RNA. *Q Rev Biophys*, *40*(2), 113-161. doi:10.1017/S0033583507004623
- Baird, N. J., Gong, H., Zaheer, S. S., Freed, K. F., Pan, T., & Sosnick, T. R. (2010). Extended structures in RNA folding intermediates are due to nonnative interactions rather than electrostatic repulsion. *J Mol Biol*, *397*(5), 1298-1306. doi:10.1016/j.jmb.2010.02.025
- Baird, N. J., Ludtke, S. J., Khant, H., Chiu, W., Pan, T., & Sosnick, T. R. (2010). Discrete structure of an RNA folding intermediate revealed by cryo-electron microscopy. *J Am Chem Soc*, *132*(46), 16352-16353. doi:10.1021/ja107492b

- Baird, N. J., Westhof, E., Qin, H., Pan, T., & Sosnick, T. R. (2005). Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J Mol Biol*, 352(3), 712-722. doi:10.1016/j.jmb.2005.07.010
- Baker, T. S., Olson, N. H., & Fuller, S. D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev*, 63(4), 862-922. doi:10.1128/MMBR.63.4.862-922.1999
- Ban, N., & McPherson, A. (1995). The structure of satellite panicum mosaic virus at 1.9 Å resolution. *Nat Struct Biol*, 2(10), 882-890. doi:10.1038/nsb1095-882
- Bao, Q., Li, X., Han, G., Zhu, Y., Mao, C., & Yang, M. (2019). Phage-based vaccines. *Adv. Drug Deliv. Rev.*, 145, 40-56. doi:10.1016/j.addr.2018.12.013
- Basnak, G., Morton, V. L., Rolfsson, O., Stonehouse, N. J., Ashcroft, A. E., & Stockley, P. G. (2010). Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid. *J Mol Biol*, 395(5), 924-936. doi:10.1016/j.jmb.2009.11.018
- Beekwilder, J. (1996). *Secondary structure of the RNA genome of bacteriophage Q β* . (Doctorate). University of Leiden,
- Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P. A., . . . Spahn, C. M. (2015). Structural snapshots of actively translating human ribosomes. *Cell*, 161(4), 845-857. doi:10.1016/j.cell.2015.03.052
- Borodavka, A., Tuma, R., & Stockley, P. G. (2012). Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. USA*, 109(39), 15769-15774. doi:10.1073/pnas.1204357109
- Borodavka, A., Tuma, R., & Stockley, P. G. (2013). A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol*, 10(4), 481-489. doi:10.4161/rna.23838
- Bruinsma, R. F., Comas-Garcia, M., Garmann, R. F., & Grosberg, A. Y. (2016). Equilibrium self-assembly of small RNA viruses. *Phys Rev E*, 93(3), 032405. doi:10.1103/PhysRevE.93.032405
- Bunka, D. H., Lane, S. W., Lane, C. L., Dykeman, E. C., Ford, R. J., Barker, A. M., . . . Stockley, P. G. (2011). Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. *J Mol Biol*, 413(1), 51-65. doi:10.1016/j.jmb.2011.07.063
- Butan, C., Lokhandwala, P. M., Purdy, J. G., Cardone, G., Craven, R. C., & Steven, A. C. (2010). Suppression of a morphogenic mutant in Rous sarcoma virus capsid

- protein by a second-site mutation: a cryoelectron tomography study. *J Virol*, 84(13), 6377-6386. doi:10.1128/JVI.00207-10
- Butler, P. J. (1972). Structures and roles of the polymorphic forms of tobacco mosaic virus protein. VI. Assembly of the nucleoprotein rods of tobacco mosaic virus from the protein disks and RNA. *J Mol Biol*, 72(1), 25-35. doi:10.1016/0022-2836(72)90065-4
- Butler, P. J., Finch, J. T., & Zimmern, D. (1977). Configuration of tobacco mosaic virus, RNA during virus assembly. *Nature*, 265(5591), 217-219. doi:10.1038/265217a0
- Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L., Knobler, C. M., & Gelbart, W. M. (2012). Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *J Virol*, 86(6), 3318-3326. doi:10.1128/JVI.06566-11
- Callanan, J., Stockdale, S. R., Shkoporov, A., Draper, L. A., Ross, R. P., & Hill, C. (2018). RNA Phage Biology in a Metagenomic Era. *Viruses*, 10(7). doi:10.3390/v10070386
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., . . . Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3, 2. doi:10.1186/1471-2105-3-2
- Cech, T. R., Zaug, A. J., & Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell*, 27(3 Pt 2), 487-496. doi:10.1016/0092-8674(81)90390-1
- Chen, Z. G., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., . . . Johnson, J. E. (1989). Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science*, 245(4914), 154-159. doi:10.1126/science.2749253
- Cheng, C. Y., Chou, F. C., & Das, R. (2015). Modeling complex RNA tertiary folds with Rosetta. *Methods Enzymol*, 553, 35-64. doi:10.1016/bs.mie.2014.10.051
- Chiu, W. (1986). Electron microscopy of frozen, hydrated biological specimens. *Annu Rev Biophys Biophys Chem*, 15, 237-257. doi:10.1146/annurev.bb.15.060186.001321
- Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., & Das, R. (2013). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods*, 10(1), 74-76. doi:10.1038/nmeth.2262

- Cielens, I., Ose, V., Petrovskis, I., Strelnikova, A., Renhofa, R., Kozlovska, T., & Pumpens, P. (2000). Mutilation of RNA phage Qbeta virus-like particles: from icosahedrons to rods. *FEBS Lett.*, *482*(3), 261-264. doi:10.1016/s0014-5793(00)02061-5
- Comas-Garcia, M. (2019). Packaging of Genomic RNA in Positive-Sense Single-Stranded RNA Viruses: A Complex Story. *Viruses*, *11*(3). doi:10.3390/v11030253
- Comas-Garcia, M., Cadena-Nava, R. D., Rao, A. L., Knobler, C. M., & Gelbart, W. M. (2012). In vitro quantification of the relative packaging efficiencies of single-stranded RNA molecules by viral capsid protein. *J Virol*, *86*(22), 12271-12282. doi:10.1128/JVI.01695-12
- Comas-Garcia, M., Datta, S. A., Baker, L., Varma, R., Gudla, P. R., & Rein, A. (2017). Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *Elife*, *6*. doi:10.7554/eLife.27055
- Comas-Garcia, M., Garmann, R. F., Singaram, S. W., Ben-Shaul, A., Knobler, C. M., & Gelbart, W. M. (2014). Characterization of Viral Capsid Protein Self-Assembly around Short Single-Stranded RNA. *J Phys Chem B*, *118*(27), 7510-7519. doi:10.1021/jp503050z
- Comas-Garcia, M., Kroupa, T., Datta, S. A., Harvin, D. P., Hu, W. S., & Rein, A. (2018). Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *Elife*, *7*. doi:10.7554/eLife.38438
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*, 561-563. doi:10.1038/227561a0
- Crothers, D. M., Cole, P. E., Hilbers, C. W., & Shulman, R. G. (1974). The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *J Mol Biol*, *87*(1), 63-88. doi:10.1016/0022-2836(74)90560-9
- Cui, Z., Gorzelnik, K. V., Chang, J. Y., Langlais, C., Jakana, J., Young, R., & Zhang, J. (2017). Structures of Qbeta virions, virus-like particles, and the Qbeta-MurA complex reveal internal coat proteins and the mechanism of host lysis. *Proc. Natl. Acad. Sci. USA*, *114*(44), 11697-11702. doi:10.1073/pnas.1707102114
- Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z. H., & Sun, R. (2017). In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature*, *541*(7635), 112-116. doi:10.1038/nature20589
- Danev, R., & Baumeister, W. (2016). Cryo-EM single particle analysis with the Volta phase plate. *Elife*, *5*. doi:10.7554/eLife.13046

