STRUCTURE OF THE ssRNA PHAGE Qβ ENABLES DESIGN OF VIRUS-LIKE PARTICLES FOR RNA DELIVERY

A Dissertation

by

Karl Victor Gorzelnik

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,	Jason J. Gill
	Ryland F. Young
	Lanying Zeng
Head of Department,	Joshua Wand

May 2020

Major Subject: Biochemistry

ABSTRACT

Single-stranded (ss) RNA viruses infect all domains of life. Qβ, an ssRNA phage specific for *Escherichia coli*, has a near *T*=3 icosahedral lattice of coat proteins assembled around the genomic RNA (gRNA). In this work, the structure of the ssRNA virus Qβ was solved using cryo-electron microscopy, including the gRNA within the capsid. In ssRNA phages there is a well-established form of translational regulation wherein the coat protein binds specifically to a stem-loop at the start of the replicase. This stem-loop is thought to be important as a nucleation site to start the assembly of these phages. When looking for this site within the electron density, there were many which fit the previous crystal structure. This result led to the hypothesis that significant portions of the genome can be replaced and the virus can still assemble. The region encoding the coat protein was replaced with GFP or a kanamycin resistance gene, with the coat protein supplied *in trans*. The 'virus-like' particles were able to 'infect' cells to deliver their cargo, seen through fluorescence or bacteria resistant to kanamycin. The previously determined secondary structure of the gRNA was modeled into and largely fit the electron density of Qβ. Site-directed mutagenesis and plaque assays were used to validate the gRNA / maturation protein contacts. The mutated viruses were subsequently purified to verify that the coat proteins assembled into a capsid. During the establishment of a new purification procedure, new capsid morphologies were discovered, and subsequently verified by purifying the wild-type virus in the same manner as the non-infectious and thus poorly expressed mutants. The new capsid

morphologies were probably lost in previous purification techniques, as they have a different density due to the ratio of encapsidated RNA to proteins in the capsid. These capsid morphologies have never been seen for a wild-type ssRNA phage. For the smaller capsid, the volume is too small for the full-length genome, so the RNA inside is either degraded viral gRNA or host RNAs. In sum, this work proposes and validates guidelines for packaging foreign RNAs and their delivery into piliated bacteria using VLPs based off of ssRNA phages.

DEDICATION

To my loving wife, Keya Mukherjee, whose support made this possible.

ACKNOWLEDGEMENTS

I would like to thank Dr. Carlos Gonzalez for pushing me to finish writing my Master's and further pursue my Ph.D. in Dr. Junjie Zhang's lab. I would like to thank Dr. Junjie Zhang, for letting me explore my ideas, and guiding them along the way. Without his help I would not have been able to work, let alone survive the past couple years. The rest of my committee has been a great help along the way. I would like to thank Dr. Lanying Zeng for her guidance. Dr. Ryland Young's enthusiasm for ssRNA phages encouraged Dr. Zhang and helped me keep working on this project instead of having to switch to another system. Dr. Jason Gill has been a mentor for a decade, through my rotation in Dr. Young's lab, my time in Paul Straight's lab, then next door to his lab when I worked for Dr. Carlos Gonzalez, and finally through Dr. Junjie Zhang's lab.

I would like to thank Jim Sacchettini, and the help of many of his lab members facilitated by Dr. Jeremy Wood. I met Jeremy on a plane for the recruiting weekend from the biochemistry department, here at Texas A&M in February of 2010, and we've been friends ever since. Thank you Jeremy for your support throughout the years. I would be remiss to thank Jeremy, without mentioning his wife Claudia, whom Keya and I met when we entered the biochemistry graduate program here at Texas A&M. Dr. Claudia Castillo Gonzalez is certainly the hardest working scientist I know.

Within Dr. Zhang's lab, I have known none as well as Mengqiu Jiang, who is the best microscopist we have. I've known her since she was a visiting scientist in 2014 and

it has been a pleasure. Jeng-Yih Chang has been an awesome lab mate, he has helped me with numerous Linux questions over the years, and has been patient throughout all of my computer problems. Some of Jeng-Yih's work directly led to new avenues within my own projects, so I would not be writing the same dissertation without him. Ran Meng and Jirapat Thongchol kept me sane throughout the many ssRNA phage infections and delivery experiments, not to mention daily life. Zhicheng Cui was there with me during long data collection sessions in Baylor.

Laith Harb probably kept me here another two years, by performing some experiments which I designed and Dr. Zhang loved. Thanks Laith… That being said, Laith is a good friend and a great microscopist. Dr. Karthik Chamakura helped me understand ssRNA phages, in general, but especially with regards to MS2, perhaps the most well studied ssRNA phage. There are many other great scientists at Texas A&M who helped me with my experiments or let me borrow anything they had, especially members of Vishal Gohil's laboratory: Dr. Shiva Theja, Donna Iadorola, Natalie Garza, Dr. Mohammad Zulkifli, and Sagnika Ghosh.

I could not have been here without the support of my family. My parents have been very supportive over the years. I have been very fortunate in having such accepting and warm in-laws, thank you Ma and Baba. My brother and brother-in-law have each helped me endure in their own ways. As this work is dedicated to my wife, I would be remiss if I didn't acknowledge all the encouragement and support Keya has given me over the years. **ধন্যবাদ শোনা।**

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professors Junjie Zhang, Lanying Zeng, and Ry Young, of the Department of Biochemistry and Biophysics, and Professor Jason Gill, of the Department of Animal Science.

Data presented in Chapter Two was collected by Karl Gorzelnik and Zhicheng Cui at Baylor College of Medicine. The data was processed by Karl Gorzelnik and Dr. Junjie Zhang. This resulted in a *PNAS* publication in 2016. The data presented in Chapter Three was also collected at BCM by Karl Gorzelnik and Zhicheng Cui and was processed by Dr. Junjie Zhang and Zhicheng Cui. This was also published in *PNAS,* in 2017. Jeng-Yih Chang utilized this data to model the gRNA of Qβ into the electron density. His work has not been published. Some of his models are shown here to better explain my experiments, which have also not yet been published. Jeng-Yih Chang's work is partially shown in Chapters Three and Four, and is noted where it is presented. Laith Harb used VLPs which I purified to infect *E. coli*; his fluorescence experiments are shown in Chapter Four. Aaron Jacobson, from the Pellois lab, used the same VLPs I purified to infect HeLa cells. Haania Kakwan, an undergraduate in the lab, assisted in some cloning. Jeng-Yih, Laith, Aaron, and Zhicheng are all 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 each time we went. All other work conducted for the dissertation was completed by the student independently.

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, and 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

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1. Overview of ssRNA Phage Biology

Single-stranded RNA (ssRNA) bacteriophages (phages) were discovered accidentally in the search for phages that were dependent on F^+ *Escherichia coli*, a specific 'mating type' (Loeb 1960, Loeb and Zinder 1961). *E. coli* can transfer genetic material through conjugation, or mating, and has three mating types, F^+, F^-, and Hfr (Cavalli, Lederberg et al. 1953), which are differentiated based on their ability to donate genetic material. The strains which produce the physical elements required for conjugation, 'fertility factor' or 'F-factor' were deemed male cells, and produce F-pili, these strains come from the F^+ or Hfr backgrounds, which either have the genes encoded on the F-plasmid or chromosome. As later studies determined, ssRNA phages bind to pili, then utilize the host retraction machinery to pull in their genetic cargo, akin to the people of Troy pulling in the Trojan Horse, to their demise (Geraets, Dykeman et al. 2015).

ssRNA phages have since been model systems in molecular biology (Pierrel 2012). In the early days of molecular biology, messenger RNA (mRNA) was hard to purify in large amounts and susceptible to degradation (Pierrel 2012). ssRNA phages were a readily available source of homogenous and easily attainable mRNA which enabled their use to decrypt the genetic code enabling scientists to understand the nature

of genes. Early work on ssRNA phages ultimately allowed the study of translation initiation, translational gene regulation, RNA-protein interactions, host-virus interactions, and the development of RNA sequencing, the first whole genome sequenced was that of the ssRNA bacteriophage MS2 (Fiers, Contreras et al. 1976, van Duin 1988, Pierrel 2012). Perhaps the best studied RNA-dependent RNA polymerase comes from the ssRNA phage Qβ, which has led to extensive studies on RNA replication (Chetverin 2018). The capsids of MS2-like phages R17 and f2 were among the first icosahedral capsids solved (Crowther, Amos et al. 1975).

All ssRNA phages depend on retractile pili to enter their host bacterium due to the nature of their genome packaging within the capsid. Unlike double-stranded DNA phages, which pump their genetic material into a preformed capsid, ssRNA phages assemble coat proteins around their genome (Feiss and Catalano 2005, Gelbart and Knobler 2009, Aksyuk and Rossmann 2011, Dykeman, Stockley et al. 2013). Thus, ssRNA phages are not pressurized and are unable to use the pressure within their capsid to forcibly inject their genome into the host (Belyi and Muthukumar 2006, Brandariz-Nunez, Liu et al. 2019). Only about 5-10% of ssRNA phage particles are infectious (Cooper and Zinder 1963, Paranchych 1975). It is conceivable that this may be due to genome packaging, such that only a fraction of the viruses have their gRNA arranged in such a way to allow it to be pulled from the capsid quickly, so as to not be degraded by free RNases upon leaving the capsid.

Of the ssRNA phages, the two best-studied phages are MS2 and Qβ, which infect *Escherichia coli* through the sex (F^+) pilus (Rumnieks and Tars 2018). MS2, initially called f2, was the first RNA phage isolated, with Qβ isolated in Japan shortly afterward (Watanabe and Okada 1964). The differences between Qβ and MS2 will be discussed in length later on, but briefly, Qβ has an ~4.2kb ssRNA genome vs. ~3.6kb for MS2. Qβ has three open reading frames producing four proteins, the maturation protein, the coat protein whose gene has a leaky stop codon which yields an extended coat protein called A_1, and the β-subunit of the RNA-dependent RNA polymerase called the replicase (Figure 1). MS2 has at least four open reading frames, encoding: the maturation protein, the coat protein, the lysis protein, and the replicase (Chamakura, Edwards et al. 2017). In Qβ there is a fifth potential open reading frame in the replicase that has a viable ribosome-binding site (Nishihara, Morisawa et al. 2004), but the gene product shows no homology with any known protein and is not required for successful infection in *E. coli* (K. Chamakura, personal communication). We do not know the original hosts for MS2 and Qβ, only that they require F-pili, encoded on the F plasmid or the chromosomal marker Hfr (Zinder 1965). As these plasmids can travel between hosts, some gene products required for propagation on one host might not be necessary for another host. MS2 can overtake Qβ if the two are co-infected into *E. coli*, possibly by binding to the F-pilus better (Manchak, Anthony et al. 2002). However, there might be another host which Qβ is better suited for infection. The extended coat protein (A_1) in Qβ is required for infection (Hofstetter, Monstein et al. 1974). Though the exact mechanism is unknown, it could be that A_1 facilitates binding to the host pilus.

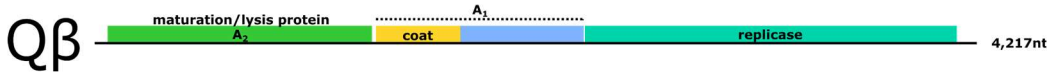

Figure I-1. Qβ genome organization
The positive-sense RNA genome codes for three genes which make four proteins. The genome organization of all ssRNA phages is conserved, starting with the maturation protein, which for Qβ has the dual role of being the lysis protein, in the case of Qβ it is named A_2 (colored green). Then comes the gene for the coat protein (colored yellow). For Qβ the gene for the coat protein has a leaky stop codon which is sometimes translated by ribosomes to an elongated version, A_1 (colored light blue). Last, is the replicase (colored teal). There are untranslated regions at the start and end of the genome, as well as between A_2 and coat, and A_1 and the replicase. The genome is drawn to scale based off of the Anc(P1) genome for Qβ (Kashiwagi et al., 2014).

1.1.1. The Infectious Cycle of Qβ

Most of the principles for ssRNA phage infection were only proven with a single phage, although it is thought that a rule learned for one generally applies to all other ssRNA phages. Qβ has an infectious cycle that can conclude as early as 40 minutes (Tsukada, Okazaki et al. 2009). Once a virion attaches to a retractable pilus of a host cell, the maturation protein / gRNA complex is taken up into the cytoplasm of the host via an unknown mechanism, as has been shown for MS2 (Krahn, O'Callaghan et al. 1972). As detached pili are unable to trigger genome withdrawal from the phage capsid (Novotny and Fives-Taylor 1974), it is thought that the Type IV secretion machinery for the F-pilus pulls the maturation protein and the gRNA out of the virion, leaving behind an empty viral capsid. Once inside the host cell, the positive-sense gRNA serves as messenger RNA (mRNA) (Weissmann 1974). The host translation machinery binds to and starts translating the coat/A_1 and the replicase right away (Weber and Konigsberg

1975). While there is a ribosomal binding site ahead of the maturation protein, this is only accessible on newly-synthesized RNA, before long-distance interactions are formed (Robertson and Lodish 1970, Kolakofsky and Weissmann 1971, Kolakofsky and Weissmann 1971, Staples, Hindley et al. 1971, Beekwilder, Nieuwenhuizen et al. 1996). This mechanism is conserved within ssRNA phages as a way to reduce the amount of maturation protein formed, as only one copy is incorporated into a mature virion (Takamatsu and Iso 1982). For Qβ, there is added reason to regulate the amount of the maturation protein, as it is the lysis protein, binding to MurA to inhibit host cell wall biosynthesis, even when the maturation protein is in a mature virion (Karnik and Billeter 1983, Winter and Gold 1983, Bernhardt, Wang et al. 2001).

Regulating lysis timing is essential for all phages to ensure maximal progeny production based on growth conditions and the number of potential future hosts (Wang, Dykhuizen et al. 1996). In the time it takes Qβ to synthesize enough A_2 to inhibit MurA, and lyse the cell, Qβ can synthesize at least 1,000 virions (Tsukada, Okazaki et al. 2009, Reed, Langlais et al. 2012, Yin and Redovich 2018). Depending on the growth conditions, by the time of lysis, there may be 10,000 to 100,000 particles released, of which estimates vary from 5-10% being infectious (Cooper and Zinder 1963, Paranchych, Krahn et al. 1970). Interestingly, ssRNA phages all appear to require active growth to lyse their host and release progeny (Lerner and Zinder 1977). ssRNA phages have tightly controlled all aspects of their growth, as twenty to thirty minutes after infection replicase production is almost stopped, while coat protein production continues

throughout production until lysis (Vinuela, Algranati et al. 1967). Replicase protein production is halted by the coat protein binding to the 'operator' at the start of the RNA for the replicase gene which inhibits initiation of translation (Robertson, Webster et al. 1968, Eggen and Nathans 1969). Eventually, the level of coat proteins is high enough that they start to assemble around the gRNA, possibly nucleating from the site of the operator, but not necessarily as there are other operator-like sequences within the genome (Beekwilder, Nieuwenhuizen et al. 1996). The coat protein binds to the operator with a high affinity (Kd ~1nM) but also to many other stem-loop structures throughout the gRNA (with Kd values of 10nM to 1000nM), leading to many potential correct assembly pathways (Carey, Cameron et al. 1983, Borodavka, Tuma et al. 2012, Garmann, Goldfain et al. 2019).

1.1.2. The Unique Scientific Role of ssRNA Bacteriophages

Mentioned briefly in the first section, ssRNA bacteriophages were widely studied due to the versatility afforded by using a microbial system to study the biochemistry and genetics of a virus (Fiers 1979). They played a key role in early molecular biology because they were a source of highly homogeneous RNA (van Duin 1988), although the reasons why RNA encapsidated in ssRNA phage coat proteins may not be entirely homogeneous will be mentioned later on in the discussion chapter of this dissertation. This readily available source of RNA, which was also mRNA, enabled researchers to crack the genetic code (Pierrel 2012). The MS2 coat protein was the first gene to be sequenced in 1972 (Min Jou, Haegeman et al. 1972) with the whole genome fully

sequenced just four years later (Fiers, Contreras et al. 1976), which is truly remarkable given the difficulty of sequencing at the time. Knowing the whole genomic sequence helped in understanding the initiation of translation, regulation of gene expression at the translational level, as well as translational control by repressor proteins. Less well-studied areas, but still vital, is the knowledge gained from RNA phages in regards to the origin of life in an RNA world, as well as compartmentalization being necessary to prevent parasitic RNAs from taking over in an RNA world. After molecular biology matured enough for other organisms to become widely used, studies on RNA phage biology dropped off, but they were widely studied as model systems for RNA viruses.

1.2. Cryo-Electron Microscopy as a Unique Tool

When trying to enclose the large volume needed for a genome, from a viral perspective it makes sense to enclose the genomic material within a large capsid composed of many copies of as small of a protein as possible (Crick and Watson 1956). This was seen in filamentous and icosahedral viruses, from Rosalind Franklin's structure of tobacco mosaic virus, with 16-17 protein subunits per helical turn of the RNA genome (49 for 3 turns), to the icosahedral tomato bushy stunt virus (Caspar 1956, Holmes and Franklin 1958). The realization by Crick and Watson that some viruses have non-crystallographic icosahedral symmetry led to major leaps in structural virology (Crick and Watson 1956). Viral symmetry has subsequently played a key role in structural biology, particularly within electron microscopy, with many advances coming as a result of honing techniques with viruses (Valentine 1958, Wang, Mukhopadhyay et al. 2018).

Cryo-electron microscopy (cryoEM) is a technique used to study the three-dimensional structure of biological specimens under near-native conditions (Zhang, Gorzelnik et al. 2016). In cryoEM a biological specimen is frozen in a thin layer of ice and imaged using a transmission electron microscope equipped with a holder that keeps the specimen frozen (Passmore and Russo 2016). By combining the images of many identical particles in different orientations, it is possible to determine the structure of the molecule to a high resolution (Bai, Fernandez et al. 2013).

The first 3D reconstruction from electron microscopic data was that of the tail sheath of phage T4, whose helical symmetry allowed reconstruction from 2D images (De Rosier and Klug 1968). Within a few years, the first icosahedral structure was reconstructed computationally from multiple micrographs and views of tomato bushy stunt virus (Crowther, Amos et al. 1970). Computationally combining multiple images into one structure demonstrated the power of single-particle transmission electron microscopy and computational analysis and the technique took off by the end of the decade (Frank, Goldfarb et al. 1978). Computational reconstructions of viruses were hindered by uneven staining, as well as distortions produced by staining, and were only possible for the portions of the macromolecule able to be stained (Rossmann 2013). However, the ability to flash freeze the sample in vitreous ice, enabled cryo-electron microscopy which advanced the technique (Adrian, Dubochet et al. 1984). CryoEM enabled microscopists to look at samples without staining or fixing, increasing the

contrast without changing the specimen (Adrian, Dubochet et al. 1984). By the ’90s better methodology had enabled researchers to solve a sub-nanometer structure of the Hepatitis B virus (Bottcher, Wynne et al. 1997). Single-particle virus reconstructions reached a near-atomic resolution within a decade, and the 3.3 Å aquareovirus showed how powerful *de novo* modeling could be to build a structure from electron density alone (Zhang, Jin et al. 2010).

Structure determination using cryoEM typically uses a technique called ‘single-particle’ cryoEM, wherein a macromolecule is imaged, hopefully from many different orientations. The different orientations of the molecule are then computationally assembled into a 3D structure (Zhang, Gorzelnik et al. 2016). As the particular angles describing each different view are unknown, they have to be determined by aligning the images to a reference structure, often a blurred out mass, with any errors in alignment reducing the overall resolution of the new 3D structure (Grigorieff 2013). Thousands to millions of particles are required to give atomic-resolution to symmetry free macromolecules (Zhang, Gorzelnik et al. 2016). As the grid moves slightly during imaging due to beam-induced movement or vibrations in the building, software is required to track the movement of particles and align them, to further increase the resolution from a given dataset (Bai, Fernandez et al. 2013).

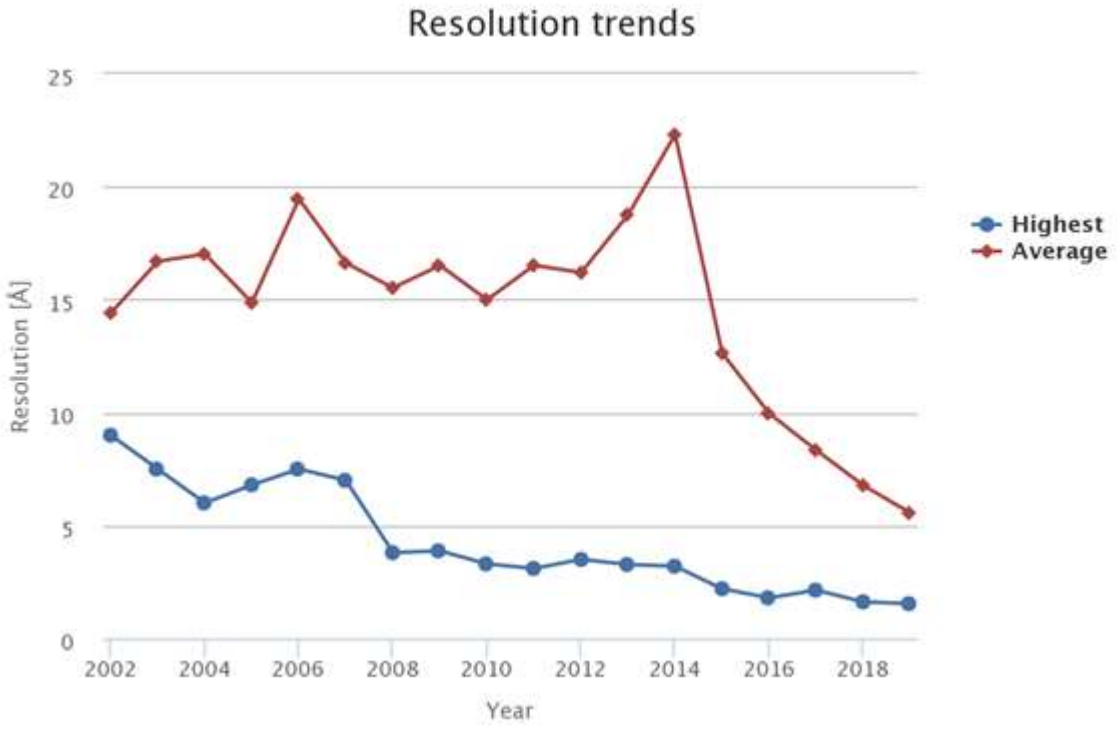

Figure I-2. Highest and average resolutions of the structures annually deposited in the EM Data Bank
It should be noted that the spikes for an increased average resolution is due to an increase in the overall numbers of structures deposited, and not a decline in the quality of structures. Data includes all structures deposited in EMDB, including tomography, micro-ED, and 2D crystal structures, as well as single-particle cryoEM. Figure accessed on September 15th, 2019. (https://www.ebi.ac.uk/pdbe/emdb/statistics_num_res.html/).