- Danev, R., Tegunov, D., & Baumeister, W. (2017). Using the Volta phase plate with defocus for cryo-EM single particle analysis. *Elife*, 6. doi:10.7554/eLife.23006
- Das, R., Karanicolas, J., & Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, 7(4), 291-294. doi:10.1038/nmeth.1433
- Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., . . . Ourmazd, A. (2014). Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci. USA*, 111(49), 17492-17497. doi:10.1073/pnas.1419276111
- de Castro, I. F., Volonte, L., & Risco, C. (2013). Virus factories: biogenesis and structural design. *Cell Microbiol*, 15(1), 24-34. doi:10.1111/cmi.12029
- de Martin Garrido, N., Crone, M. A., Ramlaul, K., Simpson, P. A., Freemont, P. S., & Aylett, C. H. S. (2020). Bacteriophage MS2 displays unreported capsid variability assembling T = 4 and mixed capsids. *Mol. Microbiol.*, 113(1), 143-152. doi:10.1111/mmi.14406
- Desvoyes, B., & Scholthof, K. B. (2000). RNA: protein interactions associated with satellites of panicum mosaic virus. *FEBS Lett*, 485(1), 25-28. doi:10.1016/s0014-5793(00)02177-3
- Doermann, A. H., Eiserling, F. A., & Boehner, L. (1973). Genetic control of capsid length in bacteriophage T4. I. Isolation and preliminary description of four new mutants. *J Virol*, 12(2), 374-385. doi:10.1128/JVI.12.2.374-385.1973
- Dong, Y., Liu, Y., Jiang, W., Smith, T. J., Xu, Z., & Rossmann, M. G. (2017). Antibody-induced uncoating of human rhinovirus B14. *Proc. Natl. Acad. Sci. USA*, 114(30), 8017-8022. doi:10.1073/pnas.1707369114
- Durham, A. C., & Klug, A. (1971). Polymerization of tobacco mosaic virus protein and its control. *Nat New Biol*, 229(2), 42-46. doi:10.1038/newbio229042a0
- Dykeman, E. C., Stockley, P. G., & Twarock, R. (2010). Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J Mol Biol*, 395(5), 916-923. doi:10.1016/j.jmb.2009.11.016
- Dykeman, E. C., Stockley, P. G., & Twarock, R. (2013a). Building a viral capsid in the presence of genomic RNA. *Phys Rev E Stat Nonlin Soft Matter Phys*, 87(2), 022717. doi:10.1103/PhysRevE.87.022717
- Dykeman, E. C., Stockley, P. G., & Twarock, R. (2013b). Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and

- geometry of the packaged genome. *J Mol Biol*, 425(17), 3235-3249. doi:10.1016/j.jmb.2013.06.005
- Eddy, S. R. (2004). How do RNA folding algorithms work? *Nat Biotechnol*, 22, 1457-1458. doi:10.1038/nbt1104-1457
- Elrad, O. M., & Hagan, M. F. (2008). Mechanisms of size control and polymorphism in viral capsid assembly. *Nano Lett*, 8(11), 3850-3857. doi:10.1021/nl802269a
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat Rev Genet*, 12(12), 861-874. doi:10.1038/nrg3074
- Fenoglio, C., Ridolfi, E., Galimberti, D., & Scarpini, E. (2013). An emerging role for long non-coding RNA dysregulation in neurological disorders. *Int J Mol Sci*, 14(10), 20427-20442. doi:10.3390/ijms141020427
- Ferreira, D., Hernandez, R., Horton, M., & Brown, D. T. (2003). Morphological variants of Sindbis virus produced by a mutation in the capsid protein. *Virology*, 307(1), 54-66. doi:10.1016/s0042-6822(02)00034-x
- Fiedler, J. D., Higginson, C., Hovlid, M. L., Kislukhin, A. A., Castillejos, A., Manzenrieder, F., . . . Finn, M. G. (2012). Engineered mutations change the structure and stability of a virus-like particle. *Biomacromolecules*, 13(8), 2339-2348. doi:10.1021/bm300590x
- Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V., & Stark, H. (2015). Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature*, 520(7548), 567-570. doi:10.1038/nature14275
- Fontana, J., Cardone, G., Heymann, J. B., Winkler, D. C., & Steven, A. C. (2012). Structural changes in Influenza virus at low pH characterized by cryo-electron tomography. *J Virol*, 86(6), 2919-2929. doi:10.1128/JVI.06698-11
- Ford, R. J., Barker, A. M., Bakker, S. E., Coutts, R. H., Ranson, N. A., Phillips, S. E., . . . Stockley, P. G. (2013). Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J Mol Biol*, 425(6), 1050-1064. doi:10.1016/j.jmb.2013.01.004
- Fraenkel-Conrat, H., & Williams, R. C. (1955). Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components. *Proc. Natl. Acad. Sci. USA*, 41(10), 690-698. doi:10.1073/pnas.41.10.690
- Frank, J. (1975). Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy*, 1(2), 159-162. doi:10.1016/s0304-3991(75)80020-9

- Frietze, K. M., Peabody, D. S., & Chackerian, B. (2016). Engineering virus-like particles as vaccine platforms. *Curr. Opin. Virol.*, *18*, 44-49. doi:10.1016/j.coviro.2016.03.001
- Gao, S., Zhang, R., Yu, Z., & Xi, Z. (2012). Antofine analogues can inhibit tobacco mosaic virus assembly through small-molecule-RNA interactions. *Chembiochem*, *13*(11), 1622-1627. doi:10.1002/cbic.201200313
- Garmann, R. F., Comas-Garcia, M., Gopal, A., Knobler, C. M., & Gelbart, W. M. (2014). The assembly pathway of an icosahedral single-stranded RNA virus depends on the strength of inter-subunit attractions. *J Mol Biol*, *426*(5), 1050-1060. doi:10.1016/j.jmb.2013.10.017
- Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, *3*(6). doi:10.1101/cshperspect.a003533
- Gaussier, H., Yang, Q., & Catalano, C. E. (2006). Building a virus from scratch: assembly of an infectious virus using purified components in a rigorously defined biochemical assay system. *J Mol Biol*, *357*(4), 1154-1166. doi:10.1016/j.jmb.2006.01.013
- Glaeser, R. M. (2007). *Electron crystallography of biological macromolecules*. Oxford ; New York: Oxford University Press.
- Golmohammadi, R., Fridborg, K., Bundule, M., Valegard, K., & Liljas, L. (1996). The crystal structure of bacteriophage Q beta at 3.5 Å resolution. *Structure*, *4*(5), 543-554. doi:10.1016/s0969-2126(96)00060-3
- Golmohammadi, R., Valegard, K., Fridborg, K., & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.*, *234*(3), 620-639. doi:10.1006/jmbi.1993.1616
- Gopal, A., Zhou, Z. H., Knobler, C. M., & Gelbart, W. M. (2012). Visualizing large RNA molecules in solution. *RNA*, *18*(2), 284-299. doi:10.1261/rna.027557.111
- Gorzelnik, K. V., Cui, Z., Reed, C. A., Jakana, J., Young, R., & Zhang, J. (2016). Asymmetric cryo-EM structure of the canonical Allovivivirus Qbeta reveals a single maturation protein and the genomic ssRNA in situ. *Proc. Natl. Acad. Sci. USA*, *113*(41), 11519-11524. doi:10.1073/pnas.1609482113
- Grant, T., & Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife*, *4*, e06980. doi:10.7554/eLife.06980

- Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. *J Struct Biol*, 157(1), 117-125. doi:10.1016/j.jsb.2006.05.004
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3), 849-857. doi:10.1016/0092-8674(83)90117-4
- Gytz, H., Mohr, D., Seweryn, P., Yoshimura, Y., Kutlubaeva, Z., Dolman, F., . . . Knudsen, C. R. (2015). Structural basis for RNA-genome recognition during bacteriophage Qbeta replication. *Nucleic Acids Res.*, 43(22), 10893-10906. doi:10.1093/nar/gkv1212
- Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*, 15(8), 509-524. doi:10.1038/nrm3838
- Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature*, 276(5686), 368-373. doi:10.1038/276368a0
- Henderson, R., & Unwin, P. N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, 257(5521), 28-32. doi:10.1038/257028a0
- Hofstetter, H., Monstein, H. J., & Weissmann, C. (1974). The readthrough protein A1 is essential for the formation of viable Q beta particles. *Biochim. Biophys. Acta.*, 374(2), 238-251. doi:10.1016/0005-2787(74)90366-9
- Hohn, T. (1976). Packaging of genomes in bacteriophages: a comparison of ssRNA bacteriophages and dsDNA bacteriophages. *Philos Trans R Soc Lond B Biol Sci*, 276(943), 143-150. doi:10.1098/rstb.1976.0105
- Hu, B., Margolin, W., Molineux, I. J., & Liu, J. (2013). The bacteriophage t7 virion undergoes extensive structural remodeling during infection. *Science*, 339(6119), 576-579. doi:10.1126/science.1231887
- Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat Med*, 21(11), 1253-1261. doi:10.1038/nm.3981
- Jiang, J., Chan, H., Cash, D. D., Miracco, E. J., Ogorzalek Loo, R. R., Upton, H. E., . . . Feigon, J. (2015). Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science*, 350(6260), 4070. doi:10.1126/science.aab4070