CryoEM has enjoyed a rapid rise as a structural biology tool, with the average resolution of structure recently reaching levels where protein secondary structures are able to be accurately fit (Figure I-2). The number of near-atomic resolution structures (<4 Å) in the EM Data Bank (http://www.ebi.ac.uk/pdbe/emdb/) rose from 36 in 2014 to 241 in 2016 and 622 in 2018 (Figure I-3). It should be noted, when looking at Figure I-3 that the number of accession codes released in the first three quarters of 2019 was

already higher than that in 2018. In the spring of 2019, EMDB shifted from a four-digit number accession code to one with five digits, as the number of structures deposited have reached a level where the EMDB ran out of codes.

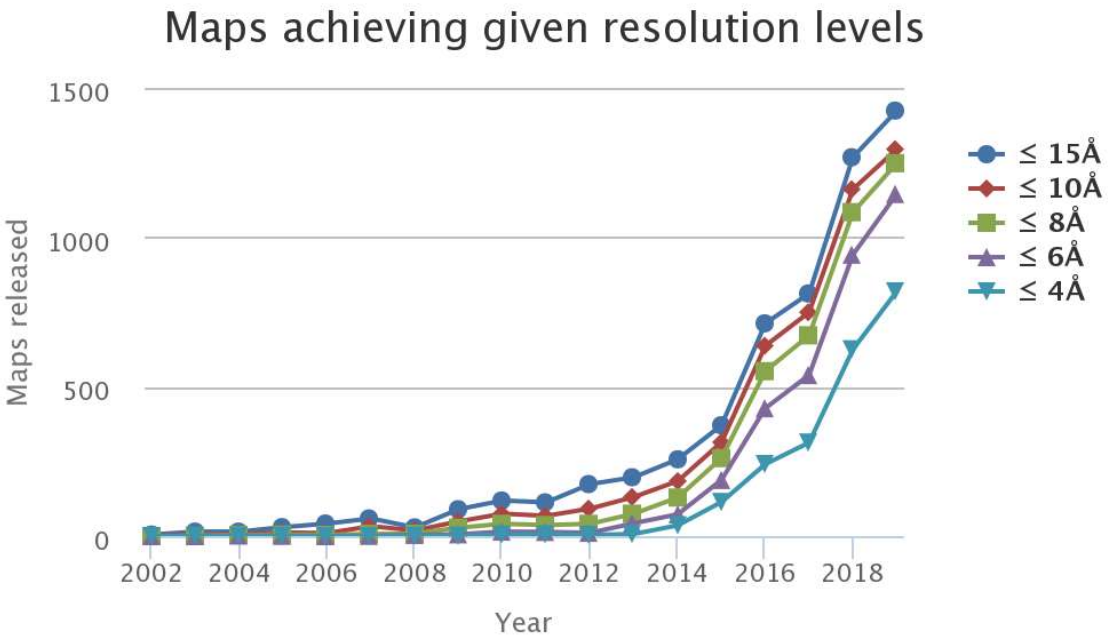

Figure I-3. Number of structures deposited each year with EMDB at a particular resolution from 2002-2019
The slope decrease from 2018 to 2019 is merely due to the data being from September 15th, 2019, not the whole calendar year.
(http://www.ebi.ac.uk/pdbe/emdb/statistics_num_res.html/)

Traditionally EM images were recorded on photographic film, which was not convenient for high-throughput methods due to the laborious process of developing the film and digitizing the images (Bai, Fernandez et al. 2013). The advent of charge-coupled device (CCD) cameras accelerated the collection of data and enabled automated data collection (Stagg, Lander et al. 2006). CCD cameras increased the workflow of

microscopes at a cost because the electrons passing through a sample had to be converted into visible light to be detected (Carroni and Saibil 2016). The introduction and spread of direct electron detectors is a critical factor in driving the recent surge in high-resolution structures (Grigorieff 2013). By detecting the electrons directly, these cameras can take pictures faster, enabling researchers to compensate for beam-induced drift (Bai, Fernandez et al. 2013). The increased size of datasets has further required computational developments to improve the resolution of structures (Zheng, Palovcak et al. 2017). These new detectors have enabled cryoEM to shift away from the traditional focus on large macromolecular complexes, such as icosahedral viruses and ribosomes, toward smaller specimens that have been neglected in crystallography, such as membrane proteins (Lyumkis 2019).

1.2.1. Viruses as a Model System for CryoEM

Due to their large size, relative ease of purifying large quantities, and symmetry, viruses have been an excellent model system for the development of new techniques in electron microscopy and three-dimensional image reconstruction (Jiang and Tang 2017). CryoEM has grown by leaps and bounds within the last decade into a fully mature structural biology technique, capable of solving 3D structures to high-resolution (Stass, Ilca et al. 2018). The growth in cryoEM is primarily due to the development of direct electron detectors and the advancement of methodology needed to process large, high-resolution datasets (Zhang, Gorzelnik et al. 2016). Direct electron detectors obtain sufficient signal to solve asymmetric structures of increasingly smaller proteins and

complexes. With microscopes, detectors, and image-processing algorithms no longer the bottlenecks for structure determination, the major hurdles for structure determination are now in the grid-preparation stages and finding researchers with the time to image a sample (Yu, Li et al. 2016).

While X-ray crystallography was capable of determining the structures of icosahedral viruses by the 1970s (Harrison, Olson et al. 1978), it was challenging to achieve high-resolution for large viruses, or flexible loops within the viral core. Additionally, crystallography requires screening many conditions, which may introduce artifacts into the sample, but it also takes a long time. During the Zika outbreak of 2015-2016 cryoEM was used to solve two structures of the viral core within months independently (Kostyuchenko, Lim et al. 2016, Sirohi, Chen et al. 2016).

CryoEM has been used extensively to determine structures of icosahedral capsids. Due to their symmetry, the number of particles required for structure determination is less than that for typical proteins, often just requiring thousands of particles to reach 3 Å resolution (Yu, Li et al. 2016). At this resolution, it is possible to recognize amino acid side-chain densities and build the protein backbone *de novo* (Laanto, Mantynen et al. 2017). As of March 2019, there are many icosahedral viruses with near-atomic resolution, surpassing the resolution achieved via crystallography and becoming the preferred approach to studying the structural components of these viruses (Harrison 2010). These include Adeno-associated virus at ~1.9 Å (Tan, Aiyer et al.

2018), human enterovirus D68 at ~2.2 Å (Liu, Sheng et al. 2018), rhinovirus at 2.3 Å (Dong, Liu et al. 2017), Hepatitis B at 2.6 Å (Bottcher and Nassal 2018), Seneca Valley Virus at 2.8 Å (Cao, Zhang et al. 2018), although all of these viruses have symmetry applied to increase the resolution.

The gRNA within icosahedral viruses are often obscured using crystallography (Schneemann 2006), or they can be erroneously averaged into artifacts when performing symmetric reconstructions in cryoEM datasets (Guo and Jiang 2014). By treating the gRNA as an asymmetric feature, it can be visualized during reconstruction (Gorzelnik, Cui et al. 2016). Groups studying ssRNA phages have been aided tremendously in this regard compared to those studying eukaryotic viruses, as the capsid symmetry in ssRNA phages is broken by a single copy of the maturation protein, which also binds to the gRNA, locking it in a dominant conformation (Gorzelnik, Cui et al. 2016, Koning, Gomez-Blanco et al. 2016, Zhong, Carratala et al. 2016, Cui, Gorzelnik et al. 2017, Dai, Li et al. 2017). Typically, these viruses are required to be at concentrations of 10^{12}-10^{14} particles/mL for cryoEM (Guo and Jiang 2014, Yu, Li et al. 2016), although the effective plaque forming concentration might be significantly lower, if they are inactivated in the purification process.

One problem in structural virology is the relationship between particles and infectious viruses, for human herpes simplex virus this ratio is 10:1, whereas for another human herpes virus, Varicella-Zoster virus, that ratio is ~40,000:1 (Harland and Brown

1998, Carpenter, Henderson et al. 2009). Only 10% of hepatitis B virions contain their gRNA (Ning, Nguyen et al. 2011). Even Ebola, one of the most feared viruses, has a particle to plaque forming unit of ~500:1 (Alfson, Avena et al. 2015). This is particularly relevant for ssRNA phages, which have a particle to PFU ratio of ~10:1 (Cooper and Zinder 1963, Paranchych 1975). One method which has been used to increase the concentration of viruses is to use cryoEM grids with an affinity for some antigen, as was the case for Tulane virus, resolved to 2.6 Å (Yu, Li et al. 2016). This method is expected to increase in popularity for viruses that are less abundant or hard to purify in cell culture systems, such as human norovirus (Ettayebi, Crawford et al. 2016). Eventually, increased hardware and software capabilities will negate the need for purifying macromolecules to near homogeneity, as in vivo images of viruses infecting a cell or assembling in the host will be able to be visualized via cryo-electron tomography. There are excellent cryoET structures of viral assembly intermediates in a cyanobacterium (Dai, Fu et al. 2013) and the capabilities to visualize within a cell will only increase. The structure of the ssRNA phage MS2 attached to its receptor, the F-pili, was recently solved using single-particle averaging taken of particles imaged at different angles to solve the preferred orientation problem that comes with having long filamentous structures (Meng, Jiang et al. 2019).

The larger the size of the viral capsid, the thicker the ice needs to be to surround the sample on the grid. The resolution available for a given sample decreases as the thickness increases past a certain point, due to the defocus gradient across the specimen

as well as the Ewald sphere effect (Stass, Ilca et al. 2018, Zhu, Wang et al. 2018). Another limitation is that the larger the virus the fewer particles you can get in a single image, increasing the data collection time exponentially. One way around this problem is to collect images at a lower magnification, but this will reduce the resolution as well. Data processing for large viruses is still more difficult than for smaller viruses, but recently there have been many large viruses determined, including human cytomegalovirus (at ~130nm in diameter), Kaposi's sarcoma-associated herpesvirus, and herpes simplex type 2 virus (~200nm in diameter) (Stass, Ilca et al. 2018). The capsid is a very large component in viruses, and symmetric capsids can obscure asymmetric features within them (Zhang, Gorzelnik et al. 2016). This can be resolved by subtracting the capsid contributions from the particle during data processing and adding them in at a later state then performing a subsequent asymmetric reconstruction (Zhang, Kostyuchenko et al. 2007).

An advantage, or disadvantage, of electron microscopy is the ability to classify particles and discard irregular particles that do not fall into classes with the 'average' macromolecule (Baker and Cheng 1996, Scheres 2012, Wang, Mukhopadhyay et al. 2018). Therefore, many particles with irregular protein distributions might be excluded for those with more 'average' features (Baker and Cheng 1996). Class averaging enables the visualization of details that would otherwise be lost in the noise of the electron micrographs.

One significant problem for the structural determination of viruses using electron microscopy is the variation of structure when looking away from icosahedral viruses, toward pleomorphic viruses such as HIV, influenza, bunyaviruses and many others (Rossmann 2013). Due to the lack of symmetry and amorphic shape, these viruses do not lend themselves to reconstructions using single-particle cryoEM. However, cryoET is becoming a popular technique for examining irregular viral structures (Jun, Ke et al. 2011, Grange, Vasishtan et al. 2017, Si, Zhang et al. 2018), as also was mentioned earlier with regards to MS2 attaching to its receptor (Meng, Jiang et al. 2019). In cryoET, the sample is imaged multiple times at different angles (often at 1-2° intervals) to give different orientations that are combined computationally to give a three-dimensional structure. Due to the limitations of rotating a grid, these angles are limited to +/- 70°, leaving a wedge of 40° which is unable to be included in the final reconstruction. Additionally, due to the radiation damage associated with the electron beam hitting the sample, the length of exposure at each projection needs to be limited, so the resolution is significantly lower compared to single-particle analysis (Rossmann 2013).

1.3. Host-Virus Relationship

The lifecycle, if one would call it that for non-metabolic organisms, of ssRNA phages can be split into two parts: outside the host, trying to get in; and inside the host, building up the resources to get out. The first step of the infectious cycle, trying to get in, can further be divided into three stages: adsorption to the pili, gRNA release, and

penetration (seen as 1 A, B, and C in Figure I-4, respectively) (Rumnieks and Tars 2018). It seems that all ssRNA phages require hosts with pili, while these pili can be very different between hosts. The pili specific for ssRNA phages all share a common feature, they need to be able to retract in order for infection to occur (Rumnieks and Tars 2012). The second part of the infectious cycle, once the phage gRNA gets pulled inside of the host (denoted as 2 in Figure I-4), is when the virus hijacks host translation machinery to make enough progeny for subsequent release and reinfection.

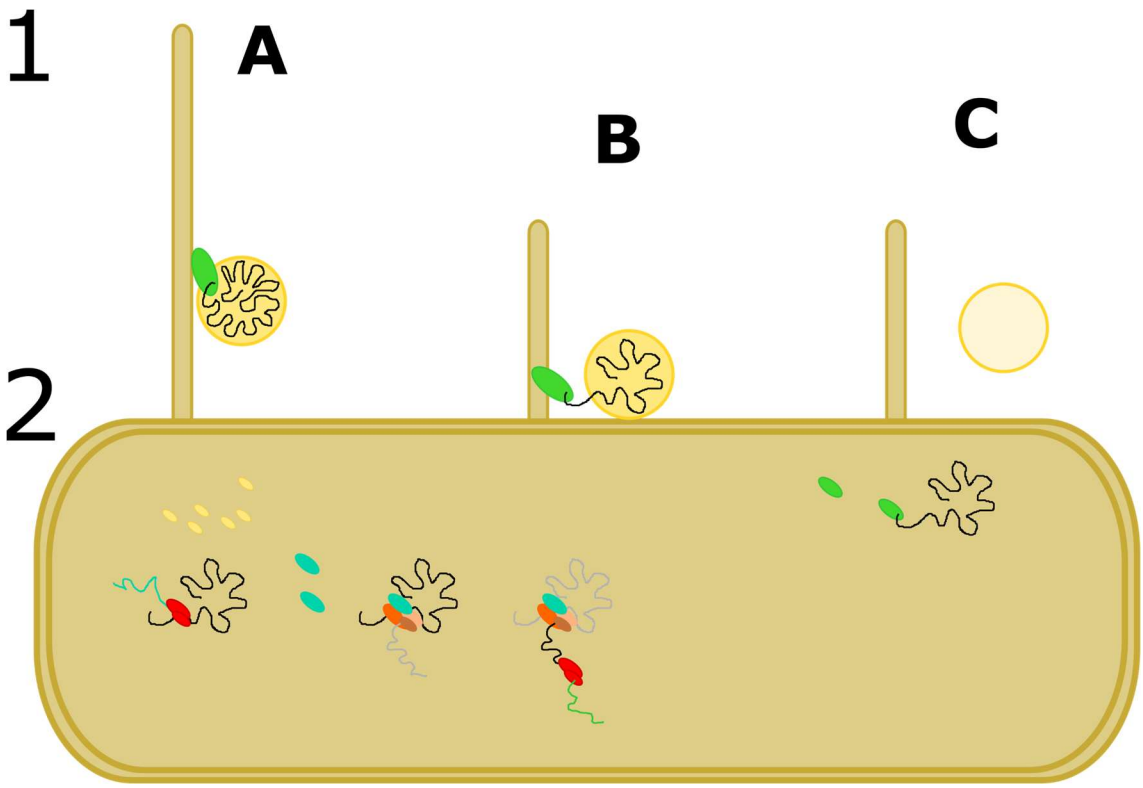

Figure I-4. Infectious process for ssRNA phages
In the first part of infection (denoted in 1A-C) the phage are outside of the cell and bind to retractile pili (the phage maturation protein is colored in green, the capsid in yellow, the genomic RNA in black); gRNA release comes next (1B) followed by penetration (1C) of the cell outer- and inner-membrane by the gRNA and possibly the maturation protein cleaved into two pieces. Once the gRNA is

inside the cell the host ribosomes (red) start translating the positive-sense gRNA as it would any other mRNA in the cell, producing coat proteins (yellow) and replicase (teal). When the replicase is in sufficient quantities it replicates the positive-sense gRNA into negative-sense gRNA (grey) which can be further replicated into positive-sense gRNA, from which the maturation protein can be made, while the gRNA is still being replicated.

1.3.1. The Infection Process

In the most comprehensive work on ssRNA phages, the 1975 book 'RNA Phages', William Paranchych starts off his chapter with the prescient statement, 'It is perhaps ironical, in view of our present understanding of the RNA phage replicative processes, that the molecular mechanisms involved in the adsorption and penetration stages of RNA phage infection are relatively poorly understood' (Paranchych 1975). Forty-plus years later little more is known about these steps in infection, while much more is known about the phages, from detailed requirements for RNA replication to virion morphogenesis to host lysis to the nucleotide sequences of these viruses and much more.

The principles for most phage processes are derived from studies on MS2 (and MS2-like phages such as R17, f2, etc.) and Qβ. Each mature virion contains ~180 copies of the coat protein, one maturation protein, and a single copy of the gRNA (Steitz 1968). The infectious process of all ssRNA phages is dependent on initial binding to retractile pili (Brinton 1965, Silverman and Valentine 1969, Novotny and Fives-Taylor 1974). This has been shown for every ssRNA phage with a known host (Rumnieks and Tars 2018). As stated previously, ssRNA phages bind to the pili with their maturation

protein (Roberts and Steitz 1967). The maturation protein is named such because without it the viruses are non-infectious, with it they are 'mature' virions (Engelhardt and Zinder 1964). A virus without the maturation protein is unable to absorb to F-pili, suggesting that this protein was required for attachment (Lodish, Horiuchi et al. 1965). Sensitivity to different phage groups has been used to classify distinct pilus types, even within F-pili (Frost, Finlay et al. 1985). As a side note, none of the viral proteins are actually required for the virus to propagate if the gRNA transformed into spheroplasts or is electroporated into the host (Engelhardt and Zinder 1964, Taketo 1989).

There was a debate within the ssRNA phage community as to when the gRNA leaves the capsid, whether this happens while the phage is attached to the middle of the pili or if it happens when the phage has reached the type IV secretion system which is in the outer membrane (Figure I-5) (Brinton 1965, Marvin and Hohn 1969, Paranchych, Ainsworth et al. 1971). Meng *et al.*, showed structures of MS2 bound to F-pili which verified that the genome remains inside the capsid up until the pilus retracts to the cell surface, proving the third model of genome entry/ejection seen in Figure I-5 (Meng, Jiang et al. 2019). When the gRNA is taken up into the host while the empty viral capsid dissociates from the pilus and drifts into the surrounding media (seen as panel C) (Silverman and Valentine 1969, Paranchych, Krahn et al. 1970). One study, using radiolabeled ^{3}H maturation proteins and ^{32}P gRNA, reported that the maturation protein is taken up into the cell and in the process cleaved into two pieces, 15kDa and 24kDa (the maturation protein for MS2 is 39kDa), along with the gRNA (Krahn, O'Callaghan et

al. 1972), but this was done with a multiplicity of infection (MOI) greater than 300-400 viral particles per cell. It could simply be that some amount of the maturation proteins are degraded by an extracellular protease and taken up into the cell by another means. In that paper, the authors reported that the gRNA leaves as many as 250 phage particles, but the gRNA which makes it into the cell is only equivalent to about 35-40 particles (Krahn, O'Callaghan et al. 1972). That leaves a lot of gRNA and maturation proteins available to extracellular RNases and proteases, which could break down the phage RNA and capsids to later be taken up by the cell for nutrients.

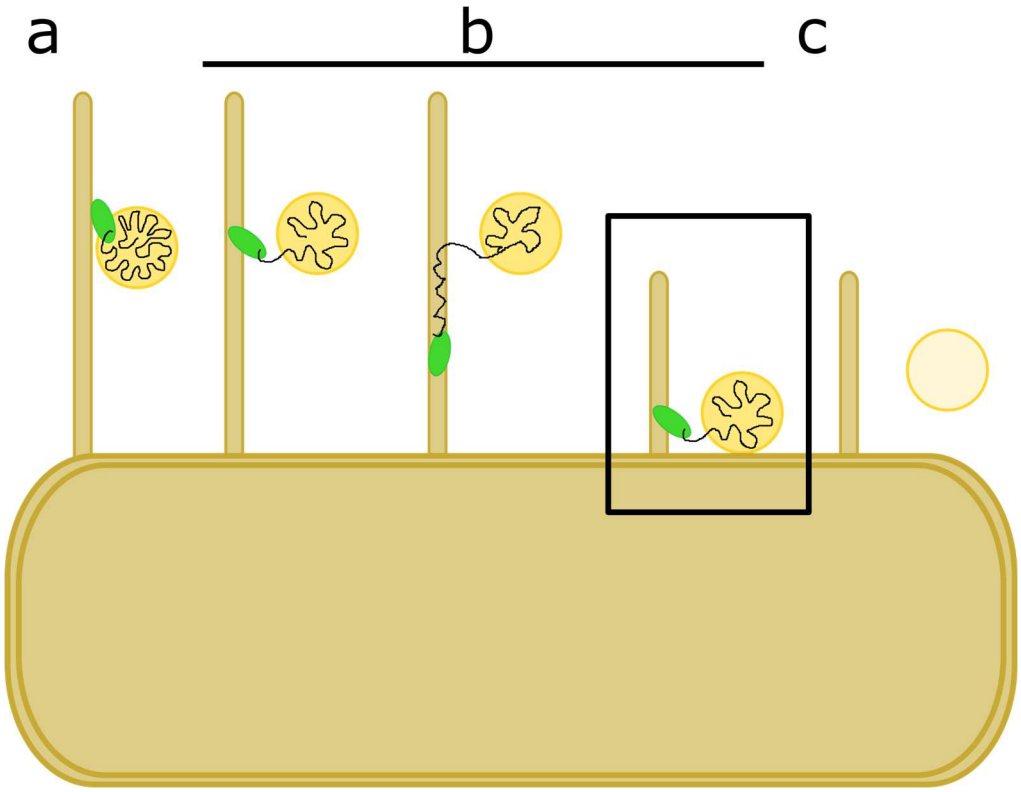

Figure I-5. RNA release in ssRNA phages
Panel A shows the initial stage of ssRNA phage binding to the host pilus. Panel B indicates three possible scenarios for gRNA release from the phage which were hypothesized, wherein the left shows the genome leaving the capsid at some time

during binding to the pilus, the middle shows the maturation protein and the genome leaving the capsid and traveling down the channel of the pilus at some point during attachment, while the right panel indicates that the genome remains inside the capsid until the phage reaches the surface of the cell, when the genome and the maturation protein are pulled into the host. Panel C illustrates that the empty capsid floats off into the media after the genome/maturation protein complex enters the cell. The third panel of B is boxed because there is structural evidence to support it.

The gRNA of MS2 was shown to bind to the maturation protein in vitro in two places, within the maturation protein-coding sequence and in the 3' UTR (Shiba and Suzuki 1981). However, a high-resolution cryoEM structure of MS2 revealed that the gRNA only comes in contact with the maturation protein in virio at the 3' end of the RNA (Dai, Li et al. 2017), with nucleotide numbers slightly different than what was reported. The nucleotide discrepancy can be ignored because every group thinks they are using wild-type phage. Each lab is really using a derivative that has mutated at least once per generation due to the high mutation rate of ssRNA viruses. The quasispecies theory that holds these viruses are merely the average of all the related viruses (Domingo, Sabo et al. 1978). The high resolution of the cryoEM electron density unambiguously shows that only the 3' end of the genome of MS2 comes in contact with the maturation protein in the mature virus. It is possible that the maturation protein interacts with other regions of the gRNA in the host, as the original observation was established in vitro without coat proteins (Shiba and Suzuki 1981). However, this is unlikely, as it would be seen in at least a fraction of the cryoEM data. Observations from the published structures of MS2, as well as from our own lab, suggests that >95%

of the dataset for MS2 gRNA falls into a dominant conformation (Jeng-Yih Chang, personal communication).

RNA release requires cell-attached pili (Brinton and Beer 1967, Danziger and Paranchych 1970) which retract (Valentine and Strand 1965), meaning the cells have to be metabolically active, as pili that are either sheared off or are inactive, such as in stationary phase, are unable to support gRNA uptake. If phages are incubated with cells at 4°C the binding to pili is reversible and the virions are still infectious after detachment from the pili (Valentine and Strand 1965, Valentine and Wedel 1965, Silverman and Valentine 1969, Paranchych, Krahn et al. 1970). Similarly, phages can be added to free pili without a loss of infectivity as seen by the fact that the structure of the MS2 bound to detached F-pili shows the gRNA within the capsid, not released (Toropova, Stockley et al. 2011, Meng, Jiang et al. 2019). The current consensus is that the phage simply sits on the pilus, binding and releasing within some kinetic parameters until the pilus retracts enough for the phage to be proximal to the membrane and the maturation protein / gRNA complex pulled in by the host. This was shown using fluorescence microscopy (Clarke, Maddera et al. 2008). Using fluorescently labeled R17 the authors could see pili retract and even could see new pili being secreted, which would not be able to happen if there was some sort of a signal transduced by phage binding that instructs the cell to retract the pili (as has been proposed) (Clarke, Maddera et al. 2008, Lang, Kirchberger et al. 2011).

Once the phage is attached to the pili, in actively growing cells, there is a period when the phage is 'sensitized' to RNases, where the genome has left the protection of the capsid before it is taken up by the host (Valentine and Wedel 1965, Zinder 1965, Paranchych 1966, Danziger and Paranchych 1970). In order for the gRNA to leave the capsid, the maturation protein must also dissociate (as is shown in Chapter 2) to make space for the bulky secondary structures to exit. Since the maturation protein binds the gRNA, as well as the F-pilin monomers which make up the F-pilus, it is thought that when the type IV secretion (T4S) system retracts the pilus one monomer at a time, it pulls the maturation protein/gRNA complex in, mistakenly instead of an F-pilin monomer (Lang, Kirchberger et al. 2011). T4S systems contain ~12 proteins to make up a multi-megadalton complex that crosses the bacterial envelope with a central channel to enable nucleotide transfer across the periplasmic space (Fronzes, Christie et al. 2009, Low, Gubellini et al. 2014, Costa, Ilangovan et al. 2016, Hu, Khara et al. 2019).

Little is known about the requirements of T4S systems that enable ssRNA phage penetration, other than the necessity of retractable pili. TraD, a cytoplasmic coupling protein within the conjugative T4S system, has been shown to be required for ssRNA phage infection, although through an unknown mechanism (Bayer, Eferl et al. 1995, Lang, Kirchberger et al. 2011). There are traD mutants that are resistant to MS2 infection, but not Qβ (Achtman, Willetts et al. 1971). A curiosity as there are few differences between these phages. An amber mutant of traW, an inner-membrane F-pilus assembly protein, was shown to convey resistance to ssRNA phages, while still

being able to complement the transfer of plasmids with defects in other tra genes (Miki, Horiuchi et al. 1978). TrbI was also shown to affect phage infection. TrbI is an inner-membrane protein whose gene is located within the *tra* genes on the F-plasmid (Maneewannakul, Maneewannakul et al. 1992). Hosts with a kanamycin insertion within the *trbI* gene showed reduced sensitivity to ssRNA phages but were still proficient for DNA transfer. These mutants expressed very long pili, which might indicate that TrbI plays a role in pilus retraction, a function necessary for ssRNA phage infection, although there may be polar effects from the insertion of a resistance marker within the gene. TraG, a transmembrane protein forming a pore in the inner-membrane as part of the conjugative complex, can have mutations in it which abolish Qβ infection, and severely weaken MS2 infection (Frost and Paranchych 1988). Otherwise, there is little experimental evidence of T4S system requirements for ssRNA phage infections.

1.3.2. Co-option of Host Proteins for Replication

Once the ssRNA phage gRNA enters the cell it acts as a regular mRNA to be translated by the host ribosomes (Rumnieks and Tars 2018). The maturation protein may or may not enter the cell, radiolabelled maturation proteins were shown to be inside of a non-radiolabelled host cell by a group in the 1970s, but there has not been any follow-up since, and the cleaved maturation protein is not believed to play any further roles (Krahn, O'Callaghan et al. 1972). Nothing is known about the secondary structure of the gRNA as it enters the cell, presumably, even if the secondary structure is broken as the gRNA leaves the capsid and then makes its way through two membranes and the

periplasmic space, the gRNA would eventually refold into an optimal conformation once inside the cytoplasm.

Of the three to four genes in the genome of ssRNA phages, only the coat protein can be translated in the native conformation of the gRNA (Smit and van Duin 1993). The maturation protein is not able to be translated, as there are long-range interactions that pair the bases of the ribosome binding site, closing off access to it (Beekwilder, Nieuwenhuizen et al. 1996, Poot, Tsareva et al. 1997). There is also evidence that the maturation protein is not detectable during the initial stages of infection, whether because the methods used were not sufficient to detect such low quantities of protein or because the secondary structure of the gRNA inhibits synthesis of the maturation protein, is unknown. Ribosomes are also not able to access the ribosome binding site of the replicase (or in the case of MS2, that for the gene of the lysis protein), due to the local secondary structure of the gRNA, which is highly branched (Berkhout, Schmidt et al. 1987). The only free ribosome binding site in either Qβ or MS2 is that of the coat protein (Smit and van Duin 1993). As the coat protein is translated the ribosomes disrupt the secondary structure of the gRNA enabling other ribosomes to access the newly freed ribosome binding site of the replicase (Berkhout and van Duin 1985).

From the initial gRNA which makes its way into the cell, the coat protein and replicase are translated until there are sufficient levels of the replicase to initiate synthesis of the negative strand of RNA. Perhaps due to the limited coding capacity

within the ssRNA phages or the coevolution of the phage with the host, the viral encoded subunit of the RNA-dependent RNA polymerase is insufficient to replicate the virus gRNA on its own (Blumenthal and Carmichael 1979). The so-called β-subunit of the replicase holoenzyme is what is commonly referred to as the replicase. The replicase hijacks ribosomal protein S1 as the α-subunit and translation factors EF-Tu and EF-Ts as the ɣ and δ subunits, respectively (their Greek abbreviations are based on the size of the protein, with S1 being the largest component of the holoenzyme).

There are two sites within the gRNA of Qβ and MS2-like phages that recruit S1, the S- and M-sites (Miranda, Schuppli et al. 1997). These sites will be discussed at length, later on, what should be stated here is that the S-site is just upstream of the gene for the coat protein and the M-site is within the RNA for the replicase. As these sites recruit the S1 protein to the viral gRNA they can also engage the ribosome. If the ribosome goes to the M-site, nothing will happen, as there are no ribosome binding sites downstream of it (Beekwilder, Nieuwenhuizen et al. 1995). Once S1 recruits the ribosome to the S-site, translation can begin for the coat protein and afterward the replicase (Miranda, Schuppli et al. 1997).

As replication proceeds 3’ to 5’, while translation goes 5’ to 3’, replication of the gRNA cannot proceed until there is a sufficient level of the replicase to bind to the S-site, preventing ribosome binding (Vasilyev, Kutlubaeva et al. 2013). Once a replicase holoenzyme binds the S-site, it becomes inaccessible for more ribosomes to bind and

initiate translation of the coat and replicase. Synthesis of the negative strand can proceed once the ribosomes detach from the gRNA. It is not known whether the replicase holoenzyme which bound the S-site slides down the gRNA until it reaches the 3' end to start replication or if another replicase initiates synthesis at the 3' end. In Qβ there are long-range interactions that bring the M-site proximal to the 3' end, facilitating the initiation of replication (Klovins, Berzins et al. 1998). Binding and release of the gRNA are the rate-limiting steps in replication, not the addition of nucleotides to a transcript, that happens extremely fast, with the 4,217nt genome replicated in ~2 minutes (Haruna and Spiegelman 1965).

The synthesis of negative-strand RNA requires the complete holoenzyme, as well as in the case of Qβ, Hfq (host factor for Qβ). Hfq is an RNA chaperone that in the host is used for pairing small RNAs for post-transcriptional regulation (Kavita, de Mets et al. 2018). In Qβ RNA replication the role of Hfq is unknown but is required for plus-strand synthesis from a negative-strand template (Barrera, Schuppli et al. 1993). Perhaps this is due to the fact that S1 is not required for plus-strand synthesis from a negative-strand template (Chetverin 2018).

Once the replicase is made and the gRNA is replicated into the negative-sense, the phage growth is near exponential. The negative-sense gRNA does not encode any essential proteins, although it has never been tested whether it encodes any gene products which might play a role in infection. As the negative-sense gRNA presumably

is not translated, it is free to be replicated repeatedly, whereas the positive-sense gRNA is in constant demand by both the replicase holoenzyme and the ribosome, leading to a 10:1 ratio of positive-sense to negative-sense gRNAs (Billeter, Libonati et al. 1966). The first replicase holoenzymes detected in cell extracts were from f2 and MS2 infections but further characterization was hindered by the instability of the complexes (August, Cooper et al. 1963, Haruna, Nozu et al. 1963). A few years later Spiegelman's group was able to purify the enzymatic complex from cells infected with Qβ (Haruna and Spiegelman 1965).

1.3.3. Overview of RNA Replication within a Qβ-infected Host

The RNA-dependent RNA polymerase (RdRp) from Qβ is one of the best characterized RdRps and certainly is the most studied of all ssRNA phages. Once incoming viral gRNA is translated the β-subunit of the replicase will hijack host proteins to form the holoenzyme RdRp (Takeshita, Yamashita et al. 2014). As stated earlier, the RdRp is composed of the ssRNA phage-encoded β-subunit and three host factors: ribosomal protein S1, and the translational elongation factors EF-Tu and EF-Ts (Kamen 1970, Kondo, Gallerani et al. 1970, Blumenthal and Carmichael 1979, Kidmose, Vasiliev et al. 2010, Takeshita, Yamashita et al. 2012, Gytz, Mohr et al. 2015). Once fully assembled the holoenzyme can replicate the positive-sense Qβ gRNA (plus strand) with the help of host factor Hfq (Su, Schuppli et al. 1997). The mechanistic details of the required protein interactions will be discussed below. Once the plus strand is replicated into the complementary negative-sense (minus) strand further replication back

to the plus strand does not require Hfq, or even S1 (Blumenthal, Landers et al. 1972, Kamen, Kondo et al. 1972, Landers, Blumenthal et al. 1974). Interestingly, the ratio of plus-strand to minus strand gRNA is 10:1 (Billeter, Libonati et al. 1966), this is due to competition between the RdRp and the ribosome for plus-strand gRNA, whereas the minus strand is not thought to be translated, so it can continuously be replicated (Ahlquist, Noueiry et al. 2003), although this imbalance has also been attributed to a lack of Hfq (Kamen 1975, Kuo, Eoyang et al. 1975).

S1 recognizes two sites in the positive-sense gRNA, the M-site within the replicase, and the S-site just upstream of the coat protein (Senear and Steitz 1976), although the S-site may be dispensable for replication (and personal communication from Charlotte Knudsen). The S1-site may just increase the affinity for the ribosome, as S1 is a part of the ribosome, and the ribosome-binding site upstream of the coat is weak (Friedman, Genthner et al. 2009). However, the coat protein is the highest translated phage product (Weissmann 1974), so this could be a mechanism to increase the recruitment of the ribosome to that region of the gRNA. S1 in complex with the replicase binding to the S-site is also a potential source of translational repression, as the RNA cannot be replicated and translated in the same site at the same time (Weber, Billeter et al. 1972). It is thought that there are long-range interactions in the gRNA of Qβ which brings the M-site closer to the 3' end of the genome in order to increase the rate of gRNA replication (Klovins, Berzins et al. 1998), which is limited not by the

speed of replicating RNA but by binding and releasing RNA templates (Chetverin 2018). The Qβ replicase can copy its genome in ~2 minutes (Weissmann, Feix et al. 1968).

The Qβ replicase is the best-studied ssRNA phage RdRP and among the best-studied RdRps. RdRps are error-prone, as a result of not having a duplicate strand of genetic information to serve for proofreading (Moustafa, Korboukh et al. 2014). Error rates are as high as 10^{-3}/nt copied, or about one mutation per genome per round of replication (Garcia-Villada and Drake 2013). This has led to the 'quasispecies' concept for viruses, wherein the population as a whole is highly heterogeneous but the defined average of the population is referred to as the wild-type sequence (Domingo, Sabo et al. 1978). All ssRNA viruses, including ssRNA phages, are quasispecies, with certain subsets of the population that are more or less infectious. This can be seen in the continually evolving flu virus, for which new vaccines are needed every year. A high mutation rate is advantageous to viruses that produce large numbers of progeny, many of which may never go on to infect another host, but those which do successfully can pass on any positive traits.

1.3.4. Co-evolution of Phage and Host

ssRNA phages are thought to have derived from a common ancestor, based on the genome structure, which in every instance consists from 5' to 3' of the maturation protein, coat, then the replicase, which have high homology to other ssRNA phage replicases (Olsthoorn and van Duin 2011). The divergences come particularly in the

lysis genes, which are spread throughout the genome in different phages, but also in the requirements for a minor capsid protein (A_1 for Qβ and SP-like phages) (Inokuchi, Jacobson et al. 1988), and the host factor requirements for binding to the positive-sense genome, where Qβ differs from MS2 (Rumnieks and Tars 2018). The maturation proteins naturally have differences in their pilus-binding region, within the β-region of the maturation protein (Gorzelnik, Cui et al. 2016, Dai, Li et al. 2017, Rumnieks and Tars 2017). There have been many discussions on the evolution and divergence of ssRNA phages (Zinder 1980, Mekler 1981, van Duin 1988, Bollback and Huelsenbeck 2001).

In order to determine the mechanism of A_2-mediated lysis by Qβ, Bernhardt and colleagues took a genetic approach wherein they overexpressed A_2 from an inducible plasmid and plated cultures for surviving colonies (Bernhardt, Wang et al. 2001). Two surviving colonies (out of ninety) were subsequently found to be resistant to Qβ, one of which was reported to accumulate viral particles at a rate comparable to wild-type cells during an infection, indicating that the defect allowing the cells to survive the initial screen was not in A_2 expression but in A_2-mediated lysis. The authors subsequently mapped the mutation to murA, which produces a conserved protein within cell wall biosynthesis, and identified the change as L138Q, on the substrate-binding cleft of MurA. This mutant, termed as RAT (resistant to A_2), was found to be dominant as if $murA^{rat}$ was expressed on a plasmid with basal levels of transcription it would prevent lysis of the cells during infection (Bernhardt, Wang et al. 2001). The authors

subsequently showed that Qβ virions were able to inhibit a crude preparation of MurA, but not $MurA^{rat}$.

While $murA^{rat}$ prevents A_2-mediated lysis, Qβ can subsequently evolve to overcome this mutation, particularly fast due to the error-prone replicase. A subsequent study identified eight mutations, termed *por* mutants (*p*lates *o*n *r*at), which were able to infect bacteria with the $murA^{rat}$ allele (Reed, Langlais et al. 2013). These mutants had missense mutations (Table I-1) could have impacted the binding interface around the RAT mutation, but were found to increase the rate of A_2 translation, with the three common mutants reported having drastically increased levels of A_2.

Table I-1. Qβ *por* mutations
Amino acid mutations are shown, with the particular nucleotide changes immediately below. Some mutants had multiple mutations, which are listed in the cells below. Adapted from Reed *et al.*, 2013.

Qβ A_2^{por} mutations (AA and nt changes)							
1	2	3	4	5	6	7	8
D52N GAU> AUU	E125G GAG> GGG	D52N GAU> AAU	L28P CUC> CCC	D52N GAU> AAU	D52N GAU> AAU	D52N GAU> AAU	L28P CUC> CCC
	F412S UUU> UCU	P66P CCA> CCG					
		H67R CAU> CGU					

A look at the secondary structures reveals that the nucleotide changes disrupted secondary structure stability of stems loops in A3 (L28P, CUC>CCC); A4 (D52N, GAU>AAU; P66P, CCA>CCG; H67R, CAU>CGU), A8 which is near the long-range

interaction base-pairing the Shine-Dalgarno sequence (E125G, GAG>GGG), and A19 which is within the S-site (F412S, UUU>UCU). All of the mutations disrupt base pairing, suggesting that less stable secondary structure leads to increased translation of the maturation protein. It is possible to envision ribosomes backed up on the gRNA like cars on a road, and if there are fewer speed bumps on the road then the cars will not break as often, enabling traffic to flow faster. The stem-loop most affected was A4, with 5/8 mutants having mutations in it, and one of those five has three mutations in it. Two other mutations L28P (CUC>CCC) and F412S (UUU>UCU) were found in biological studies of Qβ, wherein the authors were looking for mutations which affected the fitness of Qβ, increasing temperatures (F412S, U1295C) (Arribas, Cabanillas et al. 2016) or increased ability to produce offspring (L28P, U128C) (Garcia-Villada and Drake 2013). Increasing maturation protein levels would make sense in both circumstances. In the first condition, growing the bacteriophage at a higher temperature would require a higher level of maturation protein if it is thermodynamically unstable, or if more MurA is needed at that temperature. In the second condition, selecting for increased fecundity (viable progeny), if the virus is selected for mutants that produce more offspring there could be a need for more maturation protein to bind to the gRNA at a higher ratio. Or it could be that having more maturation proteins prevent fully mature viruses from being inactivated, as Reed and colleagues found that MurA binding to A_2 caused the gRNA to be released and degraded by RNases, so having more A_2 would enable MurA to bind free A_2, thus increasing the fecundity of the virus (Reed, Langlais et al. 2013).

ssRNA phage can co-opt their hosts without necessarily killing them if there are amber mutations in the maturation or lysis genes. There are several instances of Qβ being reported to have adopted a lysogeny-like state, wherein there was an amber mutation in A_2 (maturation/lysis protein) which prevented lysis in a non-permissive host but did not prevent replication and packaging of the gRNA (Watanabe, Sakurai et al. 1979). The phages were grown in the non-permissive host and within this host produced viable particles at what the authors calculated to be 10^{-3} PFU/cell, which is along the lines of the mutation rate for the replicase.

1.4. General Characteristics of ssRNA Phages

RNA phages are split into two classes, single- and double-stranded (Krishnamurthy, Janowski et al. 2016). The ssRNA phages are simpler, with the coding capacity limited by their small size, ~3,500-4,300nt for phages with known hosts (Rumnieks and Tars 2018) and potentially up to ~5,000nt, although the host for this phage is unknown (Krishnamurthy, Janowski et al. 2016). The replicase, along with the maturation protein and coat protein, are conserved in all ssRNA phages (Rumnieks and Tars 2018). They need to be able to copy their own gRNA (replicase), then package the gRNA to protect it from RNases (coat protein), finally, they need a protein that can bind the host and be pulled into the host cell along with the RNA (the maturation protein). The other genes within these phages are more varied, some have an elongated coat protein, while others have lysis proteins that are found in all regions of the genome, except after the replicase (K. Chamakura, personal communication).