- Jossinet, F., Ludwig, T. E., & Westhof, E. (2010). Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, 26(16), 2057-2059. doi:10.1093/bioinformatics/btq321
- Kainov, D. E., Tuma, R., & Mancini, E. J. (2006). Hexameric molecular motors: P4 packaging ATPase unravels the mechanism. *Cell Mol Life Sci*, 63(10), 1095-1105. doi:10.1007/s00018-005-5450-3
- Kappel, K., Liu, S., Larsen, K. P., Skiniotis, G., Puglisi, E. V., Puglisi, J. D., . . . Das, R. (2018). De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature Methods*, 15(11), 947-954. doi:10.1038/s41592-018-0172-2
- Karimi, M., Mirshekari, H., Moosavi Basri, S. M., Bahrami, S., Moghoofei, M., & Hamblin, M. R. (2016). Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. *Adv. Drug Deliv. Rev.*, 106(Pt A), 45-62. doi:10.1016/j.addr.2016.03.003
- Karnik, S., & Billeter, M. (1983). The lysis function of RNA bacteriophage Qbeta is mediated by the maturation (A2) protein. *EMBO J.*, 2(9), 1521-1526. doi:10.1002/j.1460-2075.1983.tb01617.x
- Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., . . . Summers, M. F. (2015). RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*, 348(6237), 917-921. doi:10.1126/science.aaa9266
- Khoshouei, M., Radjainia, M., Baumeister, W., & Danev, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat Commun*, 8, 16099. doi:10.1038/ncomms16099
- Klovins, J., Overbeek, G. P., van den Worm, S. H. E., Ackermann, H. W., & van Duin, J. (2002). Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J. Gen. Virol.*, 83(Pt 6), 1523-1533. doi:10.1099/0022-1317-83-6-1523
- Knapp, B., Demharter, S., Deane, C. M., & Minary, P. (2016). Exploring peptide/MHC detachment processes using hierarchical natural move Monte Carlo. *Bioinformatics*, 32(2), 181-186. doi:10.1093/bioinformatics/btv502
- Kolakofsky, D. (2015). A short biased history of RNA viruses. *RNA*, 21(4), 667-669. doi:10.1261/rna.049916.115
- Koning, R. I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., . . . Koster, A. J. (2016a). Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.*, 7, 12524. doi:10.1038/ncomms12524

- Koning, R. I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., . . . Koster, A. J. (2016b). Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat Commun*, *7*, 12524. doi:10.1038/ncomms12524
- Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, *479-480*, 2-25. doi:10.1016/j.virol.2015.02.039
- Kostyuchenko, V. A., Lim, E. X., Zhang, S., Fibriansah, G., Ng, T. S., Ooi, J. S., . . . Lok, S. M. (2016). Structure of the thermally stable Zika virus. *Nature*, *533(7603)*, 425-428. doi:10.1038/nature17994
- Kotta-Loizou, I., Peyret, H., Saunders, K., Coutts, R. H. A., & Lomonosoff, G. P. (2019). Investigating the Biological Relevance of In Vitro-Identified Putative Packaging Signals at the 5' Terminus of Satellite Tobacco Necrosis Virus 1 Genomic RNA. *J Virol*, *93(9)*. doi:10.1128/JVI.02106-18
- Krahn, P. M., O'Callaghan, R. J., & Paranchych, W. (1972). Stages in phage R17 infection. VI. Injection of A protein and RNA into the host cell. *Virology*, *47(3)*, 628-637. doi:10.1016/0042-6822(72)90552-1
- Krasilnikov, A. S., Yang, X., Pan, T., & Mondragon, A. (2003). Crystal structure of the specificity domain of ribonuclease P. *Nature*, *421(6924)*, 760-764. doi:10.1038/nature01386
- Krawczyk, K., Sim, A. Y., Knapp, B., Deane, C. M., & Minary, P. (2016). Tertiary Element Interaction in HIV-1 TAR. *J Chem Inf Model*, *56(9)*, 1746-1754. doi:10.1021/acs.jcim.6b00152
- Krishna, N. K., Marshall, D., & Schneemann, A. (2003). Analysis of RNA packaging in wild-type and mosaic protein capsids of flock house virus using recombinant baculovirus vectors. *Virology*, *305(1)*, 10-24. doi:10.1006/viro.2002.1740
- Kuznetsov, Y. G., Dowell, J. J., Gavira, J. A., Ng, J. D., & McPherson, A. (2010). Biophysical and atomic force microscopy characterization of the RNA from satellite tobacco mosaic virus. *Nucleic Acids Res*, *38(22)*, 8284-8294. doi:10.1093/nar/gkq662
- Larman, B. C., Dethoff, E. A., & Weeks, K. M. (2017). Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. *Biochemistry*, *56(16)*, 2175-2183. doi:10.1021/acs.biochem.6b01166

- Larson, S. B., Day, J., Greenwood, A., & McPherson, A. (1998). Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J Mol Biol*, 277(1), 37-59. doi:10.1006/jmbi.1997.1570
- Larson, S. B., & McPherson, A. (2001). Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr Opin Struct Biol*, 11(1), 59-65. doi:10.1016/s0959-440x(00)00166-4
- Lee, Y. J., Yi, H., Kim, W. J., Kang, K., Yun, D. S., Strano, M. S., . . . Belcher, A. M. (2009). Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science*, 324(5930), 1051-1055. doi:10.1126/science.1171541
- Leipold, B., & Hofschneider, P. H. (1975). Isolation of an infectious RNA-A-protein complex from the bacteriophage M12. *FEBS Lett.*, 55(1), 50-52. doi:10.1016/0014-5793(75)80954-9
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., . . . Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6), 584-590. doi:10.1038/nmeth.2472
- Li, Z., Ge, X., Zhang, Y., Zheng, L., Sanyal, S., & Gao, N. (2018). Cryo-EM structure of *Mycobacterium smegmatis* ribosome reveals two unidentified ribosomal proteins close to the functional centers. *Protein Cell*, 9(4), 384-388. doi:10.1007/s13238-017-0456-9
- Liao, M., Cao, E., Julius, D., & Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478), 107-112. doi:10.1038/nature12822
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., . . . Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, 75(Pt 10), 861-877. doi:10.1107/S2059798319011471
- Liekina, I., Kalnins, G., Akopjana, I., Bogans, J., Sisovs, M., Jansons, J., . . . Tars, K. (2019). Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data. *J. Nanobiotechnology*, 17(1), 61. doi:10.1186/s12951-019-0497-8
- Liu, D., Geary, C. W., Chen, G., Shao, Y., Li, M., Mao, C., . . . Weizmann, Y. (2020). Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. *Nat. Chem.*, 12(3), 249-259. doi:10.1038/s41557-019-0406-7