The viral particles of ssRNA phages contain one copy of gRNA bound to the maturation protein, presumably non-covalently since the maturation protein can dissociate from the RNA once inside the cell (Shiba and Miyake 1975). The gRNA is further encapsidated by 178 coat protein monomers, in the form of 89 coat dimers, although later in this dissertation I will show that Qβ has an internally sequestered coat protein dimer (Cui, Gorzelnik et al. 2017). Outside of the host, the only known role of the coat protein is to protect the genome from RNases (Rumnieks and Tars 2018). Single-stranded RNA phages range from 25nm to 30nm and are quasi-icosahedral, with the symmetry of the capsid broken by a single copy of the maturation protein (Robertson and Lodish 1970, Dent, Thompson et al. 2013), with the gRNA asymmetrically organized inside (Gorzelnik, Cui et al. 2016, Dai, Li et al. 2017). It has been noted that there are ~1,000 molecules of spermidine within the capsid of ssRNA phages (Fukuma and Cohen 1975) which is thought to counteract the negative charge of the RNA, this density is blurred out in the cryoEM structures.

When ssRNA phages were first studied it was determined that the maturation protein was required for infection but through an unknown mechanism. Scientists were once unsure if it was on the surface of the virus or buried inside of the capsid, just that it was required for successful infection. Later it was learned that the maturation protein bound to the pilus, and in the case of certain phages also served as the lysis protein (Olsthoorn and van Duin 2011). Once the maturation protein and gRNA are taken up by

the host the gRNA serves directly as a template for the translation of proteins (Yin and Redovich 2018). The maturation protein may or may not be able to be translated at this point in time, because the gRNA might be folded in such a way upon entry into the cell that the ribosome binding site for the maturation protein is blocked. The maturation proteins are regulated at the translational level by long-range RNA interactions to prevent the ribosome from being able to bind to the Shine-Dalgarno sequence (Beekwilder, Nieuwenhuizen et al. 1996). The maturation protein can only be translated from nascent (newly synthesized) gRNA before it has had time to form tertiary RNA:RNA interactions (Beekwilder, Nieuwenhuizen et al. 1996).

Once the replicase is copying the gRNA the phage goes into as a 'hypercycle' or an 'explosive positive-feedback loop' (Yin and Redovich 2018) wherein more phage RNA enables there to be more translation of the replicase which will make more phage RNA allowing more replicase to be translated et cetera et cetera (Eigen, Biebricher et al. 1991). The replicase can copy virtually any RNA once, but if the RNA does not have enough secondary structures then the product strand will anneal to the template strand and the replicase cannot proceed through dsRNA of more than ~200nt (Tomita, Ichihashi et al. 2015). The gRNA for Qβ and MS2 is highly branched, an evolutionary adaptation to prevent product RNA from annealing to the template RNA during replication by the replicase (Tomita, Ichihashi et al. 2015).

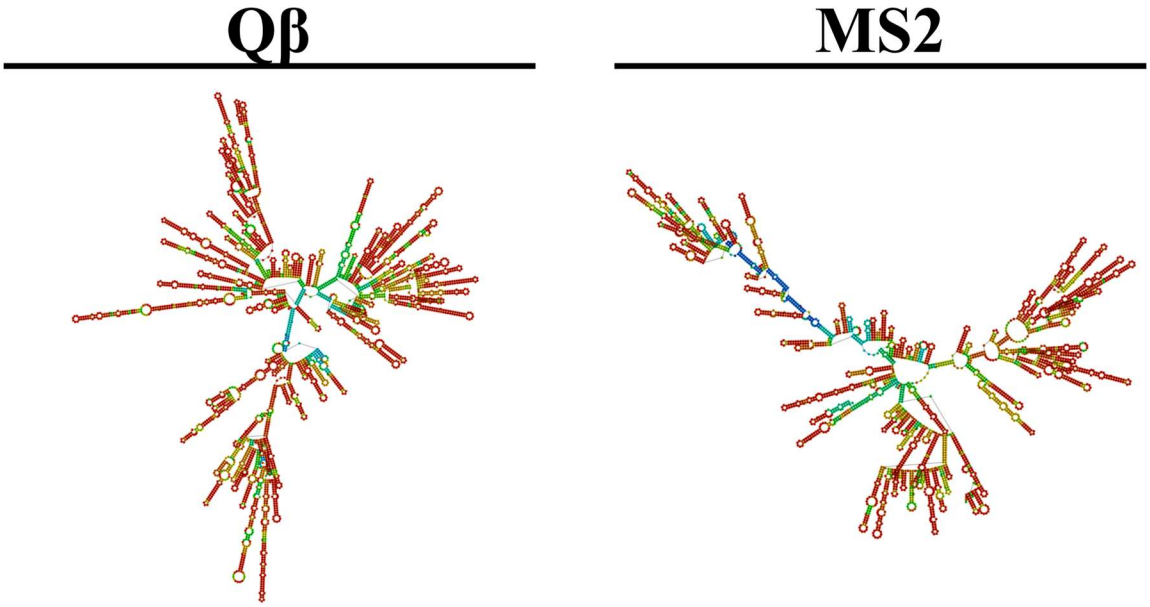

Figure I-6. The genomes of ssRNA phages are highly branched
The gRNA of Qβ and MS2 were put into RNAfold to predict the secondary structure, and show it in a concise image. The secondary structures for both phages are known and are more or less along the lines of what RNAfold predicts.

This phenomenon of non-specific RNA replication has been well characterized for Qβ, with the emergence of what is called 6S RNA, which is associated with Qβ infections (Banerjee, Rensing et al. 1969). This probably occurs for all ssRNA phage infections but was historically studied within Qβ. These RNAs are what have been termed parasitic RNAs, as they do not contribute to the 'hypercycle' but benefit from RdRp mediated replication (Bansho, Furubayashi et al. 2016). Parasitic RNAs develop from host RNAs and over time become more efficient at replication, losing gene functions that may have been encoded, as they are unnecessary for replication (Yumura, Yamamoto et al. 2017). While these parasitic RNAs will arise at each and every infection as a by-product of viral RNA replication via the RdRp, the odds are not good for the possibility of horizontal gene transfer, as is the case with newly released viral

particles (which are not truly transferring their genes from one host or another but rather reinfecting new hosts). As the hosts actively have to pull in the viral RNA, unless the parasitic RNAs can somehow bind the maturation protein and form a virus-like particle which can infect a host cell at the same time as a replicase-encoding phage, or what has been termed a 'satellite virus', then the parasitic RNA will not be propagated through subsequent hosts (a satellite virus of an RNA phage is an interesting theoretical concept though). The parasitic small RNAs which are replicated by the replicase would need to bind the maturation protein in order to reinfect new cells, because as has been stated before, all ssRNA phages with identified hosts require retractile pili, presumably to pull in the viral RNA. This has been seen from *E. coli* phages to *Pseudomonas* and *Caulobacter* phages (Rumnieks and Tars 2018).

Structurally ssRNA phages are quite similar, they range from 25-30nm, all have near icosahedral *T*=3 capsids composed of ~180 coat proteins (or in the case of *Allolevivirus*, the 180 coat proteins include 3-15 copies of the readthrough coat protein A_1) with a single maturation protein breaking the symmetry (Takamatsu and Iso 1982). The capsids for all ssRNA phages contain a single copy of the gRNA. The size of the capsid is based on the size of a coat protein monomer. The coat protein for MS2 is 130 amino acids for a monomer of 13.8kDa while the slightly larger Qβ coat protein is 133 amino acids for a monomer of 14.2kDa (Golmohammadi, Valegard et al. 1993). The capsids for the two phages correspondingly differ in size from 26nm for MS2 to 28nm for Qβ. The replicase for these phages are highly conserved and vary significantly from

RdRps from viruses of eukaryotes and archaea, in fact, it has been used as a means to identify ssRNA phage from metagenomic data (Krishnamurthy, Janowski et al. 2016, Callanan, Stockdale et al. 2020).

1.4.1. Classification of ssRNA Phages

f2, was the first ssRNA phage discovered, many subsequent RNA phages isolated were near-identical, and have been grouped together as MS2 (for Male Specific 2) (Loeb 1960, Davis, Sinsheimer et al. 1961, Loeb and Zinder 1961, Paranchych and Graham 1962, Marvin and Hoffmann-Berling 1963). A different form of ssRNA phage that was also dependent on F-pili was later discovered, Qβ (Watanabe and Okada 1964), differed serologically (Overby, Barlow et al. 1966), and later turned out to have different features, but had the same genomic organization (Weber and Konigsberg 1975). Classification of ssRNA phages was historically done first based on their host when most studied were those that infect bacteria with F-pili, notably *Escherichia coli* (Krueger 1969). ssRNA phages were found for other species such as *Caulobacter crescentus* and *Pseudomonas aeruginosa* (Bendis and Shapiro 1970, Shapiro and Bendis 1975), but these were the exception to the rule and were not studied as extensively as the *E. coli* phages MS2 and Qβ. To date, the only ssRNA phages with known hosts infect bacteria that possess retractile pili (Pumpens, Renhofa et al. 2016).

Some ssRNA phages isolated early on did not require pili encoded on plasmids, rather they infected via chromosomally-encoded pili. Examples of these phage are

Pseudomonas phages PP7 or 7s (Feary, Fisher et al. 1964) or *Caulobacter* phages such as φCb5 (Schmidt and Stanier 1965, Schmidt 1966, Bendis and Shapiro 1970, Miyakawa, Fukuda et al. 1976). Most subsequently discovered ssRNA phages do require plasmid-encoded pili: PRR1 (Olsen and Thomas 1973), T (Bradley, Coetzee et al. 1981), C-1 (Sirgel, Coetzee et al. 1981), Iα (Coetzee, Bradley et al. 1982), M (Coetzee, Bradley et al. 1983), D (Coetzee, Bradley et al. 1985), pilHα (Coetzee, Bradley et al. 1985), Hgal1 (Nuttall, Maker et al. 1987), although *Acinetobacter* phage AP205 infects via chromosomally encoded pili (Klovins, Overbeek et al. 2002).

ssRNA phages of *E. coli* (coliphages) were initially classified into four groups based on virion size, buoyant density, UV sensitivity, RNA size, and RNA base composition, then were classified based on their serological relationship to each other (Schmidt and Stanier 1965, Sakurai, Miyake et al. 1968, Krueger 1969). Eventually, coliphages were split based on the template requirements for the replicase holoenzyme (Miyake, Haruna et al. 1971). The best-known coliphages for each group are MS2 (group I), GA (group II), Qβ (group III), and SP (group IV). Upon sequencing of these and other phages, (Fiers, Contreras et al. 1976, Mekler 1981, Inokuchi, Takahashi et al. 1986, Inokuchi, Jacobson et al. 1988) it was determined that many phages from each group were more or less the same, varying at the nucleotide level but producing essentially the same proteins. This has been called the 'quasispecies' theory of RNA viruses, wherein the population as a whole has an average sequence which is selected

for, there can be a large variation within the species, as long as traits leading to replication and infection are preserved (Domingo, Sabo et al. 1978).

1.4.2. Phylogeny of ssRNA Phages

Eventually, the four groups of coliphages were broken down into two genera: *Levivirus* (groups I and II) and *Allolevivirus* (groups III and IV), within the *Levidiradae* family (Friedman, Genthner et al. 2009). Some other phages were placed in *Levivirus* based on sequence homology, such as *Pseudomonas* phage PP7 (Kannoly, Shao et al. 2012), or just left out as separate members of Leviviridae. Phages were placed into one of these two genera based on their genomic organization, with the most noticeable features being the presence of a leaky UGA stop codon in group B phages (also called groups 3 and 4 phages or *Allolevivirus*) which is read as a tryptophan 5% of the time leading to an elongated coat protein called A_1 (Weiner and Weber 1971, Hofstetter, Monstein et al. 1974), and the absence of a dedicated lysis protein, which is found in members of group A phages (groups 1 and 2, *Levivirus*). Subsequent comparisons of *Levivirus* and *Allolevivirus* will be based on their best-studied members, MS2 and Qβ, respectively (van Duin and Tsareva 2006).

In the heyday of RNA phage research most of the ssRNA phages infecting bacteria other than *E. coli*, notably *Pseudomonas* and *Caulobacter*, were just described as other than the *E. coli* phages (Shapiro and Bendis 1975). The phages of these genera are also pili dependent and contain a similar genome structure, with the maturation

protein being a minor component of the capsid, the major component being the coat protein, also encoding a non-encapsidated replicase and lysis protein (Fiers 1979). The dependence of ssRNA phages on pili for entry into their host is exemplified by the broad-host-range phage PRR1, that will infect many genera of bacteria, such as *E. coli, Pseudomonas, Salmonella, and Vibrio*, provided they express pili encoded by the R-factor (Olsen and Thomas 1973). In fact, non-piliated enterobacteria, such as Proteus, Salmonella, or Shigella, can be artificially converted ssRNA phage sensitivity if the F-factor (a plasmid containing the operon for F-pili) is introduced (Zinder 1965).

The initial *Pseudomonas* and *Caulobacter* phages, PP7 and φCb5, respectively, were also found to be dependent on retractile pili (Schmidt and Stanier 1965, Shapiro and Bendis 1975), as was the later identified *Acinetobacter* phage AP205 (Klovins, Overbeek et al. 2002). To the best of my knowledge, there is no evidence that the maturation proteins of these phages enter the cell. However, the paradigm that this was the case was established for all ssRNA phages, based on work in MS2 that the maturation protein and gRNA are taken up into the host during infection (Krahn, O'Callaghan et al. 1972, Shiba and Miyake 1975).

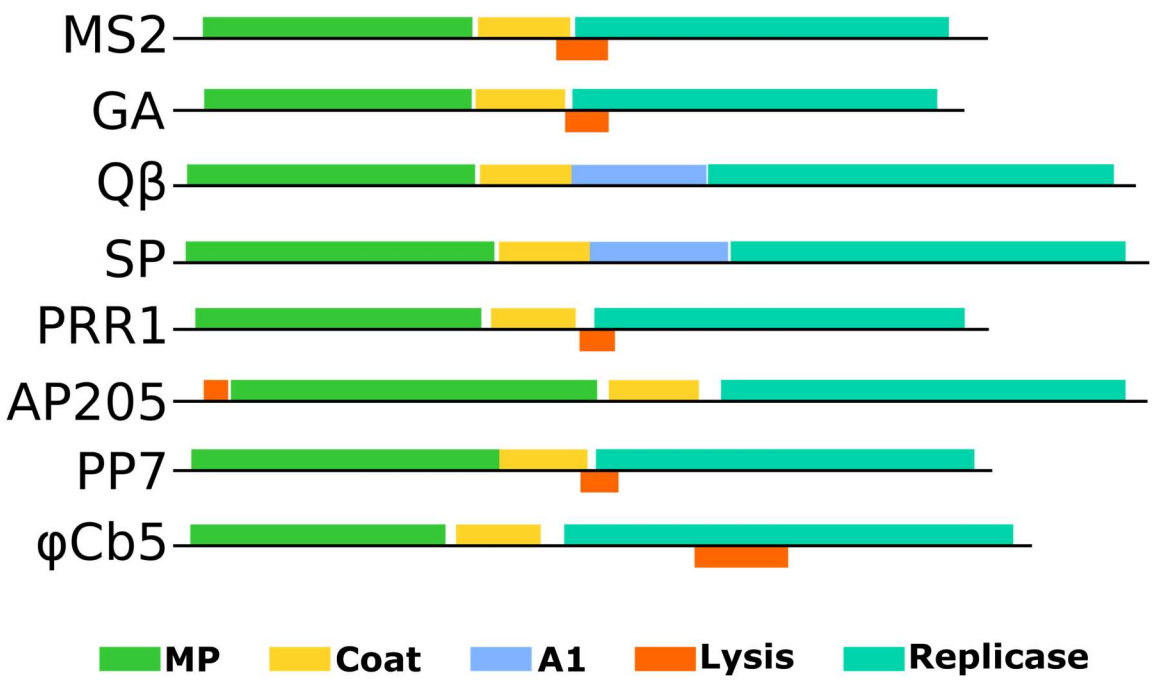

Figure I-7. Genome organization for the best studied phages within *Leviviridae*
The top four phages are the prototypical phages within the four defined groups of *Leviviridae* which infect F^+ *Escherichia coli*, with MS2, GA, Qβ, and SP being the type phages for groups 1, 2, 3, and 4, respectively. Groups 1 and 2 are also called group A phages, while groups 3 and four are called group B phages. PRR1, AP205, PP7, and φCb5 are unclassified within *Leviviridae*. The genomes are to scale. The gene for the maturation protein is in green, the gene for the coat protein is in yellow, with the readthrough of a leaky stop codon for group B phages yielding a coat protein with a C-terminal extension shown in blue, the lysis protein is in orange, and the replicase is in teal. The maturation protein for group B phages also serves as the lysis protein.

Groups 1-4 phages bind to F-pili, predominantly studied in *E. coli* (van Duin and Tsareva 2006). PRR1 infects through P-pili, mostly studied in *Pseudomonas*, but the plasmid can be found in many species (Shapiro and Bendis 1975). For the longest time there were *E. coli* phages, mostly classified into one of the four groups, or other than *E. coli* phages, with those just classified as *Leviviridae* (Kannoly, Shao et al. 2012, Rumnieks and Tars 2012, Callanan, Stockdale et al. 2018).

A study by Krishnamurthy *et al.* revealed that the initial lack of diversity in ssRNA phages may not be as great as was thought (Krishnamurthy, Janowski et al. 2016). By searching existing metagenomic datasets for sequences that were similar to the known ssRNA phage proteins, particularly the RdRp but also the maturation protein or coat proteins, the authors were able to identify 120 highly diverse sequences (Krishnamurthy, Janowski et al. 2016). More recently, another group used the same method to identify >1,000 ssRNA phage genomes with the three core genes, the MP, coat, and RdRp, and >15,000 with at least one of the three (Callanan, Stockdale et al. 2020). The RdRp of ssRNA phage is divergent from RdRps of viruses infecting other forms of life, i.e. eukaryotes and archaea, to make their use as a taxonomic identifier reliable. Many sequences with homology to ssRNA phage proteins such as the maturation, coat, RdRp, or known lysis proteins, contained genes whose coding features contain no known homologs (Krishnamurthy, Janowski et al. 2016, Callanan, Stockdale et al. 2020). This is presumably due to the great diversity of lysis proteins, as it appears that every time one of these phage jumps to a new host they develop a new lysis protein (Chamakura and Young 2019).

1.4.3. Ecology of ssRNA Coliphages

Since the initial ssRNA phages were isolated, they have been found all throughout the world (Zinder 1965, Krishnamurthy, Janowski et al. 2016, Callanan, Stockdale et al. 2020). ssRNA phages are most frequently found in sewage and the feces

of mammals, with titers approaching 10^7 PFU/ml (van Duin 1988). There are geographic distinctions within the isolation of different groups of ssRNA phage, with preference for one group over another, although the variations could be due to the particular sewage facilities used for the study in one particular area (Furuse, Ando et al. 1981, Osawa, Furuse et al. 1981, Osawa, Furuse et al. 1981). It was suggested that there is a north-south gradient for the isolation of ssRNA phage, with group III (as well as I and IV) propagating well at 40°C but not at 20°C, whereas the reverse is true for group II (Furuse, Ando et al. 1981). As groups I and II belong to *Levivirus* and groups III and IV are in *Allolevivirus*, there can be no determination on whether one genera or the other is better adapted to certain temperatures. While coliphages have been isolated from a variety of mammals, the exact host organism is unknown; however, ssRNA phages have been shown to propagate stably in the intestines of gnotobiotic (germ-free) mice if *E. coli* is present, indicating that *E. coli* can 'naturally' sustain phage growth (Ando, Furuse et al. 1979). ssRNA phages also have a natural habitat in sewage and they have been used as an indicator species to monitor sewage treatment, as far as the efficiency of virus inactivation or removal (Moce-Llivina, Muniesa et al. 2003).

1.5. Genome Organization of ssRNA Phages

The genomes for ssRNA phages often only encode three to four proteins (Figure I-8). The gRNA has several untranslated regions (UTRs), most notably stretches of at least 60 nucleotides at the 5' and 3' ends, which play a role in replication, translational control, and, later in this dissertation, the role of the 3' UTR will be discussed in terms

of packaging and maturation of the virion. Whether because MS2 has more genetic space available, due to the absence of the readthrough protein in Qβ, or if MS2 has had more time to specialize the packaging of its genome, the UTRs for MS2 and members of *Levivirus* are larger than the UTRs of Qβ and the *Allolevivirus* (Figure I-8). Again, in a subsequent chapter the experimental implications are discussed from this phenomena.

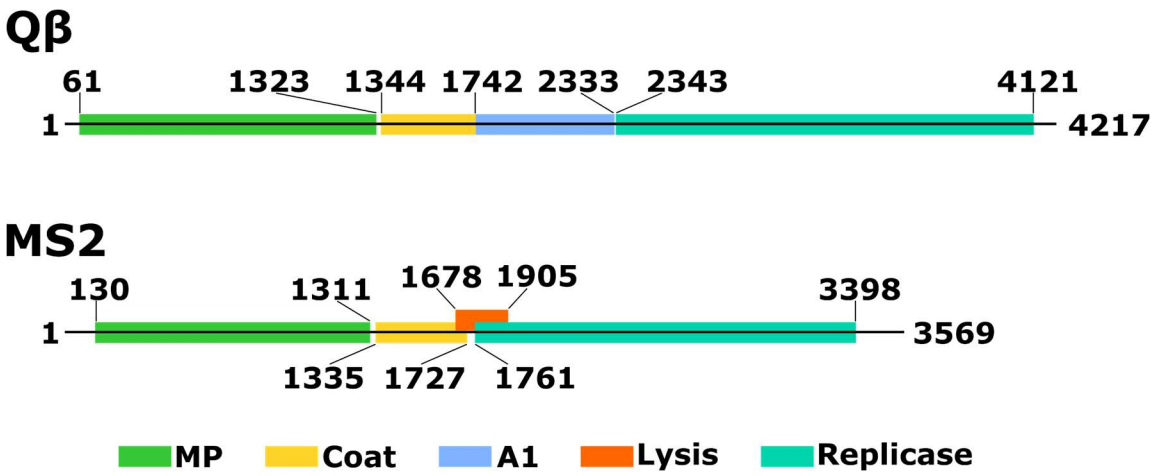

Figure I-8. Genome organization and nucleotide numbers for Qβ and MS2
The genomes are drawn to scale, with the nucleotide numbers shown to illustrate the 5' and 3' untranslated regions (UTRs) for Qβ are smaller than those of MS2. The genes are colored as before, with a legend at the bottom.