- Liu, H., & Cheng, L. (2015). Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science*, *349*(6254), 1347-1350. doi:10.1126/science.aaa4938
- Liu, Y., Sheng, J., van Vliet, A. L. W., Buda, G., van Kuppeveld, F. J. M., & Rossmann, M. G. (2018). Molecular basis for the acid-initiated uncoating of human enterovirus D68. *Proc. Natl. Acad. Sci. USA*, *115*(52), E12209-E12217. doi:10.1073/pnas.1803347115
- Liu, Z., Gutierrez-Vargas, C., Wei, J., Grassucci, R. A., Sun, M., Espina, N., . . . Frank, J. (2017). Determination of the ribosome structure to a resolution of 2.5 Å by single-particle cryo-EM. *Protein Sci*, *26*(1), 82-92. doi:10.1002/pro.3068
- Lorenz, R., Bernhart, S. H., Honer Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms Mol Biol*, *6*, 26. doi:10.1186/1748-7188-6-26
- Losdorfer Bozic, A., & Podgornik, R. (2018). Varieties of charge distributions in coat proteins of ssRNA+ viruses. *J Phys Condens Matter*, *30*(2), 024001. doi:10.1088/1361-648X/aa9ded
- Loughrey, D., Watters, K. E., Settle, A. H., & Lucks, J. B. (2014). SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res*, *42*(21), e165-e165. doi:10.1093/nar/gku909
- Lu, Z., & Chang, H. Y. (2016). Decoding the RNA structurome. *Curr Opin Struct Biol*, *36*, 142-148. doi:10.1016/j.sbi.2016.01.007
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., . . . Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, *47*(W1), W636-W641. doi:10.1093/nar/gkz268
- Makino, D. L., Day, J., Larson, S. B., & McPherson, A. (2006). Investigation of RNA structure in satellite panicum mosaic virus. *Virology*, *351*(2), 420-431. doi:10.1016/j.virol.2006.03.028
- Mandelkow, E., Holmes, K. C., & Gallwitz, U. (1976). A new helical aggregate of tobacco mosaic virus protein. *J Mol Biol*, *102*(2), 265-285. doi:10.1016/s0022-2836(76)80053-8
- Mandelkow, E., Stubbs, G., & Warren, S. (1981). Structures of the helical aggregates of tobacco mosaic virus protein. *J Mol Biol*, *152*(2), 375-386. doi:10.1016/0022-2836(81)90248-5

- Mastrorade, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.*, *152*(1), 36-51.
doi:10.1016/j.jsb.2005.07.007
- Mathews, D. H., Moss, W. N., & Turner, D. H. (2010). Folding and finding RNA secondary structure. *Cold Spring Harbor perspectives in biology*, *2*.
doi:10.1101/cshperspect.a003665
- Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, *288*(5), 911-940. doi:10.1006/jmbi.1999.2700
- Mathews, D. H., & Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol*, *16*(3), 270-278.
doi:10.1016/j.sbi.2006.05.010
- McGreevy, R., Teo, I., Singharoy, A., & Schulten, K. (2016). Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods*, *100*, 50-60.
doi:10.1016/j.ymeth.2016.01.009
- McMullan, G., Faruqi, A. R., Clare, D., & Henderson, R. (2014). Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy*, *147*, 156-163.
doi:10.1016/j.ultramic.2014.08.002
- Meng, R., Jiang, M., Cui, Z., Chang, J. Y., Yang, K., Jakana, J., . . . Zhang, J. (2019). Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nat. Commun.*, *10*(1), 3130. doi:10.1038/s41467-019-11126-8
- Minary, P., & Levitt, M. (2010). Conformational optimization with natural degrees of freedom: a novel stochastic chain closure algorithm. *J Comput Biol*, *17*(8), 993-1010. doi:10.1089/cmb.2010.0016
- Minary, P., & Levitt, M. (2014). Training-free atomistic prediction of nucleosome occupancy. *Proc. Natl. Acad. Sci. USA*, *111*(17), 6293-6298.
doi:10.1073/pnas.1404475111
- Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, *142*(3), 334-347.
doi:10.1016/s1047-8477(03)00069-8
- Mirkov, T. E., Mathews, D. M., Du Plessis, D. H., & Dodds, J. A. (1989). Nucleotide sequence and translation of satellite tobacco mosaic virus RNA. *Virology*, *170*(1), 139-146. doi:10.1016/0042-6822(89)90361-9

- Mishra, S., Ahmed, T., Tyagi, A., Shi, J., & Bhushan, S. (2018). Structures of Mycobacterium smegmatis 70S ribosomes in complex with HPF, tmRNA, and P-tRNA. *Sci Rep*, 8(1), 13587. doi:10.1038/s41598-018-31850-3
- Moriya, T., Saur, M., Stabrin, M., Merino, F., Voicu, H., Huang, Z., . . . Gatsogiannis, C. (2017). High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *J. Vis. Exp.*(123). doi:10.3791/55448
- Morton, V. L., Dykeman, E. C., Stonehouse, N. J., Ashcroft, A. E., Twarock, R., & Stockley, P. G. (2010). The impact of viral RNA on assembly pathway selection. *J Mol Biol*, 401(2), 298-308. doi:10.1016/j.jmb.2010.05.059
- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., . . . Scheres, S. H. W. (2020). Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832), 152-156. doi:10.1038/s41586-020-2829-0
- Nam, K. T., Kim, D. W., Yoo, P. J., Chiang, C. Y., Meethong, N., Hammond, P. T., . . . Belcher, A. M. (2006). Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science*, 312(5775), 885-888. doi:10.1126/science.1122716
- Navarro, J. A., Sanchez-Navarro, J. A., & Pallas, V. (2019). Key checkpoints in the movement of plant viruses through the host. *Adv Virus Res*, 104, 1-64. doi:10.1016/bs.aivir.2019.05.001
- Nguyen, T. H., Galej, W. P., Bai, X. C., Savva, C. G., Newman, A. J., Scheres, S. H., & Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, 523(7558), 47-52. doi:10.1038/nature14548
- Onoa, B., & Tinoco, I. (2004). RNA folding and unfolding. *Current Opinion in Structural Biology*, 14, 374-379. doi:10.1016/j.sbi.2004.04.001
- Ortega-Esteban, A., Mata, C. P., Rodriguez-Espinosa, M. J., Luque, D., Irigoyen, N., Rodriguez, J. M., . . . Caston, J. R. (2020). Cryo-EM structure, assembly, and mechanics show morphogenesis and evolution of human picobirnavirus. *J Virol*. doi:10.1128/JVI.01542-20
- Patel, N., White, S. J., Thompson, R. F., Bingham, R., Weiss, E. U., Maskell, D. P., . . . Stockley, P. G. (2017). HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nat Microbiol*, 2, 17098. doi:10.1038/nmicrobiol.2017.98
- Peabody, D. S. (1997). Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2. *Mol Gen Genet*, 254(4), 358-364. doi:10.1007/s004380050427

- Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, *92*(11), 3817-3829. doi:10.1529/biophysj.106.097782
- Perlmutter, J. D., & Hagan, M. F. (2015). Mechanisms of virus assembly. *Annu Rev Phys Chem*, *66*, 217-239. doi:10.1146/annurev-physchem-040214-121637
- Polikanov, Y. S., Blaha, G. M., & Steitz, T. A. (2012). How hibernation factors RMF, HPF, and YfiA turn off protein synthesis. *Science*, *336*(6083), 915-918. doi:10.1126/science.1218538
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290-296. doi:10.1038/nmeth.4169
- Qin, H., Sosnick, T. R., & Pan, T. (2001). Modular construction of a tertiary RNA structure: the specificity domain of the *Bacillus subtilis* RNase P RNA. *Biochemistry*, *40*(37), 11202-11210. doi:10.1021/bi010076n
- Qu, F., & Morris, T. J. (1997). Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J Virol*, *71*(2), 1428-1435. doi:10.1128/JVI.71.2.1428-1435.1997
- Rao, A. L. (2006). Genome packaging by spherical plant RNA viruses. *Annu Rev Phytopathol*, *44*, 61-87. doi:10.1146/annurev.phyto.44.070505.143334
- Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*, 129. doi:10.1186/1471-2105-11-129
- Rohovie, M. J., Nagasawa, M., & Swartz, J. R. (2017). Virus-like particles: Next-generation nanoparticles for targeted therapeutic delivery. *Bioeng. Transl. Med.*, *2*(1), 43-57. doi:10.1002/btm2.10049
- Rolfsson, O., Middleton, S., Manfield, I. W., White, S. J., Fan, B., Vaughan, R., . . . Stockley, P. G. (2016). Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *J Mol Biol*, *428*(2 Pt B), 431-448. doi:10.1016/j.jmb.2015.11.014
- Rossmann, M. G. (2013). Structure of viruses: a short history. *Q Rev Biophys*, *46*(2), 133-180. doi:10.1017/S0033583513000012
- Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., . . . et al. (1985). Structure of a human common cold virus and