Other regions which are untranslated play a role in the initiation of translation (ribosome binding sites) and regulation of protein synthesis (notably the coat protein binding sites, termed 'operator,' at the start of every replicase) (Beckett and Uhlenbeck 1988) or replication (M-site within the replicase) (Klovins, Berzins et al. 1998). Later in this dissertation evidence will be presented that the 3' UTR plays a role in packaging and delivery. While much more is known about the function of the operator than say the 5'

and 3' UTRs, single point mutations in these regions can be lethal for a phage or confer a severe disadvantage, which means the phage retains them for a reason and this has been proposed to be to control translation and replication (Flavell, Sabo et al. 1975, Domingo, Flavell et al. 1976, Sabo, Domingo et al. 1977, Taniguchi, Palmieri et al. 1978). As a community, we have subsequently learned that a major portion of the 5' UTR in Qβ and MS2 is dedicated toward translational control of the maturation protein (Beekwilder, Nieuwenhuizen et al. 1996, Poot, Tsareva et al. 1997). The first stem-loop in the genome, A1 (not to be confused with A_1, the readthrough product of the coat protein from Qβ) is conserved among ssRNA phages (Beekwilder, Nieuwenhuizen et al. 1996). This stem-loop has been proposed to be important in positive-strand gRNA synthesis, as it forms secondary structure right away, so as to not bind back on the negative-sense template, which would inhibit further rounds of replication.

There are cases where the phage encodes multiple proteins with the same coding sequence, with the minor coat protein of Qβ, A_1, or the case for many lysis proteins, such as that of MS2, which is encoded in a different reading frame from the coat protein and replicase (Fiers, Contreras et al. 1976, Fiers 1979). Krishnamurthy and colleagues identified many potential ORFs within ssRNA phages found within metagenomic data which would either be novel lysis proteins or have unknown functions, as they are not found in the well studied ssRNA phages (Krishnamurthy, Janowski et al. 2016).

There may be open reading frames, but these products are often small, with no homology to other proteins, and while may play a role in another host, are not essential for growth in the laboratory host. A group reported the presence of a cryptic lysis gene in Qβ (Nishihara, Morisawa et al. 2004), although this has not been subsequently followed up. During Qβ infection of *E. coli*, only three genes encoding four proteins are required. There are many more potential open reading frames (Mekler 1981), some of which have ribosome binding sites, but none are essential for infection and lysis in *E. coli*, although some might lyse other hosts. When growing f2 (an MS2-like ssRNA phage) in *Bacillus stearothermophilus* extracts at different temperatures Lodish and unnamed colleagues noticed that at higher temperatures there were several proteins expressed that were not present when expressing at lower temperatures (Lodish 1971). The author(s) attributed this to reduced gRNA secondary structure, opening up sites within the RNA which would not have otherwise been accessible to ribosomes.

The secondary structure of the genome may not be as permanent as we think, Kramer and Mills found when the Qβ replicase was used to replicate a famous satellite RNA, MDV-1, the structure had transient properties which reformed new helices during later stages of RNA elongation (Kramer and Mills 1981). The Qβ gRNA is much longer than MDV-1 but has many regions of long-range interactions, which until they are base paired with sequences much further downstream could allow for transient interactions. Transient helices present during RNA replication but absent in the fully completed structure may play a role in preventing the product RNA from annealing back to the

template RNA (Kramer and Mills 1981), which has subsequently been shown to inhibit RNA replication using an RNA-dependent RNA polymerase (Tomita, Ichihashi et al. 2015).

The RNA encoding the maturation proteins for ssRNA phages is often cited for the interactions which repress maturation protein expression once their downstream regions can base pair. When the positive-sense gRNA is replicated from the negative-sense gRNA the regions encoding the Shine-Dalgarno sequence for these proteins is unpaired and able to bind ribosomes (Beekwilder, Nieuwenhuizen et al. 1996, Poot, Tsareva et al. 1997). Once the gRNA is replicated enough so that the downstream regions bind to their complementary upstream regions the ribosomes can no longer efficiently access the Shine-Dalgarno site for the maturation protein, and translation rates drop dramatically. There could be some transient interactions in this region, but we would never know unless we looked for them. In any event, the regulation of the maturation proteins has been worked out quite extensively by this scenario, but it could play a factor in accessing these cryptic open reading frames. While there are potential Shine-Dalgarno sequences available for many of these genes, they are non-essential for growth in *E. coli*. The possibility exists that one or more of these genes is a lysis protein for a yet undiscovered host. To my knowledge, no one has gone hunting for additional hosts for Qβ.

When the entire genome of MS2 was sequenced, Fiers *et al.* noticed that there was an absence of the AUA codon for isoleucine in the coat protein, which is made exclusively of the codons AUU and AUC, whereas AUA is present in much higher levels in the genes for the maturation protein and replicase (Fiers, Contreras et al. 1976). This is thought to ease translation of the highest translated protein by reducing the necessity of rare tRNAs. The positive or negative selection pressure for tRNAs is thought to 'brake' or 'accelerate' protein synthesis (Weissmann, Billeter et al. 1973). This has been called translational modulation, wherein the presence of rare codons within a gene further enhances the effect of low rates of initiation for those genes, i.e. the maturation protein, lysis protein, and replicase (Gussin 1966, Lodish and Zinder 1966, Robertson and Lodish 1970, Kolakofsky and Weissmann 1971).

Early on during infection, the Qβ RNA serves as messenger RNA, to be translated. However, there is competition between ribosomes and the replicase for template RNA, as the ribosomes and replicase move in opposite directions down the RNA, only one can be actively working at a time, as the replicase cannot displace actively translating ribosomes (Kolakofsky and Weissmann 1971). While S-site binding by S1 may not play a role in replication, in complex with the replicase S1 binding to the S-site certainly helps reduce the number of ribosomes that can begin translation of the most produced protein product, the coat protein (Kolakofsky and Weissmann 1971). Ribosomes do not attach to the ribosome binding site ahead of the replicase when the

translation of the coat protein has ceased (Gussin 1966), which means that all translation has stopped and the gRNA is able to serve as a template for replication.

The Qβ RNA-dependent RNA polymerase holoenzyme (RdRp) is perhaps the best-studied RdRp and was so well characterized because it was able to be stably purified and was active with artificial templates as well as its own genome, unlike the replicase from MS2 (Weissmann 1974). Once the replicase is synthesized it recruits three host factors to form a holoenzyme capable of replicating the viral gRNA (Haruna and Spiegelman 1965). The components of the replicase were initially named based on their size. The phage-encoded protein is referred to as the "β-subunit' as it is the second-largest component, after ribosomal protein S1 (Wahba, Miller et al. 1974). The other two proteins which form the holoenzyme are translation elongation factors EF-Ts and EF-Tu (Blumenthal, Landers et al. 1972), whose normal functions inside the cell is to bind amino-acyl tRNAs and deliver them to the ribosome (EF-Tu). The ribosomal protein S1 is a translation initiation factor that contains six consecutive oligonucleotide binding (OB) domains, of which the first two N-terminal domains bind to the ribosomal small subunit, or in the case of the RdRp bind to the β-subunit of the replicase (Gytz, Mohr et al. 2015). Only those two domains of S1 are required for replicase activity (Chetverin 2018). This was visualized recently in a crystal structure of the complex (Kidmose, Vasiliev et al. 2010). The S1 protein is not required for the enzymatic activity of the "replicase core" consisting of the β subunit, EF-Tu, and EF-Ts (Kamen 1970). The replicase core can form a dimer in solution, and the salt-bridges responsible

for the dimerization are crucial for activity, indicating that this may be important for the proximal binding of the gRNA during replication (Kidmose, Vasiliev et al. 2010).

S1 has had many predicted functions over the years. S1 was initially determined to be required to recognize the phage gRNA in order to initiate replication (Kamen, Kondo et al. 1972). The sites where S1 binds to the gRNA are called the S-site and M-site (Weber, Billeter et al. 1972). Although there was a report in the mid-'70s of S1 binding the 3' untranslated region (ref), this has not been confirmed by a separate group and has not been mentioned in the literature in decades. The S-site is located upstream of the sequence for the coat protein. In Qβ this sequence is at nucleotides 1248-1347, occupying the 3' end of the gene encoding the maturation protein and the start codon of the gene for the coat protein/A_1 (Miranda, Schuppli et al. 1997). The S-site is named as such not because it recruits S1 when it is in complex with the ribosome or replicase, but rather binding to this high-affinity site requires moderate salt concentrations but does not require magnesium ions (Meyer, Weber et al. 1981). This site is not required for replication (ref, and Charlotte Knudsen, personal communication) but has been shown to competitively balance the binding of the ribosome and that of the replicase, as replication cannot proceed on a strand which is actively being translated (Kolakofsky and Weissmann 1971). While there is not a canonical strong ribosome binding site ahead of the coat gene: in Qβ it only has a GGG 11nt away from the start codon, rather than an AGGAGG 8nt away from the start codon (Friedman, Genthner et al. 2009), the coat is the highest translated protein of the ssRNA phages. The translation from a weak

RBS is presumably enhanced by the ribosome interactions with the S-site via S1 and either sliding down the RNA or binding and releasing the S-site then by proximity interacting with the RBS ahead of the coat gene. The stability of the RNA structure at the ribosome binding site is thought to determine access to the downstream genes, inasmuch as the secondary structure at that site can prevent the ribosome from binding (Schmidt, Berkhout et al. 1987, Skripkin, Adhin et al. 1990).

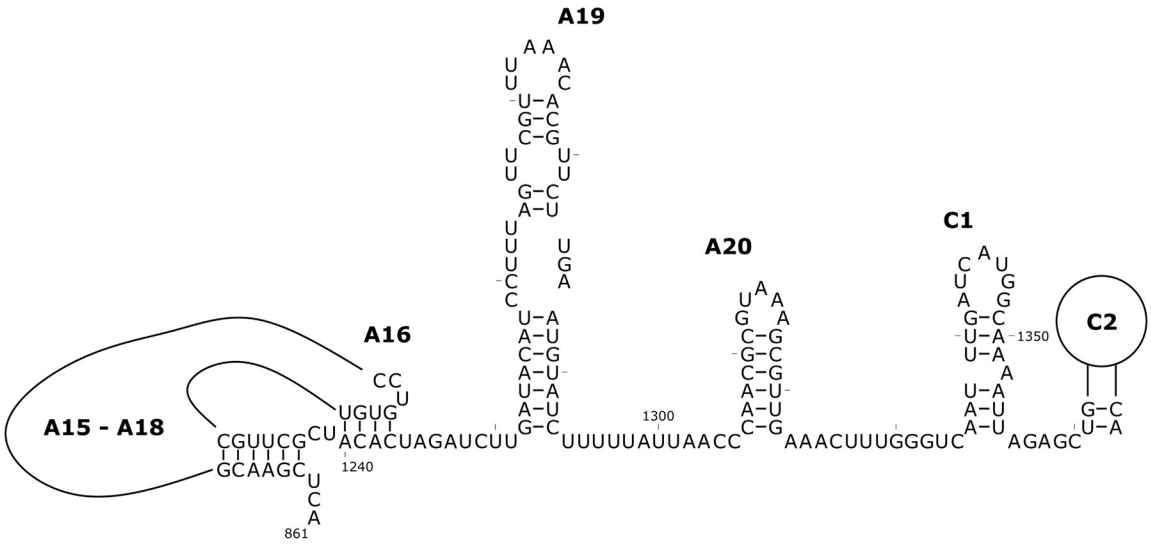

Figure I-9. The RNA secondary structure of the S-site for Qβ
The S-site, for salt-dependent site, of Qβ lies at the 3' end of the gene for A_2 and is immediately upstream of the coat gene. The RNA stem-loops are labeled as in Jules Beekwilder's papers, while the figure is adapted from his dissertation.

While the RBS ahead of the coat is a relatively weak one, it is always free. For Qβ and MS2, the only ssRNA phages whose RNA secondary structures are known, the region immediately upstream of the coat protein is relatively unstructured. The rest of the genome is high in secondary structure, and there have been a number of papers on

how the maturation proteins for each of these phages are controlled by long-distance interactions (Beekwilder, Nieuwenhuizen et al. 1996). In folded phage gRNA the only translation initiation site that is available for translation by ribosomes is that of the coat gene, while the maturation and replicase genes are buried in secondary structure (van Duin and Tsareva 2006). The M-site is named as such for its requirement of magnesium ions in order for the replicase to bind to and show activity (Meyer, Weber et al. 1981). The M-site is located within the replicase gene, from nucleotides 2550-2870, and unlike the S-site, the M-site is required for gRNA replication, with its removal resulting in a drastic loss of template activity (Schuppli, Miranda et al. 1998). While the M-site is located >1,000 nt from the 3' end of the genome the end of the genome is brought into proximity by two long-range interactions and a short-range interaction (Klovins, Berzins et al. 1998, Klovins and van Duin 1999).

The Qβ replicase core was visualized using negative-stain electron microscopy in 1988 which confirmed the core complex has four subunits (Berestowskaya, Vasiliev et al. 1988). Initial crystal structures of the β-subunit with EF-Tu and EF-Ts, with or without an illegitimate RNA template revealed that while the core of the RdRp has significant similarity with other viral RdRps, the ssRNA phage RdRp has evolved features to take advantage of the bacterial EF-Tu to split the template from product strands of RNA (Kidmose, Vasiliev et al. 2010, Takeshita and Tomita 2010, Takeshita, Yamashita et al. 2012). These structures were solved without S1, a crucial component in recognizing the positive-sense gRNA (Gytz, Mohr et al. 2015). The hijacked ribosomal

protein S1 is not required for negative-strand replication but is for positive-strand replication (Kutlubaeva, Chetverina et al. 2017). S1 has six oligonucleotide binding domains, the three N-terminal domains were found to be necessary for RNA replication within the Qβ replicase holoenzyme (Vasilyev, Kutlubaeva et al. 2013). Later, it was found that the two N-terminal oligonucleotide binding domains of S1 anchored it to the replicase, while the third (of six) OB-fold protrudes from the side of the replicase and interacts with the RNA templates (Takeshita, Yamashita et al. 2014). One the ribosome, S1 functions in translation initiation by binding to mRNAs and possibly unwinding the secondary structure (Duval, Korepanov et al. 2013).

CHAPTER II

STRUCTURE OF Qβ[1]

2.1. Prior Knowledge of Qβ Structures

The structures of Qβ and MS2 have both been studied extensively. As with all ssRNA phages identified to date, the capsid of Qβ is composed of ~180 copies of the coat protein which form an icosahedral capsid with *T*=3 quasi-symmetry (Takamatsu and Iso 1982). MS2 was among the first viruses which had a structure reconstructed from electron microscopy images (Crowther, Amos et al. 1975). The structure of MS2 was found to have icosahedral quasi-symmetry as established by Caspar and Klug (Caspar and Klug 1962). Icosahedral shells have a combination of 5-fold, 3-fold, and 2-fold symmetry axes which result in 60-fold redundancy (Figure II-1) (Parent, Schrad et al. 2018). Viruses with capsids that have more than 60 coat protein monomers require those coat proteins to have conformational plasticity, wherein they have quasi-equivalent properties allowing each monomer to occupy somewhat different geometric spaces. These proteins were grouped based on the similarity of their contacts, with six subunits together forming a hexameric ring, while five subunits together form a pentameric ring (Rossmann 2013). The pentameric rings form 12 icosahedral vertices along with the 20 vertices from hexameric rings to form the viral capsid. For MS2 and Qβ, as well as all

[1] The basis for this chapter comes from 'Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qβ reveals a single maturation protein and the genomic ssRNA in situ' by Karl V. Gorzelnik, Zhicheng Cui, Catrina A. Reed, Joanita Jakana, Ry Young, and Junjie Zhang *Proceedings of the National Academy of Sciences of the United States of America* 2016, 113:41 11519-11524. Figures are republished with permission.

other viruses with similar symmetries, the 3- and 2-fold axes are within the hexameric ring. Caspar and Klug defined the *T* number (triangulation number) as the number of similarly structured subunits per icosahedral asymmetric unit, or the different conformations the same subunit could occupy (Rossmann 2013). The *T* number is mathematically defined as $T = h^2 + hk + k^2$, wherein *h* and *k* are the number of hexameric rings between pentameric rings along the *h* and *k* axes of a hexagonal array. As *h* and *k* are numbers of rings along an axis, they have to be integers, for that reason when using the equation listed in the previous sentence, the *T* number can never be equal to 2, rather *T* = 1, 3, 4, 7, 9, 13, 16, 21, etc.. The majority of viruses on earth form icosahedral shells with 5:3:2 symmetry (Parent, Schrad et al. 2018). For Qβ and MS2, and all known ssRNA phage, *T*=3.

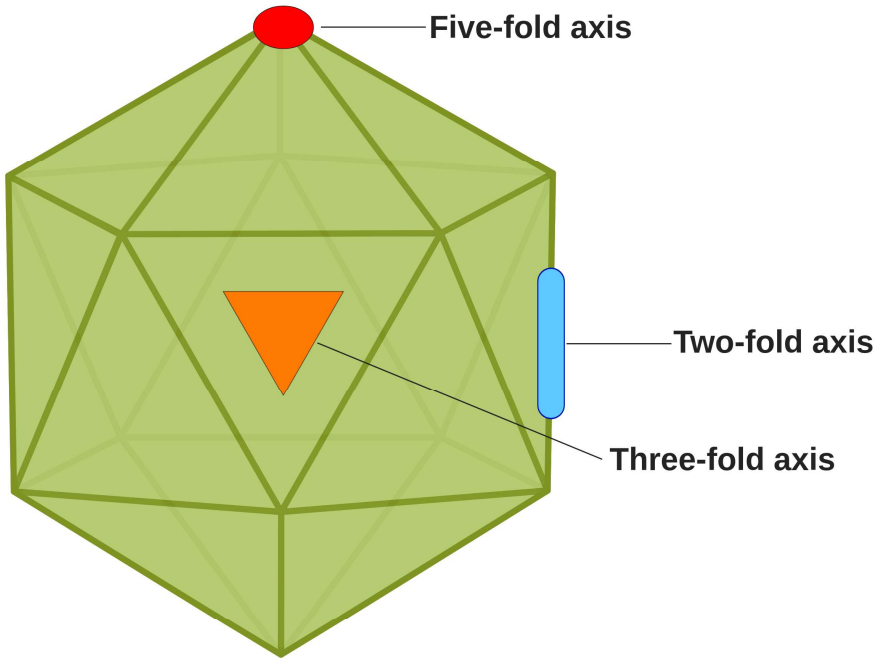

Figure II-1. How the axes are defined in a *T*=3 icosahedral virus
The *T* number represents the 'triangulation' number representing the different conformations the same protein can adopt. There are twenty faces for a *T*=3 icosahedral virus. The virus has two-fold, three-fold, and five-fold axes.

The flexibility of protein subunits allows for proteins to be in different conformations within the viral capsid, enabling the 'quasi-symmetry' for identical backbones to occupy divergent geometric spaces. Different geometries of the identical subunit suggest that there is a 'conformational switching' event during the assembly of these viruses (Kellenberger, 1976). This event is when two proteins come to bind each other, they induce a conformational change in one or both, which prevents a third chemically identical molecule to bind to either of the first two in the same way, as that conformation no longer exists (Rossmann 2013). Successive assembly of the capsid

must, therefore, be controlled by distinct events, a theory which has been extensively studied by mathematicians, particularly using MS2 as a model virus (Dykeman, Grayson et al. 2011, Dykeman, Stockley et al. 2013, Twarock, Bingham et al. 2018).

The symmetry of the *T*=3 capsids is disrupted in Qβ and other group B ssRNA by the readthrough coat protein they encode (Hofstetter, Monstein et al. 1974). Most of the coat proteins in Qβ have 132 residues, but up to thirteen subunits have a C-terminal extension of an additional 196 amino acids, caused by a readthrough of the leaky stop codon (Weiner and Weber 1971, Takamatsu and Iso 1982). The coat proteins in this form are referred to as A_1, as this protein is present in a higher abundance than A_2, the maturation protein for Qβ, and was mistakenly thought to be the maturation protein (the A protein in MS2, hence the nomenclature for Qβ) (Steitz 1968, Steitz 1968, Garwes, Sillero et al. 1969, Strauss and Kaesberg 1970, Weber and Konigsberg 1975). As the N-terminal portion of A_1 is the coat protein, A_1 can theoretically be anywhere within the capsid. There have been no studies as to the location of A_1 within the virion, and the protein:protein contacts that facilitate capsid formation are located exclusively within the coat protein domain (Golmohammadi, Fridborg et al. 1996). A_1 is required for infection, particles reconstituted with the coat protein alone or with A_2 are noninfectious while the reconstitution of particles with the coat protein, A_1, and A_2 are able to infect *E. coli* (Hung and Overby 1969, Hofstetter, Monstein et al. 1974). The required number of A_1 copies in a capsid and their location within a capsid for a particle to be infectious is

unknown. Even the length of A_1 required for successful infection is unknown (Rumnieks and Tars 2011).

The coat protein, the read-through coat protein A_1, and the maturation protein A_2 are not actually required for infection (Zinder 1965), they merely protect the gRNA when the virus is outside of the host and in the case of A_1 and A_2, facilitate infection (Staples, Hindley et al. 1971, Hofstetter, Monstein et al. 1974). The gRNA is infectious on its own, once it gets inside the cell the host ribosomes will treat it as any other mRNA and start translating (Weissmann 1974). Once replicase levels are high enough, the replicase will bind the S-site, preventing further translation, and the replicase will subsequently be able to initiate translation from the 3'-5' direction. The maturation protein subsequently also serves as the lysis protein, preventing cell wall biosynthesis by inhibiting MurA (Bernhardt, Wang et al. 2001). This is in contrast to group A phages, which have their own dedicated lysis protein that does not have a structural role in the mature virion (Valegard, Liljas et al. 1990).