- functional relationship to other picornaviruses. *Nature*, 317(6033), 145-153.
doi:10.1038/317145a0
- Rother, M., Rother, K., Puton, T., & Bujnicki, J. M. (2011). ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res*, 39(10), 4007-4022. doi:10.1093/nar/gkq1320
- Routh, A., Domitrovic, T., & Johnson, J. E. (2012). Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc. Natl. Acad. Sci. USA*, 109(6), 1907-1912. doi:10.1073/pnas.1116168109
- Rumnieks, J., & Tars, K. (2012). Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. *BMC Microbiol.*, 12, 277. doi:10.1186/1471-2180-12-277
- Rumnieks, J., & Tars, K. (2014). Crystal structure of the bacteriophage Qbeta coat protein in complex with the RNA operator of the replicase gene. *J Mol Biol*, 426(5), 1039-1049. doi:10.1016/j.jmb.2013.08.025
- Ruskin, R. S., Yu, Z., & Grigorieff, N. (2013). Quantitative characterization of electron detectors for transmission electron microscopy. *J Struct Biol*, 184(3), 385-393. doi:10.1016/j.jsb.2013.10.016
- Saxena, P., & Lomonosoff, G. P. (2014). Virus infection cycle events coupled to RNA replication. *Annu. Rev. Phytopathol.*, 52, 197-212. doi:10.1146/annurev-phyto-102313-050205
- Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, 180(3), 519-530. doi:10.1016/j.jsb.2012.09.006
- Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., . . . Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102(5), 615-623. doi:10.1016/S0092-8674(00)00084-2
- Scholthof, K. B., Jones, R. W., & Jackson, A. O. (1999). Biology and structure of plant satellite viruses activated by icosahedral helper viruses. *Curr Top Microbiol Immunol*, 239, 123-143. doi:10.1007/978-3-662-09796-0_7
- Schroeder, S. J., Stone, J. W., Bleckley, S., Gibbons, T., & Mathews, D. M. (2011). Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophys J*, 101(1), 167-175. doi:10.1016/j.bpj.2011.05.053

- Seetin, M. G., & Mathews, D. H. (2012). *RNA Structure Prediction: An Overview of Methods* (Vol. 905). Totowa, NJ: Humana Press.
- Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., . . . Skiniotis, G. (2016). 2.8-Å Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite *Leishmania*. *Cell Rep*, *16*(2), 288-294. doi:10.1016/j.celrep.2016.06.014
- Shi, H., & Moore, P. B. (2000). The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA*, *6*(8), 1091-1105. doi:10.1017/S1355838200000364
- Shiba, T., & Suzuki, Y. (1981). Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochim. Biophys. Acta.*, *654*(2), 249-255. doi:10.1016/0005-2787(81)90179-9
- Sim, A. Y., Levitt, M., & Minary, P. (2012). Modeling and design by hierarchical natural moves. *Proc. Natl. Acad. Sci. USA*, *109*(8), 2890-2895. doi:10.1073/pnas.1119918109
- Simon, A. E., Roossinck, M. J., & Havelda, Z. (2004). Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu Rev Phytopathol*, *42*, 415-437. doi:10.1146/annurev.phyto.42.040803.140402
- Singharoy, A., Teo, I., McGreevy, R., Stone, J. E., Zhao, J., & Schulten, K. (2016). Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife*, *5*. doi:10.7554/eLife.16105
- Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T. C., Rossmann, M. G., & Kuhn, R. J. (2016). The 3.8 Å resolution cryo-EM structure of Zika virus. *Science*, *352*(6284), 467-470. doi:10.1126/science.aaf5316
- Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., & Weeks, K. M. (2015). Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc*, *10*(11), 1643-1669. doi:10.1038/nprot.2015.103
- Sorger, P. K., Stockley, P. G., & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J Mol Biol*, *191*(4), 639-658. doi:10.1016/0022-2836(86)90451-1
- Sougrat, R., Bartesaghi, A., Lifson, J. D., Bennett, A. E., Bess, J. W., Zabransky, D. J., & Subramaniam, S. (2007). Electron tomography of the contact between T cells and SIV/HIV-1: implications for viral entry. *PLoS Pathog*, *3*(5), e63. doi:10.1371/journal.ppat.0030063

- Stewart, H., Bingham, R. J., White, S. J., Dykeman, E. C., Zothner, C., Tuplin, A. K., . . . Harris, M. (2016). Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Sci Rep*, 6, 22952. doi:10.1038/srep22952
- Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., . . . Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J. Mol. Biol.*, 369(2), 541-552. doi:10.1016/j.jmb.2007.03.020
- Stockley, P. G., Twarock, R., Bakker, S. E., Barker, A. M., Borodavka, A., Dykeman, E., . . . Tuma, R. (2013). Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys.*, 39(2), 277-287. doi:10.1007/s10867-013-9313-0
- Stoltenburg, R., Reinemann, C., & Strehlitz, B. (2007). SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng*, 24(4), 381-403. doi:10.1016/j.bioeng.2007.06.001
- Tan, Y. Z., Aiyer, S., Mietzsch, M., Hull, J. A., McKenna, R., Grieger, J., . . . Lyumkis, D. (2018). Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. *Nat Commun*, 9(1), 3628. doi:10.1038/s41467-018-06076-6
- Tan, Y. Z., Cheng, A., Potter, C. S., & Carragher, B. (2016). Automated data collection in single particle electron microscopy. *Microscopy (Oxf)*, 65(1), 43-56. doi:10.1093/jmicro/dfv369
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007a). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.*, 157(1), 38-46. doi:10.1016/j.jsb.2006.05.009
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007b). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*, 157(1), 38-46. doi:10.1016/j.jsb.2006.05.009
- Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M., & Johnson, J. E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat Struct Biol*, 8(1), 77-83. doi:10.1038/83089
- Tars, K. (2020). ssRNA Phages: Life Cycle, Structure and Applications. In G. Witzany (Ed.), *Biocommunication of Phages* (pp. 261-292). Cham: Springer International Publishing.

- Tars, K., Fridborg, K., Bundule, M., & Liljas, L. (2000). The three-dimensional structure of bacteriophage PP7 from *Pseudomonas aeruginosa* at 3.7-Å resolution. *Virology*, *272*(2), 331-337. doi:10.1006/viro.2000.0373
- Taylor, K. A., & Glaeser, R. M. (1976). Electron microscopy of frozen hydrated biological specimens. *J Ultrastruct Res*, *55*(3), 448-456. doi:10.1016/s0022-5320(76)80099-8
- Team, R. C. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Tomita, K., Ichihashi, N., & Yomo, T. (2015). Replication of partial double-stranded RNAs by Qbeta replicase. *Biochem. Biophys. Res. Commun.*, *467*(2), 293-296. doi:10.1016/j.bbrc.2015.09.169
- Turina, M., Desvoyes, B., & Scholthof, K. B. (2000). A gene cluster encoded by panicum mosaic virus is associated with virus movement. *Virology*, *266*(1), 120-128. doi:10.1006/viro.1999.0069
- Turina, M., Maruoka, M., Monis, J., Jackson, A. O., & Scholthof, K. B. (1998). Nucleotide sequence and infectivity of a full-length cDNA clone of panicum mosaic virus. *Virology*, *241*(1), 141-155. doi:10.1006/viro.1997.8939
- Turner, D. R., & Butler, P. J. (1986). Essential features of the assembly origin of tobacco mosaic virus RNA as studied by directed mutagenesis. *Nucleic Acids Res*, *14*(23), 9229-9242. doi:10.1093/nar/14.23.9229
- Uchida, S., & Dimmeler, S. (2015). Long noncoding RNAs in cardiovascular diseases. *Circ Res*, *116*(4), 737-750. doi:10.1161/CIRCRESAHA.116.302521
- Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J., & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*, *371*(6498), 623-626. doi:10.1038/371623a0
- van Duin, J., & Tsareva, N. (2006). Single-stranded RNA phages. In R. Calendar (Ed.), *The Bacteriophages* (Second ed., pp. 175–196). New York: Oxford University Press.
- Wang, R. Y., Song, Y., Barad, B. A., Cheng, Y., Fraser, J. S., & DiMaio, F. (2016). Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife*, *5*. doi:10.7554/eLife.17219
- Watanabe, H., & Watanabe, M. (1970). Comparative biology of five RNA phages, R23, f2, Q beta, R34, and R40. *Can. J. Microbiol.*, *16*(9), 859-864. doi:10.1139/m70-145

- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. doi:10.1093/bioinformatics/btp033
- Weiner, A. M., & Weber, K. (1971). Natural read-through at the UGA termination signal of Q-beta coat protein cistron. *Nat. New Biol.*, *234*(50), 206-209. doi:10.1038/newbio234206a0
- White, J. M. (1990). Viral and cellular membrane fusion proteins. *Annu Rev Physiol*, *52*, 675-697. doi:10.1146/annurev.ph.52.030190.003331
- Witherell, G. W., Gott, J. M., & Uhlenbeck, O. C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol*, *40*, 185-220. doi:10.1016/s0079-6603(08)60842-9
- Witherell, G. W., Wu, H. N., & Uhlenbeck, O. C. (1990). Cooperative binding of R17 coat protein to RNA. *Biochemistry*, *29*(50), 11051-11057. doi:10.1021/bi00502a006
- Woodson, S. A. (2010). Compact intermediates in RNA folding. *Annu Rev Biophys*, *39*, 61-77. doi:10.1146/annurev.biophys.093008.131334
- Wriggers, W. (2012). Conventions and workflows for using Situs. *Acta Crystallogr. D Biol. Crystallogr.*, *68*(Pt 4), 344-351. doi:10.1107/S09074444911049791
- Xu, J., Gui, M., Wang, D., & Xiang, Y. (2016). The bacteriophage varphi29 tail possesses a pore-forming loop for cell membrane penetration. *Nature*, *534*(7608), 544-547. doi:10.1038/nature18017
- Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C., & Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*, *349*(6253), 1182-1191. doi:10.1126/science.aac7629
- Yang, K., Chang, J. Y., Cui, Z., Li, X., Meng, R., Duan, L., . . . Zhang, J. (2017). Structural insights into species-specific features of the ribosome from the human pathogen *Mycobacterium tuberculosis*. *Nucleic Acids Res*, *45*(18), 10884-10894. doi:10.1093/nar/gkx785
- Yip, K. M., Fischer, N., Paknia, E., Chari, A., & Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature*, *587*(7832), 157-161. doi:10.1038/s41586-020-2833-4
- Zeltins, A. (2013). Construction and characterization of virus-like particles: a review. *Mol. Biotechnol.*, *53*(1), 92-107. doi:10.1007/s12033-012-9598-4

- Zeng, Y., Larson, S. B., Heitsch, C. E., McPherson, A., & Harvey, S. C. (2012). A model for the structure of satellite tobacco mosaic virus. *J Struct Biol*, *180*(1), 110-116. doi:10.1016/j.jsb.2012.06.008
- Zhang, J., Minary, P., & Levitt, M. (2012). Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc. Natl. Acad. Sci. USA*, *109*(25), 9845-9850. doi:10.1073/pnas.1205945109
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol*, *193*(1), 1-12. doi:10.1016/j.jsb.2015.11.003
- Zhang, K., Keane, S. C., Su, Z., Irobalieva, R. N., Chen, M., Van, V., . . . Chiu, W. (2018). Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. *Structure*, *26*(3), 490-498 e493. doi:10.1016/j.str.2018.01.001
- Zhang, X., Ding, K., Yu, X., Chang, W., Sun, J., & Zhou, Z. H. (2015). In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature*, *527*(7579), 531-534. doi:10.1038/nature15767
- Zhang, X., Lai, M., Chang, W., Yu, I., Ding, K., Mrazek, J., . . . Zhou, Z. H. (2016). Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. *Nat Commun*, *7*, 13223. doi:10.1038/ncomms13223
- Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., & Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA*, *105*(6), 1867-1872. doi:10.1073/pnas.0711623105
- Zhao, L., Kopylov, M., Potter, C. S., Carragher, B., & Finn, M. G. (2019). Engineering the PP7 Virus Capsid as a Peptide Display Platform. *ACS Nano*, *13*(4), 4443-4454. doi:10.1021/acsnano.8b09683
- Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, *14*(4), 331-332. doi:10.1038/nmeth.4193
- Zhong, Q., Carratala, A., Nazarov, S., Guerrero-Ferreira, R. C., Piccinini, L., Bachmann, V., . . . Kohn, T. (2016). Genetic, Structural, and Phenotypic Properties of MS2 Coliphage with Resistance to ClO₂ Disinfection. *Environ Sci Technol*, *50*(24), 13520-13528. doi:10.1021/acs.est.6b04170

- Zhu, L., Sun, Y., Fan, J., Zhu, B., Cao, L., Gao, Q., . . . Wang, X. (2018). Structures of Coxsackievirus A10 unveil the molecular mechanisms of receptor binding and viral uncoating. *Nat Commun*, 9(1), 4985. doi:10.1038/s41467-018-07531-0
- Zimmern, D. (1977). The nucleotide sequence at the origin for assembly on tobacco mosaic virus RNA. *Cell*, 11(3), 463-482. doi:10.1016/0092-8674(77)90065-4
- Zimmern, D., & Butler, P. J. (1977). The isolation of tobacco mosaic virus RNA fragments containing the origin for viral assembly. *Cell*, 11(3), 455-462. doi:10.1016/0092-8674(77)90064-2

APPENDIX A

THE STRUCTURE OF THE HOLIN-ANTI-HOLIN COMPLEX

A.1 Introduction

Lysis is the last step in the infection cycle of a lytic bacteriophage (phage) that causes the infected cell to burst. For double-strand DNA (dsDNA) phages, lysis is controlled by holins, a diverse group of small membrane proteins, which harmlessly accumulate in the inner membrane in the beginning. At a precise genetically programmed time, holins suddenly form fatal holes in the inner membrane. These holes disrupt the cell energy metabolism and allow endolysins which can cleave the bacterial cell wall to cross and initiate lysis (Wang, Smith, & Young, 2000; R. Young, 2014).

In bacteriophage T4, the length of the infection cycle is controlled by holin T, which forms holes in a transition from monomers to dimers to oligomers and finally to protein rafts (Moussa, Lawler, & Young, 2014). The size and topology of T are different from those of most holins. While most holins consist of around 60~120 amino acids, T has 218 amino acids which is unusually large (Grundling, Blasi, & Young, 2000a, 2000b; Park, Struck, Deaton, & Young, 2006; Ramanculov & Young, 2001). Based on their structural topology, holins can be separated into at least three classes: I, II, and III (Tran, Struck, & Young, 2005). Most holins belong to Class I and II, which contain three or two transmembrane domains (TMDs) respectively. T belongs to Class III, which has only one TMD with N-terminus in the cytoplasm and C-terminus in the periplasm

(Figure 1). T has a cytoplasmic domain of 34 amino acids, a TMD of 21 amino acids, and a periplasmic domain of 163 amino acids. Mutagenesis shows that the cytoplasmic domain might be related to the timing of hole formation and the structure of the holes and the periplasmic domain is important to the oligomerization of T to form the hole. The periplasmic domain of T also can directly interact with the T regulatory protein, antiholin RI. RI can inhibit the T oligomerization and delays lysis (Moussa, Kuznetsov, Tran, Sacchettini, & Young, 2012; Moussa et al., 2014).

RI is an inner membrane protein and has one TMD that can spontaneously escape from the membrane as a signal-anchor-release domain without proteolytic processing (Tran, Struck, & Young, 2007). Previous studies show that the soluble periplasmic domain of T (sT) and the soluble periplasmic domain of RI (sRI) can form a heterodimeric complex (Moussa et al., 2012). sRI is more active compared to the full-length RI in inhibiting lysis *in vivo*. Moreover, secreting sT to the periplasm can partially inhibit the function of RI (Tran et al., 2005).

The delayed lysis caused by RI is called lysis inhibition (LIN). Normally, after infection, T4 lyses *Escherichia coli* (*E. coli*) cells within around 25 minutes (Mukai, Streisinger, & Miller, 1967). However, if a T4 phage infected cell gets co-infected with other T4-like phages (superinfection), the lysis can be delayed for several hours (Doermann, 1948; Dressman & Drake, 1999). At superinfection, the central tail tube of the second infecting phage can penetrate the outer membrane but cannot puncture the

inner membrane. So, the DNA and protein molecules in the second phage can only be injected into the host periplasm. These substances by the fruitless injection may activate RI to inhibit the oligomerization of T, blocking the formation of membrane holes and delaying lysis (Tran et al., 2005). The LIN state is unstable and can eventually be terminated by lysis. After LIN, around 100-fold more progeny virions can be released, as compared to normal lysis state (Doermann, 1948).

Genetic studies have revealed the key players of lysis in the phage T4. However, due to the absence of structure information, the interactions between T-T and T-RI in the T4 lysis system are still not unknown, and the modulation of these interactions is not clear. Therefore, we used single-particle cryo-electron microscopy (cryo-EM) to solve the structure of the holin-antiholin complex, the sT-sRI complex. Although the structure is unlikely to be physiologically relevant, it still shows that in solution, besides heterotetramers, the sT and sRI can form a higher-order complex that consists of four sT-sRI heterotetramers.

A.2 Results and Discussion

A.2.1 Single-particle cryo-EM structure of the sT-sRI complex

Using single-particle cryo-EM, 508 movie stacks of the sT-sRI complex were collected, yielding 35,432 particles. A symmetry-free reconstruction of these particles generated a map of the sT-sRI complex at 9.4-Å resolution (Figure A-1A). Four sT-sRI heterotetramers (PDB ID: 6PXE) can be fitted into the map. The structure is a circle in

the top view, but it is like a small helical rise in the side view. The sRI at the top right of the first heterotetramer binds to the sRI at the bottom left of the fourth heterotetramer and four tetramers form a circle. The long helix at the N-terminal of each sT interacts with each other at the middle of the circle.

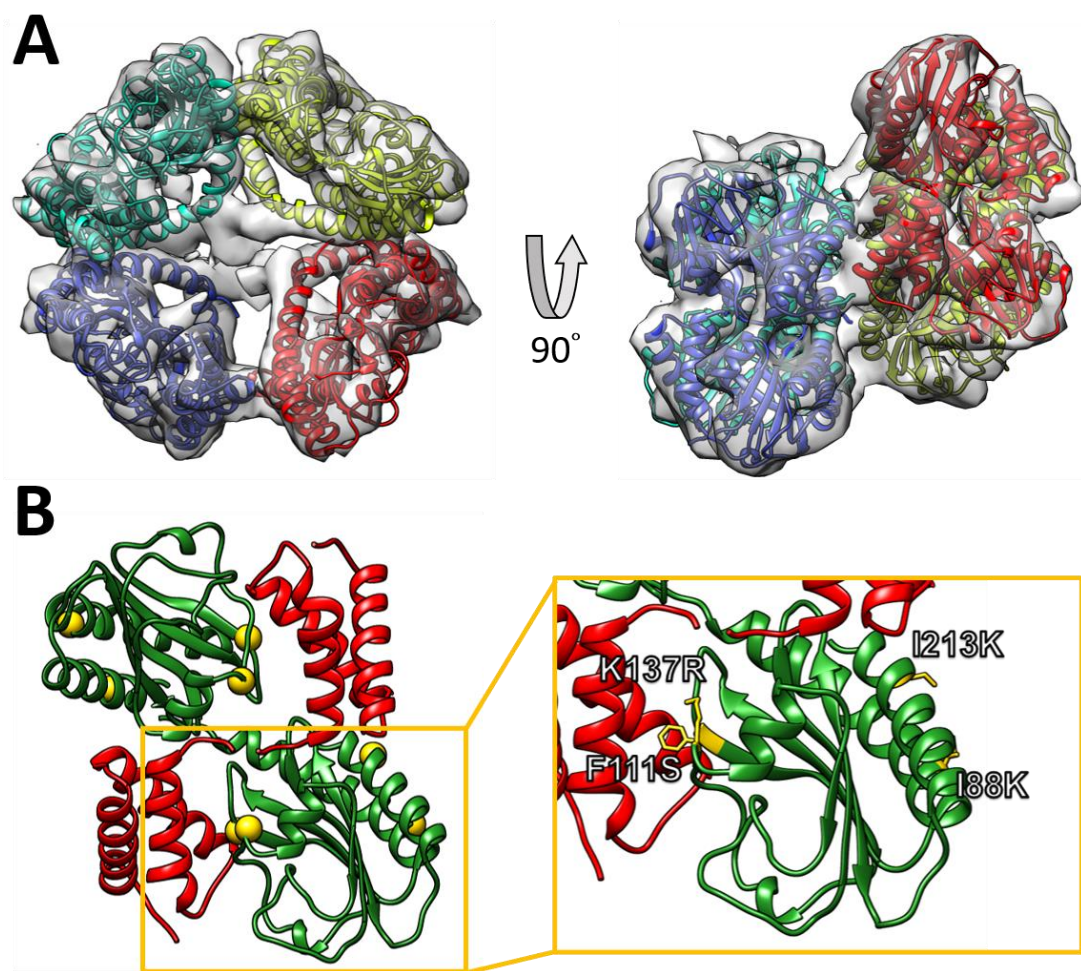


Figure V-2. The structure of sT-sRI complex.

(A) The cryo-EM map of the sT-sRI complex. Each color represents one sT-sRI heterotetramer. (B) The crystal structure of sT (Green) and sRI (Red) heterotetramer. Yellow points indicate four lysine-defective mutation sites of sT.

However, the structure of the four heterotetramers is not physiologically relevant, because the N-terminal of sT should link with the TMD in the T, but in the complex, the N-terminals of sT form a stable interaction in the middle of the circle. There is no space to put the extra TMD and cytoplasm domain. Therefore, the complex should not happen in a living cell. Nevertheless, the arrangement of the complex can explain some of the results in the mutagenesis experiments. Random mutagenesis experiments showed that four sT mutants (I88K, F111S, K137R, and I213K) lose the binding ability with RI (Figure A-1B) (Moussa et al., 2014). Two of these mutants (F111S, K137R) are at the sT-sRI binding surface in the heterotetramer, but the other two mutants (I88K, I213K) are not. The complex of the four heterotetramers shows that these two mutants (I88K, I213K) may also play important roles in forming a higher-order sT-sRI complex, while they are at the binding surface between the heterotetramers.

A.2.2 Cryo-EM image of the sT oligomer

Before lysis, holins accumulate and form oligomers in the inner membrane (Ry Young, 2002). Holin S105 of bacteriophage λ could form at least a hexamer. The oligomerization of lysis-defective S105 mutants was blocked at different stages of the hole formation (Grundling, Blasi, et al., 2000b; Grundling, Smith, Blasi, & Young, 2000). Similar to S105, T would form oligomers which might even higher-order for the membrane hole formation. Compared to lysis-defective mutants, wild-type T had one additional peak eluted at ~990 kDa (~ 34 copies of T) in gel-filtration chromatography (Moussa et al., 2014).

Previous studies showed that S105 formed a large ring made of approximately 72 subunits (Savva et al., 2008). We might get similar ring-like conformations of sT oligomers which are shown in the cryo-EM images (Figure A-2). However, the concentration of sT oligomers is too low because sT is easy to aggregate. Some mutagenesis experiments can be tested to decrease the level of aggregation. The binding interface of sT and sT should be similar to that of sT and sRI because the sRI can compete with sT in binding with another sT.