The coat protein adopts three conformations to facilitate the *T*=3 quasi-symmetry (Golmohammadi, Fridborg et al. 1996). These three conformations are similar, with the differences being seen by packing interactions with other coat proteins once in the capsid (Golmohammadi, Fridborg et al. 1996). The three conformations, termed A, B, and C, form dimers of A/B or C/C (Golmohammadi, Fridborg et al. 1996). The Stockley group, which has worked on MS2 for decades, has found that the MS2 coat protein undergoes

conformational changes before capsid assembly, with C/C dimers changing conformation to A/B once binding RNA stem-loops (Stockley, Rolfsson et al. 2007, Basnak, Morton et al. 2010, Morton, Dykeman et al. 2010, Rolfsson, Toropova et al. 2010, Borodavka, Tuma et al. 2012, Dykeman, Stockley et al. 2013).

The Qβ coat protein has two antiparallel β-strands which fold over the five-stranded antiparallel β-sheet followed by two α-helices (Golmohammadi, Fridborg et al. 1996). These β-strands are named A-G, based on their location within the gene, the two solitary strands are A and B, while the five strands which make up the β-sheet are C-G. It was suggested that this fold is not particularly stable as a monomer, but once two subunits come together the five stranded β-sheet becomes a ten stranded antiparallel β-sheet (Golmohammadi, Fridborg et al. 1996). When comparing A/B dimers with C/C dimers, the conformations are very similar to each other, with the C_α backbones being less than 0.7 Å root mean squared deviation difference between the two conformers (Golmohammadi, Fridborg et al. 1996). In both conformers, the loop between the A and B β-strands interacts with the first α-helix of the other subunit of the dimer pair (Golmohammadi, Fridborg et al. 1996).

There are two cysteines within the coat protein at amino acids 74 and 80, which are thought to form disulfide bonds between dimers at the five-fold and three-fold axes (Golmohammadi, Fridborg et al. 1996). I am not sure how this would work, but it is mentioned repeatedly for capsids of ssRNA phages (Golmohammadi, Fridborg et al.

1996, Cielens, Ose et al. 2000, Ashcroft, Lago et al. 2005, Caldeira and Peabody 2007, Bundy and Swartz 2011, Fiedler, Higginson et al. 2012). The reason why this should not be the case is that the cytoplasm is a reducing environment (Hatahet, Boyd et al. 2014). The only disulfide bonds present are in proteins in the cytoplasm that gain them during the course of their enzymatic catalysis or during signal transduction, such as ribonucleotide reductase or the transcription factor OxyR (Stewart, Aslund et al. 1998). In general, disulfide bond formation in *E. coli* is only found in the periplasmic space (Rietsch and Beckwith 1998). That being said, several groups which reported disulfide bonds in Qβ or VLPs composed of Qβ coat proteins show fairly conclusive evidence that disulfide bonds are being formed, through disrupting the capsids with reducing agents such as DTT, only heating the proteins a little to leave disulfide bonds intact when running protein gels, or measuring thermal stability with or without the reducing agents (Cielens, Ose et al. 2000, Ashcroft, Lago et al. 2005, Caldeira and Peabody 2007, Bundy and Swartz 2011, Fiedler, Higginson et al. 2012).

Wild-type MS2 two cysteines in its coat protein, but they are sequestered in the interior, so they cannot form disulfide bonds (Fiers, Contreras et al. 1976, Valegard, Murray et al. 1994, Peabody 2003). Introduction of cysteines into the surface of MS2 increased the stability of the mutant capsid, versus wild-type, but did not uniformly result in disulfide bond formation, with only ~3% of surface-exposed cysteines forming disulfide bonds (Peabody 2003). This perhaps is more likely for Qβ, with only some coat protein monomers forming disulfide bonds. At the 3.5 Å resolution for the capsid

of Qβ (Golmohammadi, Fridborg et al. 1996) the disulfide bonds cannot accurately be seen, but are predicted. The cytoplasm of *E. coli* is reducing due to the millimolar levels of glutathione (Smirnova, Muzyka et al. 2012). However, it is possible that there are local pockets of varying concentrations, especially given the crystalline packing of ssRNA phage during infection (Figure II-2) which could prevent the even distribution of the enzymes responsible for glutathione production (Schwartz and Zinder 1963).

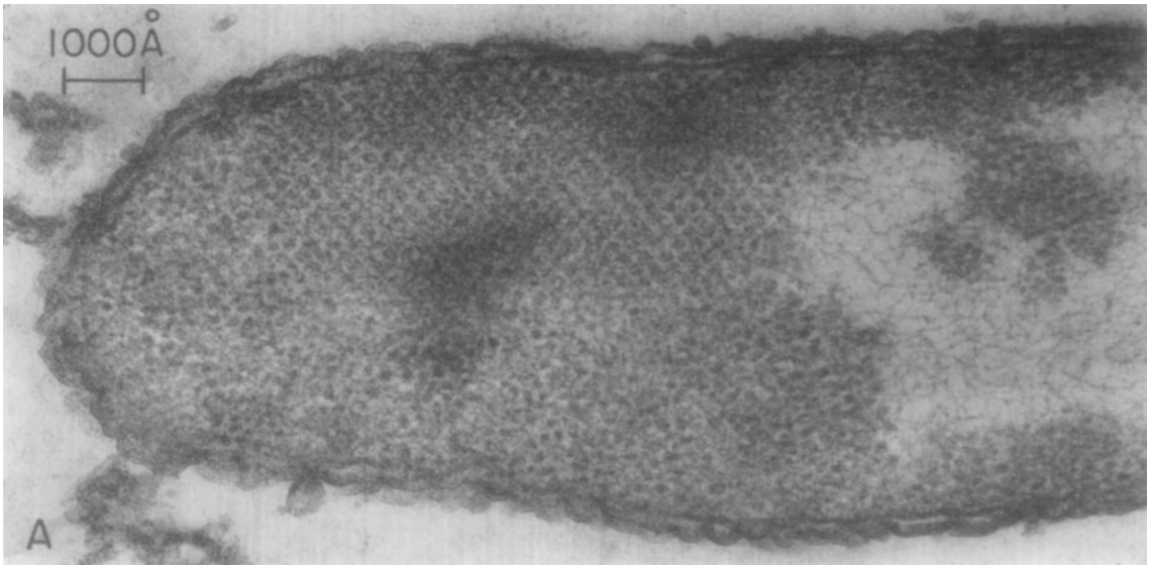

Figure II-2. Crystalline packing of f2 (MS2) within *E. coli*
During infection, of *E. coli,* f2 (MS2) produces ~20,000 particles, which can be visualized by negative staining. These particles occupy ~3.5% of the host cell volume (although the authors who derived this number used too small of a diameter for the phage particle) and can be seen to form a crystalline-like matrix of particles, wherein the local conditions could differ from that of the rest of the cytoplasm. Image used with permission from *Virology* (Schwartz and Zinder 1963).

Given that local concentrations of reducing agents within the cytoplasm could vary, and that disulfide bonds could form after release of the phage particles from the

lysed cell, there is the possibility that some coat protein monomers are linked by disulfide bonds in Qβ, but it should not be assumed that all coat protein monomers have two disulfide bonds. While not all coat protein monomers may be disulfide bond linked in the capsid, the viral capsid is very stable. The model proposed by Golmohammadi and colleagues shows two unique interdimer salt bridges and thirteen unique hydrogen bonds (Golmohammadi, Fridborg et al. 1996).

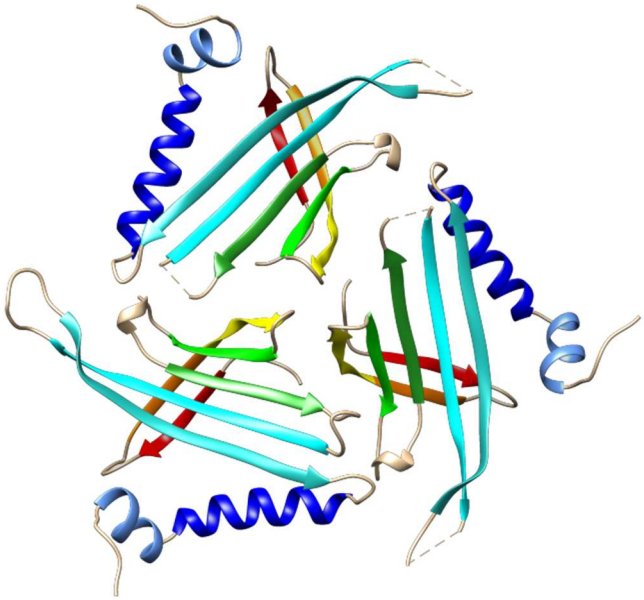

Figure II-3. Crystal structure of the three conformations of the Qβ coat protein monomer
From the crystal of the capsid (PDB id: 1QBE) by Golmohammadi and colleagues (Golmohammadi, Fridborg et al. 1996). The view seen is from the inside of the capsid with the exterior away from the page. Conformers A, B, and C start at the top and go counterclockwise. A and B form A/B dimers, while C forms C/C dimers. The seven β-strands, referred to as a-g start, from the N-terminus and are colored orange-red (a), orange (b), yellow (c), green (d), forest green (e), cyan (f), and light sea green (g), while the two α-helices are blue (A) and cornflower blue (B). The structure is missing the EF and FG loops (named based on the β-strands adjoining the loop) for conformers A and B due to flexibility in the electron density.

The cysteines potentially involved in disulfide bonds are at residues 74 and 80, within the FG loop that is missing from the electron density in conformations A and B. Cysteines in these positions are present in all sequences of ssRNA phages within *Allolevivirus* but are absent in *Levivirus* (Golmohammadi, Fridborg et al. 1996). These cysteines are also present in the Pseudomonas ssRNA phage PP7 (Tars, Fridborg et al. 2000). PP7 virus-like particles, formed in vitro, show disulfide-bonds that protect against thermal denaturation (Caldeira and Peabody 2007).

2.2. Materials and Methods

2.2.1. Bacterial Strains and Growth Conditions

Escherichia coli was used to propagate Qβ. Qβ and its hosts were a gift of Karthik Chamakura in the Young lab. Qβ was initially produced off of a cDNA plasmid, pQBm100 (Reed, Langlais et al. 2012) in ER2738. While ER2738 is a male (F^+) strain, with pQBm100 this ER2738 was effectively female *E. coli*, as it was resistant to Qβ and MS2 (not shown). While XL1-Blue and HfrH were also tried, ER2738, without the Qβ cDNA plasmid, was the most effective for infections. *E. coli* was grown at an air to liquid ratio of 1:4, i.e. 500ml culture for a 2L flask. Cultures were grown with regular LB at 37°C, shaking at 200 r.p.m. LB was supplemented with 100 µg/ml ampicillin, 50 µg/ml kanamycin, or 5 µg/ml tetracycline. Tetracycline was tetracycline hydrate, as opposed to tetracycline chloride, was used for ER2738, as tetracycline chloride delivered

two orders of magnitude lower titers (Figure II-4). This could be for a variety of reasons, but was not investigated further.

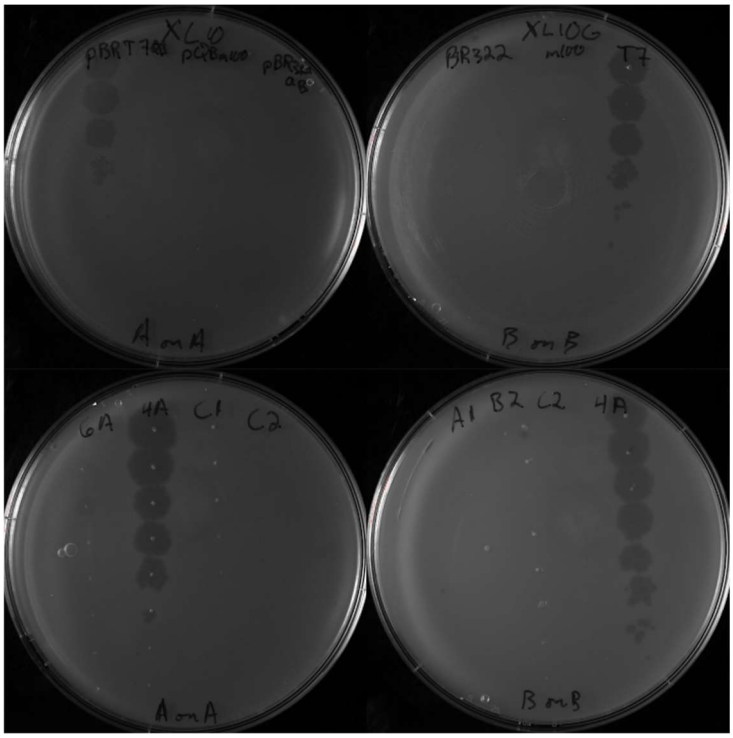

Figure II-4. Different tetracyclines affect the titer of Qβ
The titers were done on the same day, with the same serial dilutions of Qβ, but different tetracyclines, either tetracycline HCl (designated as A at the bottom of the plates) or tetracycline hydrate (designated as B at the bottom of the plates). Notice the difference for pBRT7QB (top left) versus T7 (top right; the same mutant, just written differently on the plate). The same holds true for 4A (bottom panels), a mutant generated, which plaqued to the 10^{-6} dilution on A (tetracycline HCl) but 10^{-8} on B (tetracycline hydrate).

2.2.2. Purification of Qβ

An overnight culture of this strain produced a titer of ~10^{11} PFU/ml (plaque forming units). While ER2738 is a male (F^+) strain, with pQBm100 this ER2738 was effectively female *E. coli*, as it was resistant to Qβ, MS2, and M13 (an ssDNA phage

which binds to the tip of the F-pilus). Qβ produced from overnights of this bacteria were used to infect ER2738 without the plasmid (male *E. coli*). ER2738 without pQBm100 was grown overnight in standard LB broth, with 5 µg/ml tetracycline, at 37°C, then inoculated 1:1,000 into fresh media. Once the cultures were at an OD_{600} = ~0.3 – 0.5, the cells were infected at a MOI (multiplicity of infection) = ~1-2. Growth curves were taken compared to uninfected samples. The infection was allowed to proceed for ~5 hours. The cells were spun down at 8,000 x *g* for 30min. The supernatant was filtered through a bottle-top 0.22µm filter then 0.5M EDTA pH 8.0 was added to 5mM and the sample was stored at 4°C for at least 4 hours or overnight.

The cell pellet was saved and either resuspended in Qβ buffer (150mM NaCl, 5mM EDTA, 50mM TrisHCl pH 8.0) and lysed in a French Press right away (p.s.i. = 25,000) or frozen at -20°C until being resuspended later and lysed. Once cells were lysed, the cell debris was cleared by spinning at 20,000 x *g* for 30 minutes. The supernatant was added to the supernatant of the cultures, while the pellet was discarded.

The phages in the supernatant of the cultures and the lysed cells were precipitated with 280g of ammonium sulfate per liter of sample, after the samples were at 4°C. Once the ammonium sulfate was dissolved, the samples were spun down at 8,000 x *g* for 30 minutes. The supernatant of this was discarded and the ammonium sulfate precipitation was resuspended in Qβ buffer (~30ml / L of initial culture volume). The remaining ammonium sulfate was dialyzed away through 20kDa cutoff dialysis tubing. The higher

the cutoff, the better the results, 3kDa tubing was not as effective at removing the ammonium sulfate. At least three rounds of dialysis were used for at least two hours each time. DNase and RNase were added to the dialyzed sample, at 10 U/ml. After this the samples had CsCl added to 0.52g/ml, the samples were put in Quick Seal tubes and loaded into a Ti 70.1 rotor, then spun down for at least 24hrs at 45,000 r.p.m. The longer the spin, the closer to equilibrium the samples would be, and thus the further apart the bands and the tighter the bands. For Qβ, but not MS2, there were multiple bands, presumably due to differing amounts of A_1 in the capsid. The lower band had the highest titer, and was thus used. Both bands were subjected to mass spectrometry of the protein components (Protein Chemistry Lab, Texas A&M University). There must have been some A_1 in the higher band, as it was infectious, but not enough to be detectable via mass spectrometry. The picked bands were dialyzed first against 1M NaCl, 50mM TrisHCl pH 8.0 overnight at 4°C. They were then dialyzed against Qβ buffer three times for at least four hours each time at 4°C, serial diluted in Qβ buffer and titered against ER2738.

2.2.3. CryoEM Data Collection

Qβ was diluted to ~5mg/ml protein, as measured by nanodrop at A_{280}, which corresponded to ~10^{12} PFU/ml. 3μL was applied to a 1.2/1.3 C-Flat grid at 20°C, with 100% humidity in a Vitrobot Mark III. Data was collected at Baylor College of Medicine using their JEM32000FSC cryo-electron microscope with a 300kV JEOL field emission gun. Each micrograph was collected manually using SerialEM and recorded

with a Gatan K2 Summit direct detection camera, in super-resolution electron counting mode. A nominal magnification of 30,000X was used, which gave a 0.6 Å subpixel size. The beam intensity was adjusted to a dose rate of ~7 e^-/pixel/sec. 50 frame movie stacks were taken, with 0.2sec / frame, for a total exposure time of 10sec per image. An in-column energy filter with a slit width of 29eV was used.

2.2.4. Data Processing

The initial Qβ structure (the data used for this chapter) was solved using 712 movie stacks which were initially binned by 2, then aligned, and filtered using Unblur (Grant and Grigorieff 2015). The defocus values for each summed micrograph were determined using CTFFIND4 (Rohou and Grigorieff 2015). EMAN2 was used to semi-automatically pick particles, with a box size of 320 x 320 pixels, yielding 111,507 particles. The particles were then shrunk 8 times to a pixel size of 9.6 Å and screened for high-contrast particles. The 51,815 clean particles were then refined in Relion. The starting map was a 60 Å low-pass filtered version of the crystal structure of the Qβ capsid. The icosahedral reconstruction yielded a map of 3.7 Å resolution. Of the 51,815 particles, only 12,975 were able to be used for the asymmetric reconstruction, when using an unsupervised classification in Relion. Upon reviewing the particles after the 3D classification it appears that the difference between the particles used in the final asymmetric reconstruction and those which were in other classes was that the gRNA was more stable or homogenous in the dominant class.

2.3. Results

The structure for Qβ was initially reconstructed from 712 micrographs collected on a direct detection camera in super-resolution mode, yielding 51,815 particles used for the icosahedral reconstruction Figure II-5. We subsequently collected more data on Qβ, which will be discussed later. The next section is based on data from the initial structure. At 30,000X magnification on a JEOL3200FSC, between 0-300 usable particles were able to be visualized per image, with some holes having crystalline-like packing of phage while some others might only have a handful of virions, even within the same grid square. Overlapping viruses were able to be used for the icosahedral reconstruction, but not the asymmetric reconstruction.

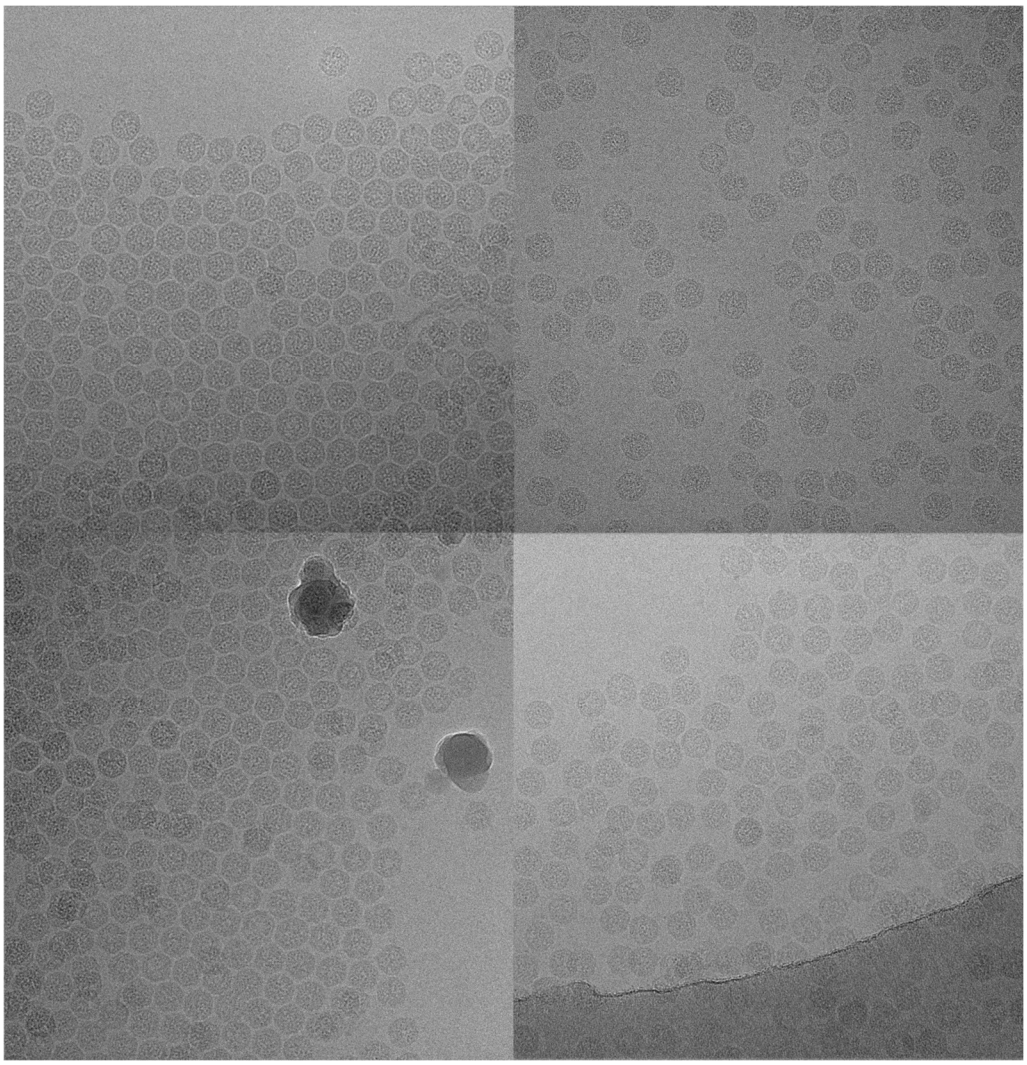

Figure II-5. Representative micrographs of Qβ
While all four images were taken from the same grid, concentrations per hole differed. Overlapping particles were not used for the asymmetric reconstructions, nor were those that were on the carbon or overlapping with the edge/ice contamination.

With symmetry applied to the coat protein shell, the resolution reached 3.7 Å, as seen in Figure II-6. The three conformers of the coat protein are shown in red, green, and blue for A, B, and C, respectively. The α-helices and β-strands are clearly resolved from the electron density, with most of the bulky side-chains visible and within the

electron density (Figure II-7). The cryo-EM structure is consistent with the previous crystal structure (PDB: 1QBE) (Golmohammadi et al., 1996). The cryoEM electron density fits the model of the crystal structure well, without the need for readjustments (Figure II-7).

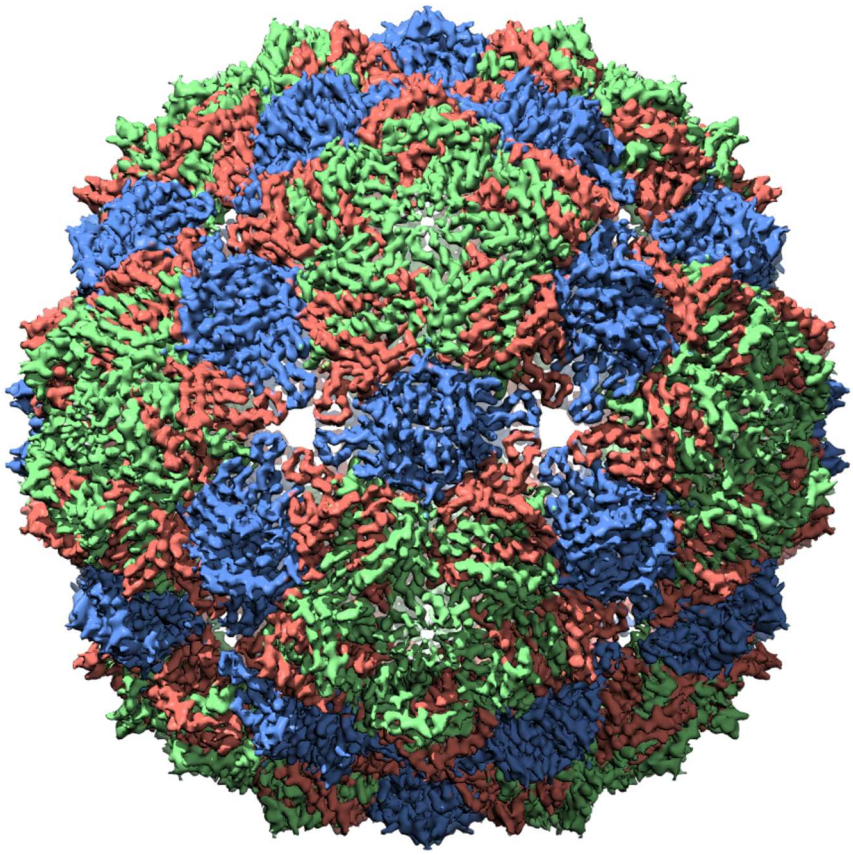

Figure II-6. Qβ capsid with symmetry applied
The three conformations of the coat protein (A, B, and C) are colored red, green, and blue, respectively. Coat protein dimers are composed of A/B conformers or C/C conformers (red/green or blue/blue in the figure).

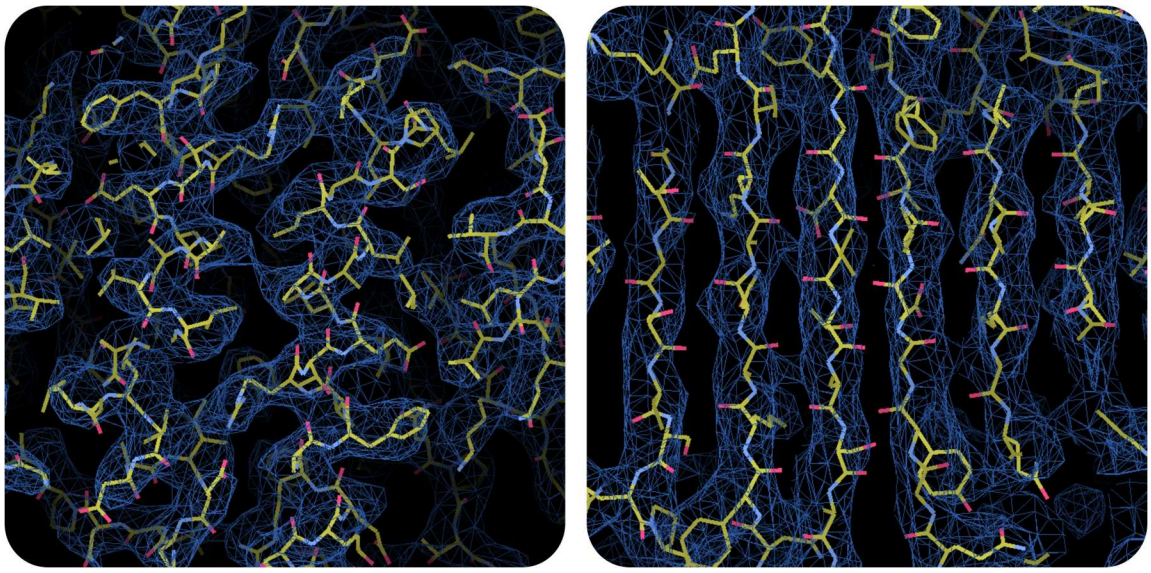

Figure II-7. Fitting of the crystal structure of the capsid for Qβ to the cryoEM electron density map for the symmetric reconstruction
The α-helices (left) and β-sheets (right) are clearly resolved within the 3.7 Å symmetric reconstruction of Qβ.

The model for the coat protein from the cryoEM is overlaid onto that from the crystal structure, with the cryo structure in blue and the crystal in tan (Figure II-8). The major difference between the cryoEM symmetric structure and the crystal structure is the FG loop, which was not able to be resolved in the crystal structure, but is able to be seen in the cryoEM density. The FG loop is between the F and G β-strands, these amino acids, residues 76-79 (ANGS), are in between two cysteines (residues 74 and 80, CTANGSC). These two cysteines were proposed to form disulfide bonds, but the electron density was smeared in the crystal structure, so the authors were not able to identify the positions of the disulfide sulfur atoms (Golmohammadi, Fridborg et al. 1996).

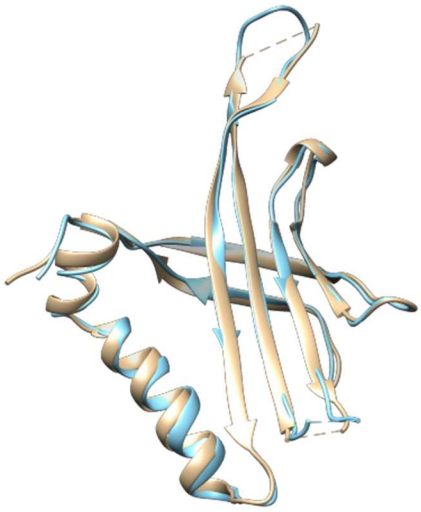

Figure II-8. Fitting of the ribbon models from crystal structure and the cryoEM structure of the coat protein from Qβ
The crystal structure, colored in beige, is overlaid onto the ribbon model from the cryoEM symmetric reconstruction of the Qβ capsid (light blue). The dashed lines in the crystal structure reflect areas where the protein was flexible and thus not able to be built in, whereas these regions are able to be seen in the cryoEM electron density.

The electron density present enabled the protein backbone to be modeled in, connecting the F and G loops, which were unable to be resolved in the crystal structure, due to weak electron density, but are seen in the cryoEM structure (Figure II-9). The FG-loop forms pores at the fivefold and threefold vertices within the capsid. The FG-loop was able to be resolved within C/C dimers in the crystal structure but now is seen for all interactions. These residues form an ~15-Å pore at the threefold vertex, in agreement with its predicted size (Golmohammadi, Fridborg et al. 1996). At the fivefold vertex, between C/C dimers, there is an ~8-Å pore, which restricted the movement of the FG-loop, allowing it to be resolved in the crystal structure. Also, of interest are the EF-

loops, which were able to be resolved in the crystal structure, but with weaker density than the rest of the backbone, and play a prominent role in the interpretation of the asymmetric reconstruction.

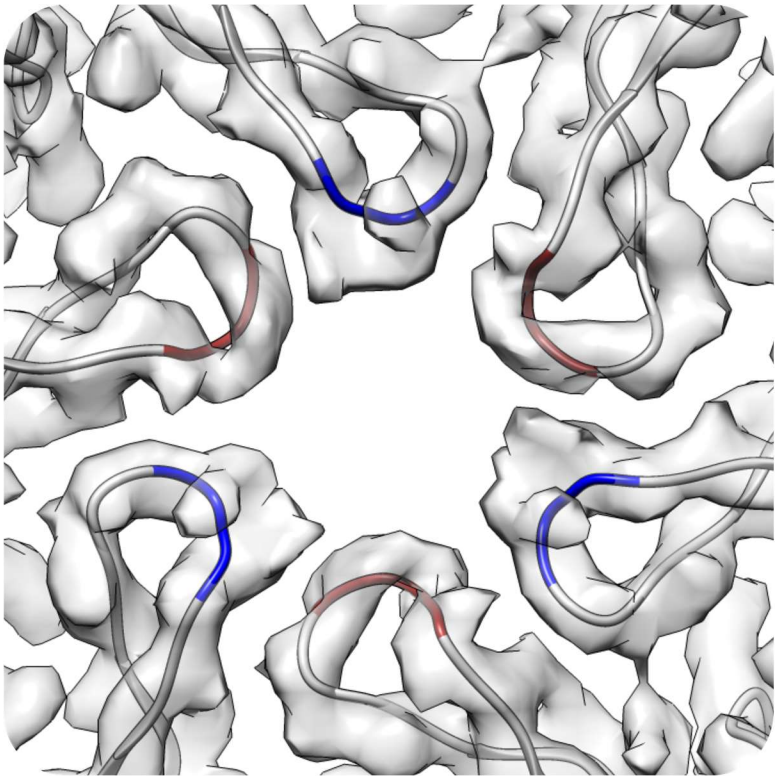

Figure II-9. FG-loops at a three-fold axis in the symmetric reconstruction of Qβ
The electron density (gray) clearly shows the FG-loops which were unable to be visualized in the crystal structure. The amino acids which were missing from the crystal structure are colored according to the coat protein conformer (red for A and blue for C conformers).

To determine the structures of the maturation protein and the gRNA within the Qβ virion symmetry had to be released. After 3D classification of the particles to select for those with the maturation protein and the gRNA in a defined conformation, 12,975

virions were selected for the asymmetric reconstruction (Figure II-10). This lowered the resolution of the virus to 7-Å but revealed the genomic RNA and the maturation protein (Figure II-10). Masking out the gRNA enabled the resolution of the coat proteins and maturation protein to reach ~6.5-Å. Overall, the gRNA near the capsid has a much higher resolution than that in the interior of the virus. This is presumably because the capsid stabilizes the gRNA, or limits its mobility. The gRNA is able to be resolved to the point where grooving can be visualized (Figure II-10, right). The capsid stabilizes the gRNA by the coat proteins interacting with the positively-charged internal EF-loop. Thus, the gRNA proximal to the capsid is at a higher resolution than the gRNA further from the capsid (Figure II-10).

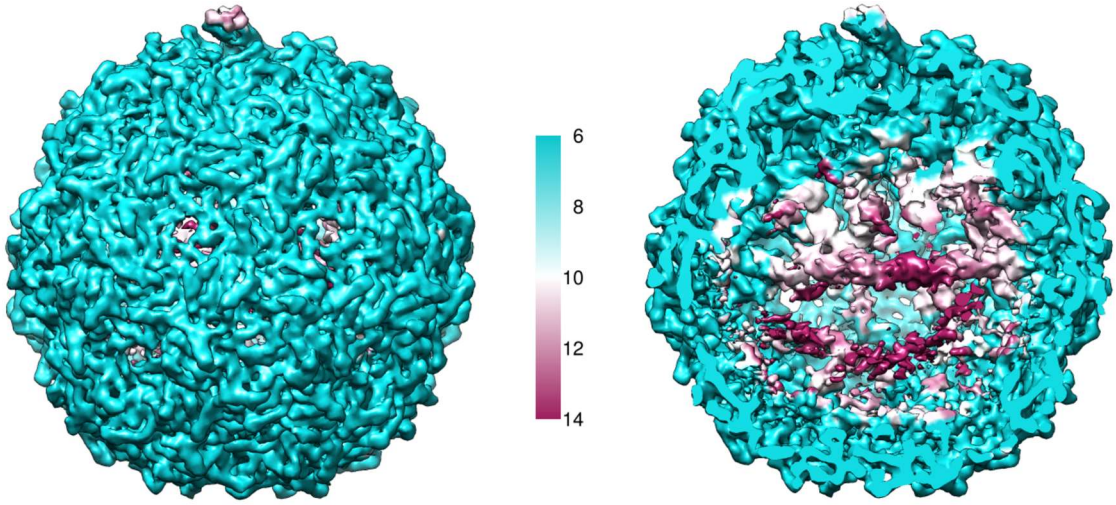

Figure II-10. Local resolution for the asymmetric reconstruction of Qβ
The local resolution is color-coded, with dark blue being the best resolution (lowest, or 6 Å) with the poorer resolution being colored in increasingly darker red.

The gRNA in the northern hemisphere of the asymmetric particles is also stabilized and thus able to be resolved to a higher resolution (Figure II-10). Within the asymmetric structure there is protein density in this region. The capsids of ssRNA phage are thought to contain ~180 copies of the coat protein, with Qβ and other *Allolevivirus* having a small amount of readthrough coat proteins (A_1, ~5-15 copies per virion), and a single copy of the maturation protein. The electron density shows that the coat proteins which make up the capsid are displaced at a two-fold axis. As there is only one disruption within the capsid and only one maturation protein within the virion, it would be logical that this position is occupied by the maturation protein. Furthermore, the N-terminal domain of A_1 is the coat protein, so this protein density cannot be A_1. If A_1 associated with other coat proteins based on the same contacts that the coat proteins use to dimerize then A_1 could be anywhere within the capsid. After further examining the electron density, by docking in a crystal structure of the readthrough domain of A_1, the protein displacing a pair of coat protein dimers at a two-fold axis is not A_1 and therefore it stands to reason that the protein density is that of A_2.

The electron density within this position interacts with the gRNA in a manner unlike the known interactions of the coat protein / operator complex. As it is known that there is one copy of the maturation protein (A_2 in the case of Qβ, A in the case of MS2), we worked on the assumption that this protein was the maturation protein. By fitting the crystal structure of the capsid into the electron density of the asymmetric structure it is easy to tell the gRNA from the coat protein components of the capsid (Figure II-11).

The electron density belonging to the capsid is colored as before in Figure II-6, with the different conformers in their respective colors. The electron density on the inside of the capsid, to which a trained eye (Dr. Junjie Zhang) can tell is not protein density, is colored in yellow. The protein density which displaces a pair of coat protein dimers is colored in pink. In subsequent work, it was determined that the whole of the protein density in this location is not the maturation protein, but this will be discussed later.

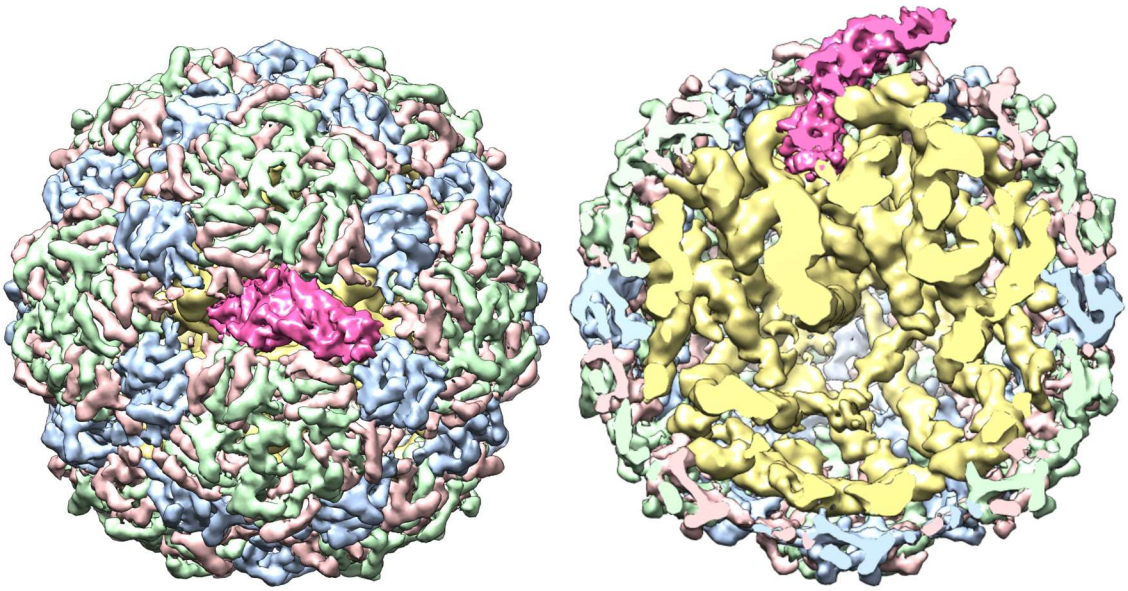

Figure II-11. Asymmetric reconstruction of Qβ with the protein components colored to better visualized the gRNA
The coat proteins are colored according to their conformer: A, B, and C, are colored in light red, light green, and light blue. The protein component which did not belong to the coat proteins, and did not fit the crystal structure for the readthrough coat protein A_1, is colored in hot pink. The presumed gRNA is colored in yellow. The left image is a top-down view from the sole break in capsid symmetry, while the right image is the same reconstruction, rotated 90 degrees and sliced open to better display the gRNA.

What is readily seen in the asymmetric structure is that the gRNA adopts a dominant conformation. If the gRNA adopted many different conformations the electron density would be blurred out, as is the case in the 'Southern Hemisphere' of the virus, the area further away from the putative maturation protein. This is presumably because the maturation protein stabilizes that gRNA proximal to it, or constricts its movement. The paradigm for packaging was that the maturation protein binds to the 5' and 3' end of the gRNA, constricting the RNA within a space such that the coat proteins are able to bind to and collapse the gRNA into a volume that would fit inside of the capsid. The resolution from this first structure of Qβ is not high enough to be able to identify the RNA sequence proximal to the maturation protein. However, the resolution was high enough to identify secondary structure elements within the putative maturation protein (Figure II-12). There are five segments that are clear α-helices, closely matching the secondary structure prediction for A_2 (secondary structure prediction done using I-TASSER). The secondary structure prediction of the maturation protein shows many α-helices, the longest of which is about 60 amino acids, this is the length of the longest putative α-helix in the electron density.

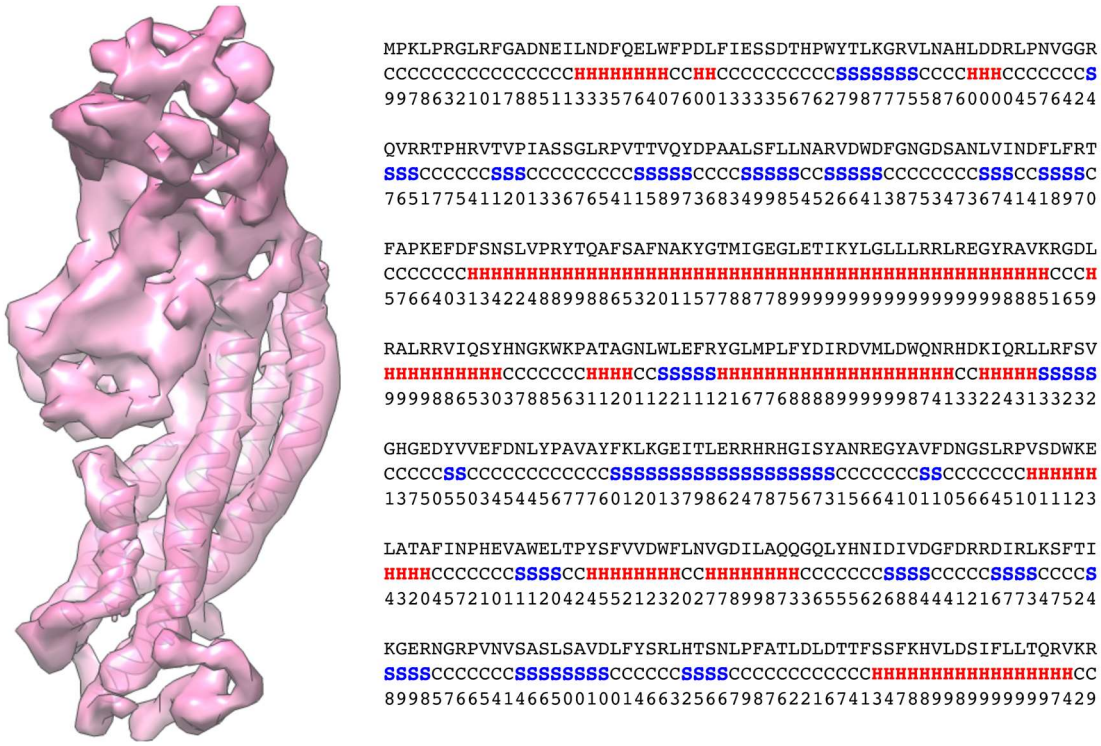

Figure II-12. Electron density of the putative maturation protein A_2, from the asymmetric structure of Qβ
The putative maturation protein has two regions, the α- and β- regions on the top and bottom. The α-region is proximal to the gRNA in the structure, while the β-region is exposed outside of the viral capsid. α-helices are able to be modeled into the α-region, with lengths similar to that of the I-TASSER predicted secondary structure. The resolution of the electron density is not high enough to model in the side-chains of particular amino acids, so the protein backbone is not able to be accurately modeled in.