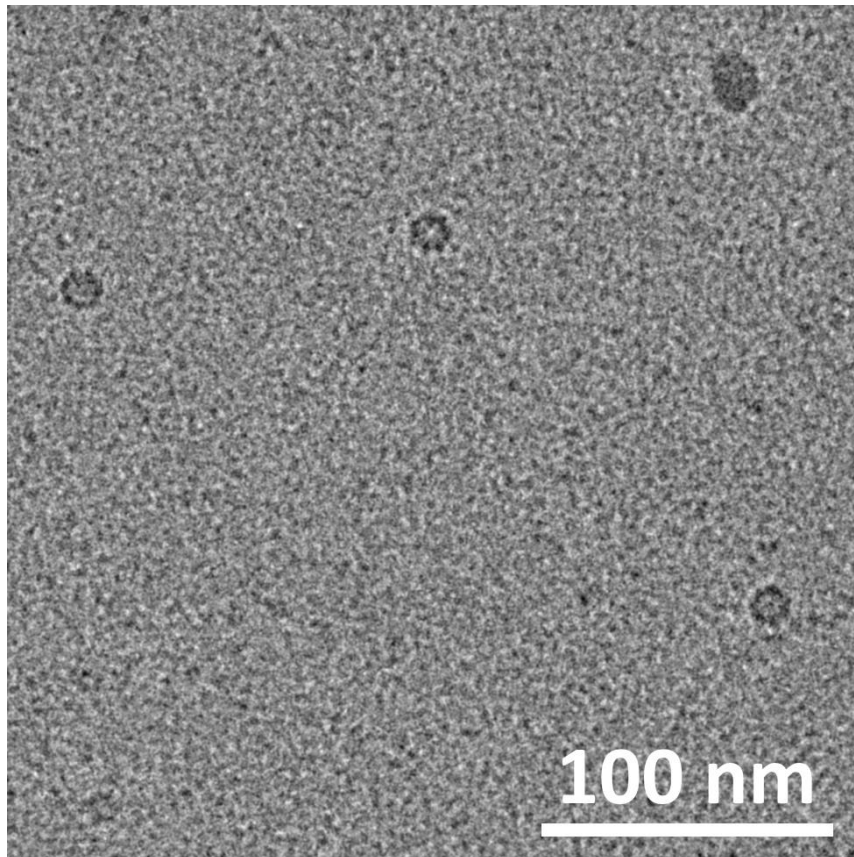


Figure V-3. Cryo-EM images of sT oligomers

The ring structure may be the sT oligomers which have a diameter around 14 nm.

A.3 Materials and Methods

A.3.1 Cryo-EM data collection

The specimen of the sT-sRI complex is from Dr. Inna Krieger in the Department of Biochemistry and Biophysics at Texas A&M University. Three microliters of the sT-sRI specimen (0.1 mg/ml) were applied to a C-Flat 1.2/1.3 holey carbon grid at room temperature with 100% relative humidity and vitrified using a Vitrobot Mark III (FEI). The prepared grids were imaged under an FEI Tecnai F20 transmission electron microscope operated at 200 kV. Data were recorded on a Gatan K2 Summit with electron-counting mode. A nominal magnification of $25,000\times$ was used, yielding a pixel size of 1.5 \AA . The beam intensity was adjusted to the dose rate of $\sim 10\text{ e}^-/\text{pixel/s}$ on the camera. A 40-frame movie stack was recorded for each picture, with 0.2-s per frame, for a total exposure time of 8 s.

A.3.2 Cryo-EM image processing and model building

Five hundred eight collected image stacks were aligned and filtered based on the electron dose using Unblur (Grant & Grigorieff, 2015). The defocus value of each image stack was determined by Gctf (Zhang, 2016). Particles were semi-automatically picked using e2boxer.py in EMAN2 (Tang et al., 2007). In total, 139,834 particles were picked with a box size of 144×144 pixels. The 2D and 3D classifications were performed using RELION (Scheres, 2012) to get clean and homogeneous particles for the final 3D reconstruction. After the post-processing, the final refined map from 35,432 particles

yielded a resolution of 9.4 Å. The particle images were downscaled by 2-fold in all processing procedures. Refinement using the particles without downscaling did not improve the resolution of the refined map.

To build the model, the crystal structure of the sT-sRI tetramer was first fitted into the EM density. In total, the EM map can accommodate four tetramers. To have a better fitting, the initial model of the four tetramers was then refined using the molecular dynamics flexible fitting with secondary structure restraints (Trabuco, Villa, Mitra, Frank, & Schulten, 2008).

A.4 References

- Doermann, A. H. (1948). Lysis and Lysis Inhibition with Escherichia coli Bacteriophage. *J Bacteriol*, 55(2), 257-276. doi:10.1128/JB.55.2.257-276.1948
- Dressman, H. K., & Drake, J. W. (1999). Lysis and lysis inhibition in bacteriophage T4: rV mutations reside in the holin t gene. *J Bacteriol*, 181(14), 4391-4396. doi:10.1128/JB.181.14.4391-4396.1999
- Grant, T., & Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife*, 4, e06980. doi:10.7554/eLife.06980
- Grundling, A., Blasi, U., & Young, R. (2000a). Biochemical and genetic evidence for three transmembrane domains in the class I holin, lambda S. *J Biol Chem*, 275(2), 769-776. doi:10.1074/jbc.275.2.769
- Grundling, A., Blasi, U., & Young, R. (2000b). Genetic and biochemical analysis of dimer and oligomer interactions of the lambda S holin. *J Bacteriol*, 182(21), 6082-6090. doi:10.1128/jb.182.21.6082-6090.2000

- Grundling, A., Smith, D. L., Blasi, U., & Young, R. (2000). Dimerization between the holin and holin inhibitor of phage lambda. *J Bacteriol*, *182*(21), 6075-6081. doi:10.1128/jb.182.21.6075-6081.2000
- Moussa, S. H., Kuznetsov, V., Tran, T. A., Sacchettini, J. C., & Young, R. (2012). Protein determinants of phage T4 lysis inhibition. *Protein Sci*, *21*(4), 571-582. doi:10.1002/pro.2042
- Moussa, S. H., Lawler, J. L., & Young, R. (2014). Genetic dissection of T4 lysis. *J Bacteriol*, *196*(12), 2201-2209. doi:10.1128/JB.01548-14
- Mukai, F., Streisinger, G., & Miller, B. (1967). The mechanism of lysis in phage T4-infected cells. *Virology*, *33*(3), 398-404. doi:10.1016/0042-6822(67)90115-8
- Park, T., Struck, D. K., Deaton, J. F., & Young, R. (2006). Topological dynamics of holins in programmed bacterial lysis. *Proc Natl Acad Sci U S A*, *103*(52), 19713-19718. doi:10.1073/pnas.0600943103
- Ramanculov, E., & Young, R. (2001). Genetic analysis of the T4 holin: timing and topology. *Gene*, *265*(1-2), 25-36. doi:10.1016/s0378-1119(01)00365-1
- Savva, C. G., Dewey, J. S., Deaton, J., White, R. L., Struck, D. K., Holzenburg, A., & Young, R. (2008). The holin of bacteriophage lambda forms rings with large diameter. *Mol Microbiol*, *69*(4), 784-793. doi:10.1111/j.1365-2958.2008.06298.x
- Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, *180*(3), 519-530. doi:10.1016/j.jsb.2012.09.006
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*, *157*(1), 38-46. doi:10.1016/j.jsb.2006.05.009
- Trabuco, L. G., Villa, E., Mitra, K., Frank, J., & Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, *16*(5), 673-683. doi:10.1016/j.str.2008.03.005
- Tran, T. A. T., Struck, D. K., & Young, R. (2005). Periplasmic domains define holin-antiholin interactions in T4 lysis inhibition. *Journal of Bacteriology*, *187*, 6631-6640. doi:10.1128/JB.187.19.6631-6640.2005
- Tran, T. A. T., Struck, D. K., & Young, R. (2007). The T4 RI antiholin has an N-terminal signal anchor release domain that targets it for degradation by DegP. *Journal of Bacteriology*, *189*, 7618-7625. doi:10.1128/JB.00854-07

- Wang, I. N., Smith, D. L., & Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annual review of microbiology*, 54, 799-825. doi:10.1146/annurev.micro.54.1.799
- Young, R. (2002). Bacteriophage Holins : Deadly Diversity Real-Time PCR Bioinformatics and Data. *Journal of Molecular Microbiology and Biotechnology*, 4, 21-36.
- Young, R. (2014). Phage lysis: three steps, three choices, one outcome. *J Microbiol*, 52(3), 243-258. doi:10.1007/s12275-014-4087-z
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol*, 193(1), 1-12. doi:10.1016/j.jsb.2015.11.003