The maturation protein protrudes from the capsid at an ~120° angle from the capsid. This angle is to presumably increase the surface area of the maturation protein that is available to bind to the pilus. The maturation protein not only replaces a coat protein dimer at a 2-fold axis but also displaces the coat proteins proximal to the maturation (Figure II-13). The displacement of coat protein dimers proximal to the maturation protein is best seen by looking at the ribbon models of the proteins. By

comparing where the center of a coat protein dimer is in the icosahedral (symmetric) structure with that from the same dimer in the asymmetric structure, the coat proteins proximal to the maturation protein are displaced more than those further away (Figure II-13). Whether this displacement is due to the maturation protein or the genomic RNA, is unknown. More likely, it is a combination of both factors. The displacement of coat proteins amounts to a small crack in the capsid, with the displacement more on one side of the maturation protein than on the other. This could allow the genome to be better pulled out of the capsid. The area of the slot created by a displaced coat protein dimer is ~60 x 40 Å. While the opening is certainly large, it might not be big enough to allow folded gRNA to leave the capsid rapidly. The crack in the capsid might prevent disulfide bonds from forming between dimers, allowing for 'breathing' of the capsid which could expand slightly when regions of the gRNA with high secondary structure or 'knots' are pulled out of the capsid.

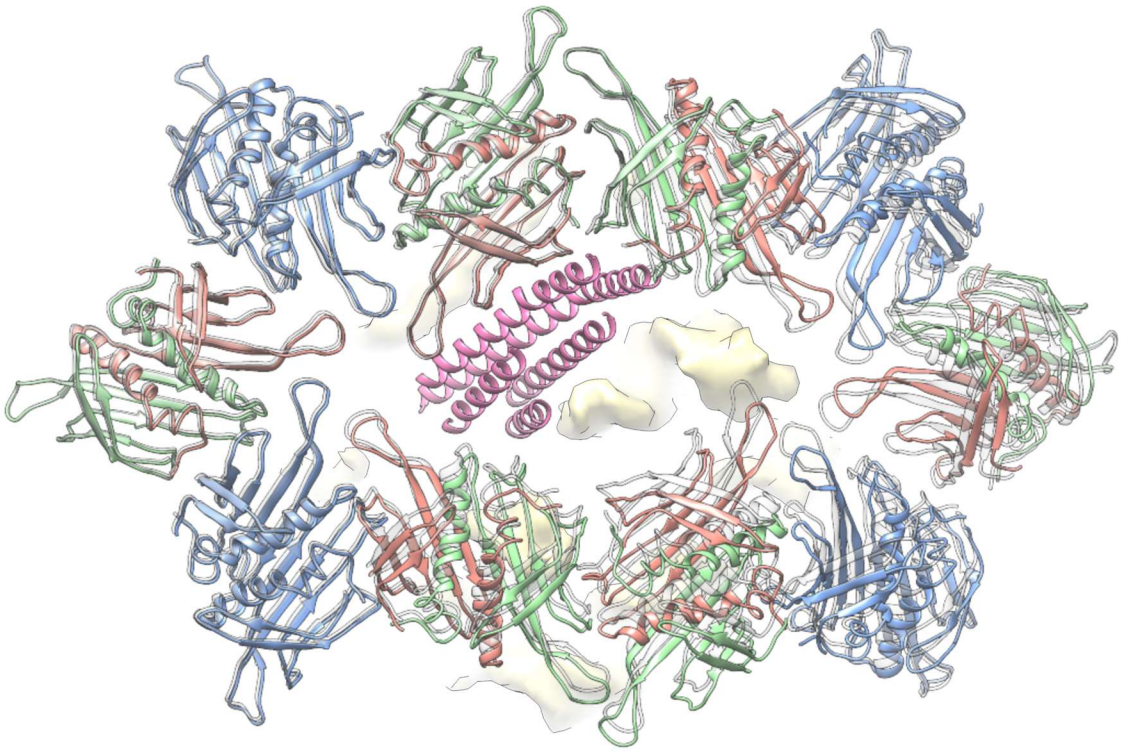

Figure II-13. Displacement of the coat proteins by the maturation protein around the opening of the virus
The coat protein monomers from the asymmetric structure are colored and docked into the 2-fold axis of the symmetric structure (where the maturation protein A2 displaces a coat dimer). The coat protein dimers from the symmetric structure are colored in grey to show the displacement of coat protein dimers by the maturation protein. The coat protein monomers from the asymmetric structure are colored as per their conformer in Figure 2pX2, with A, B, and C conformers colored as red, green, and blue. The coat protein dimers on the left of the opening are displaced less than those on the right side of the opening, as can be seen in the difference between the colored (asymmetric) and grey (symmetric) ribbons from the protein backbone.

The displacement of coat proteins is not restricted to those in the immediate vicinity of the maturation protein. When docking the symmetric with asymmetric capsids, many coat protein dimers are shown to be displaced by the maturation protein and gRNA (Figure II-14). The coat protein dimers are more displaced in the immediate

vicinity, but also in the concentric ring of coat protein dimers around those. The further away from the maturation protein, the less the displacement of coat protein dimers, to the point where there is little movement once one gets to the 'equator' of the virus, and there is no movement at the 'South Pole' of the capsid.

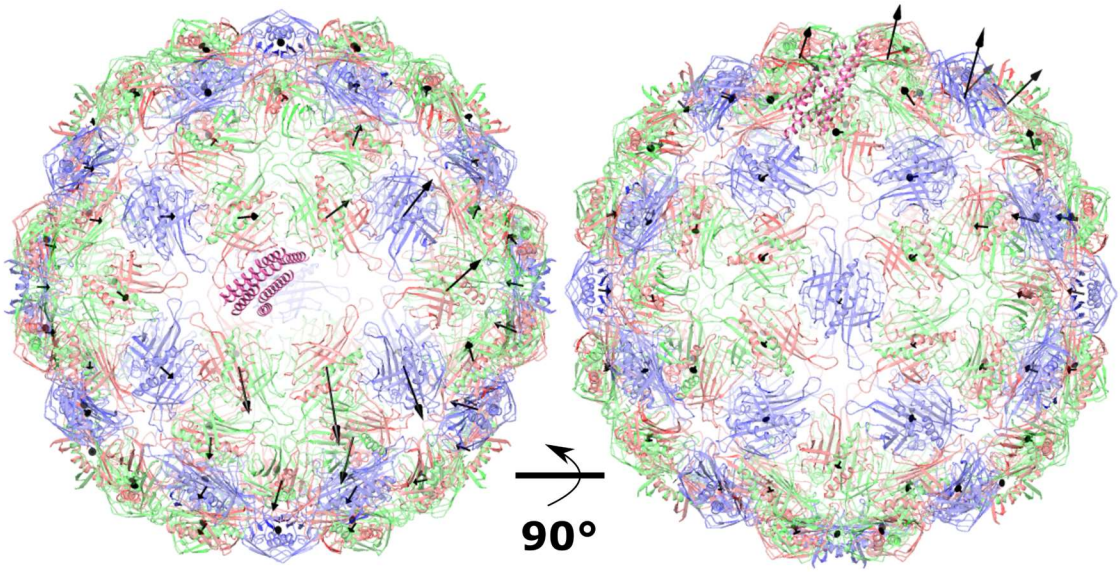

Figure II-14. Coat protein dimer deviation between the asymmetric and icosahedral cryoEM reconstructions of Qβ
The symmetric and asymmetric reconstructions were overlaid onto each other, with the crystal structure for the coat protein dimers docked into each one. The difference in positions between the asymmetric and symmetric is depicted by black arrows which are proportional to the distance difference between the two structures. The maturation protein and proximal gRNA causes the displacement, and this displacement is greater closer to the maturation protein. The left image is the top-down view of the virus, while the right is rotated by 90 degrees. The displacement decreases further away from the maturation protein, as seen by smaller arrows. The capsid shown is the asymmetric reconstruction with the proteins depicted as ribbon models. The coat proteins are colored according to their conformer, A, B, and C, are red, green, and blue, respectively. The α-region of the maturation protein is shown as a 5-helix bundle. The β-region is omitted.

The displacement is not only caused by the maturation protein, but also the gRNA, which goes through the capsid, occupying some of the space from the displaced coat protein dimer, but not significantly outside of the capsid. The gRNA could theoretically go further outside of the capsid and be blurred out of our electron density if it was flexible. However, if this was the case then it would be exposed to RNases and could get degraded and the viruses would be noninfectious. The gRNA sticking through the capsid, in the space that the displaced coat protein dimer would have been, is visible in Figure II-15. The interactions of the gRNA and the maturation protein, are very stable, as the electron density of the gRNA is strong. This presumably will play a role in the uptake of the gRNA (and potentially the maturation protein).

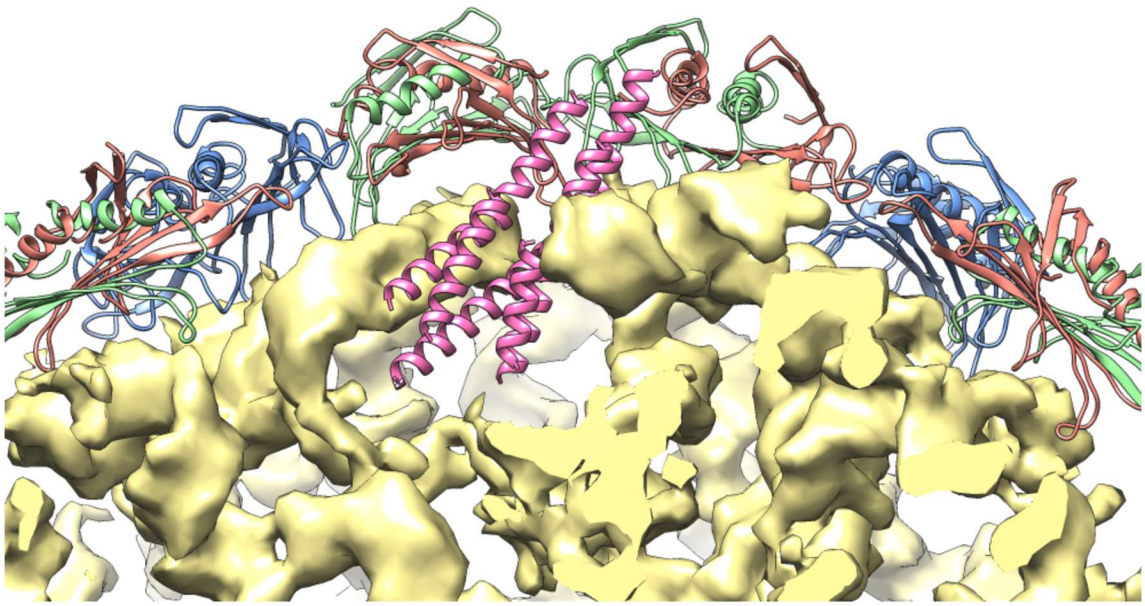

Figure II-15. Maturation protein and gRNA complex at the pole of Qβ
A cross-section of the asymmetric reconstruction of Qβ shows how the gRNA (yellow electron density) pushed through the capsid when a coat protein dimer is displaced by the maturation protein. The coat proteins are shown as ribbon

models, while only the α-region of the maturation protein is shown, also as a ribbon model. The coat protein ribbon models are colored based on their conformer type, with A, B, and C, colored as red, green, and blue, respectively. The maturation protein ribbons are colored in pink.

While the fissure in the capsid of Qβ should facilitate the release of the gRNA, the maturation protein binding to the gRNA certainly plays a role. The maturation protein of MS2 was shown to be taken up into the host along with the gRNA (Krahn, O'Callaghan et al. 1972). Subsequently, the maturation protein of MS2 was shown to bind the gRNA in vitro in two places, although in virio that is disputed (Shiba and Suzuki 1981, Dai, Li et al. 2017). While there is no further biochemical evidence of a maturation protein / gRNA complex, our structure of Qβ, and that of MS2 (Gorzelnik, Cui et al. 2016, Dai, Li et al. 2017) suggests that the binding of the gRNA by the maturation protein is very specific. The strength of the gRNA/MP interaction is evident by the higher resolution of the RNA electron density proximal to the MP. The further away from the maturation protein, the weaker the electron density of the gRNA (Figure II-16). The electron density is stronger in the 'northern hemisphere' of the virus, both in our structure and that of MS2. A striking difference between our structure and that of MS2, is that Qβ has many long-distance interactions, previously hypothesized but not experimentally verified (Beekwilder, Nieuwenhuizen et al. 1995, Klovins, Berzins et al. 1998, Klovins and van Duin 1999). After more data was collected and the resolution of the gRNA increased, this long helix was determined to be a three-way junction between multiple stem-loops (Jeng-Yih Chang, personal communication).

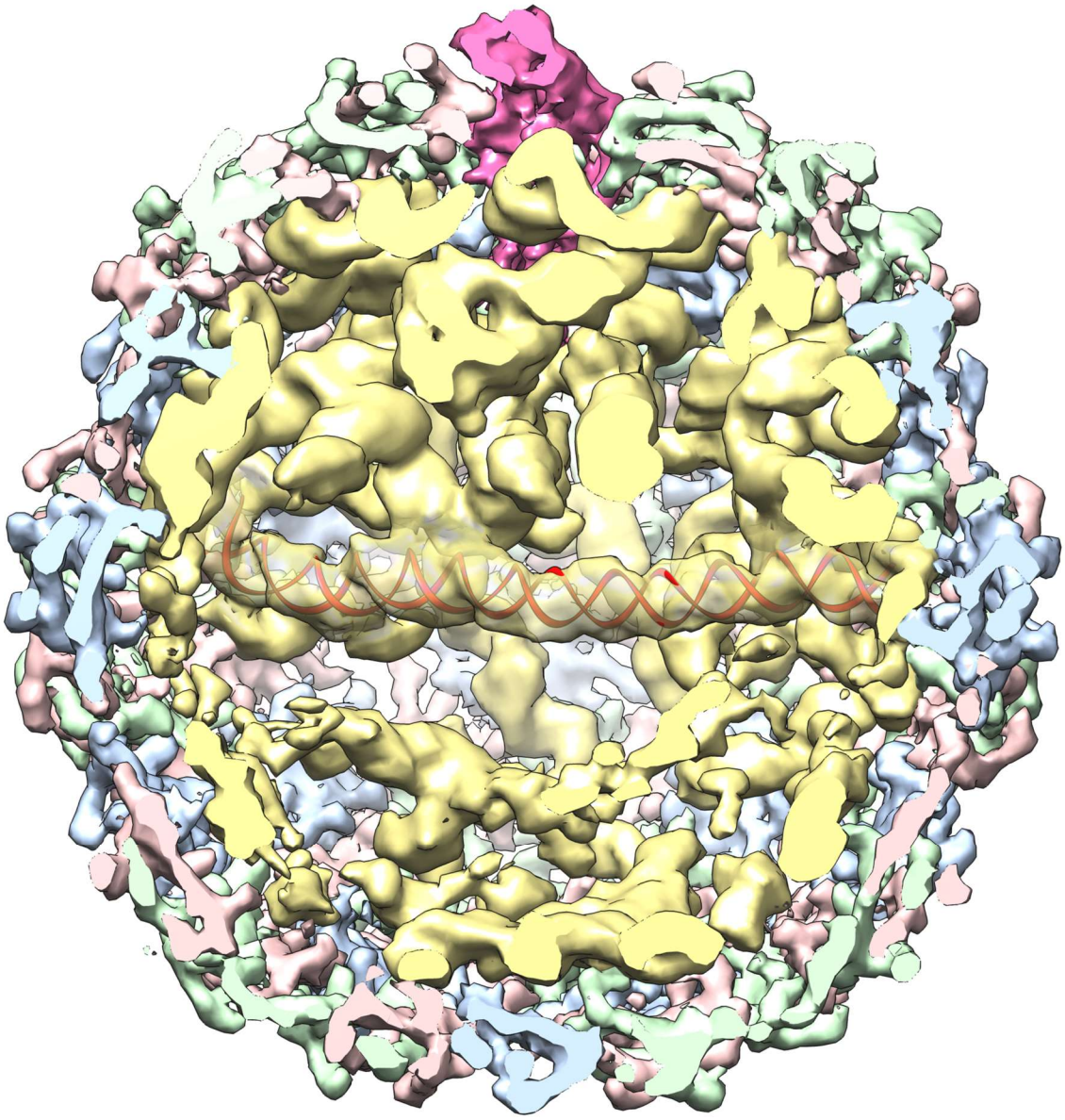

Figure II-16. Asymmetric reconstruction of Qβ showing the gRNA within the capsid

The coat protein dimers are colored according to their conformer type, A, B, and C monomers are colored red, green, and blue, respectively. The maturation protein is shown in pink. The gRNA inside the capsid is colored yellow, with a long helix modeled into the electron density. The electron density for the gRNA is more stable in the 'northern hemisphere' proximal to the maturation protein, while further away from the maturation protein the RNA is more flexible and thus appears to be less dense, this is just from the cutoff threshold used. The whole capsid is filled with RNA, where it is more flexible the electron density is less,

and so at the threshold used it appears empty, when in fact there is RNA there. An RNA helix is modeled into electron density that crosses the middle of the virion.

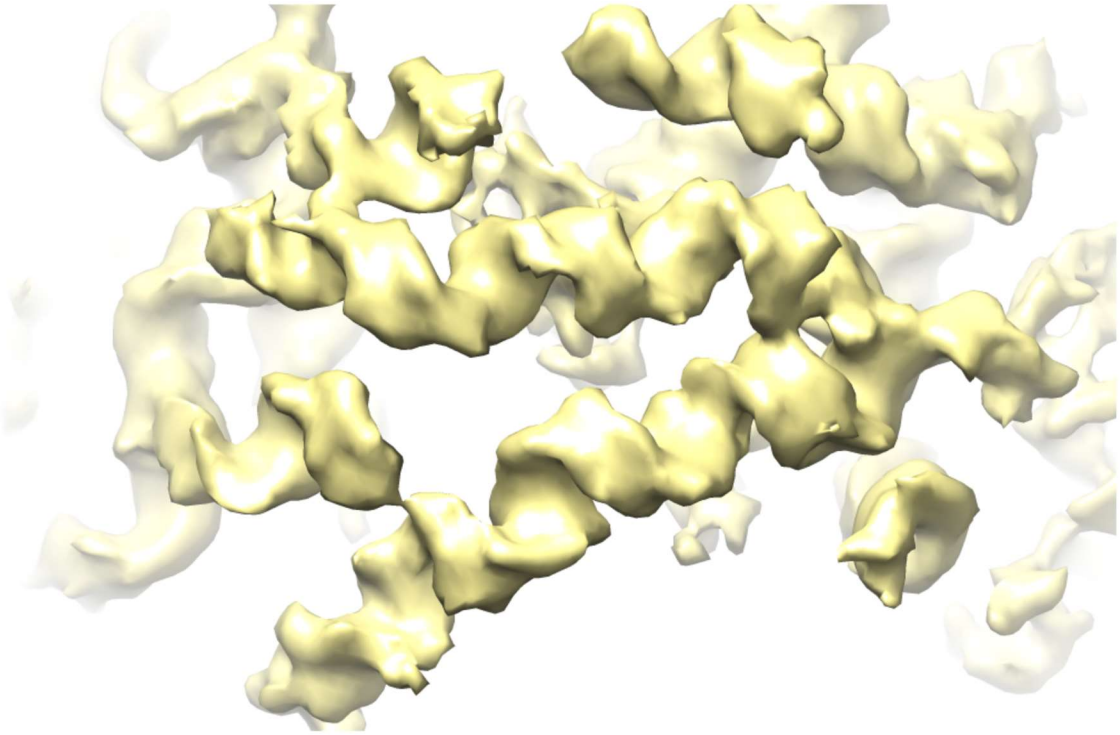

Figure II-17. RNA helices within the asymmetric reconstruction of Qβ
The capsid is computationally removed. The interior electron density is colored in yellow. The electron density shows clear grooving, akin to is seen for nucleic acids and not proteins.

The gRNA electron density is more stable proximal to the capsid and the maturation protein (Figure II-16). The major and minor grooves of RNA helices are clearly visible (Figure II-17). While there are many long helices in the genome, about ⅔ of the coat protein dimers touch stem-loops, whether operator-like or those with reverse-handedness from the operator. The coat protein for Qβ has an internal-facing loop with

positive charges, the EF-loop, named for the β-strands before and after it (Golmohammadi, Fridborg et al. 1996). The positive charges in this loop are conserved in phage from the genus *Allolevivirus*, whereas phage from the related genus *Levivirus*, such as MS2, do not have positive charges in this loop (Figure II-18) (Golmohammadi, Valegard et al. 1993, Golmohammadi, Fridborg et al. 1996).

```
                  aaaa aa     bbb bbcccccc     dddddd          eee
MS2       1 ASNFTQFVLV DNGG..TGDV TVAPSNFA.. NGVAEWISS. NSRSQAYKVT
fr        1 ASNFEEFVLV DNGG..TGDV KVAPSNFA.. NGVAEWISS. NSRSQAYKVT
GA        1 .ATLRSFVLV DNGG..TGNV TVVPVSNA.. NGVAEWLSN. NSRSQAYRVT
Qβ        1 .AKLETVTLG NIGKDGKQTL VLNPRGVNPT NGVASLSQAG AVPALEKRVT
SP        1 .AKLNQVTLS KIGKNGNQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
E. coli             L    G           P       NGVA              +VT
PRR1      1 .AQLQNLVLK DREAT.PNDH TFVPRDIRD. NVGEVVEST. GVPIGESRFT
PP7       1 ...SKTIVLS VGEAT.RTLT EIQST..... ADRQIFEEKV GPLVGRLRLT
All               L                                          + T

             eee          ffffffff fffff           gggg ggggggggg
MS2      46 CSVRQSSA.. QNRKYTIKVE VPKVATQTVG .GVELPVAAW RSYLNMELTI
fr       46 CSVRQSSA.. NNRKYTVKVE VPKVATQVQG .GVELPVAAW RSYMNMELTI
GA       46 ASYRASGA.. DKRKYAIKLE VPKIVTQVVN .GVELPGSAW KAYASIDLTI
Qβ       50 VSVSQPSRNR KNYKVQVKIQ NPTACTANG. ..SCDPSVTR QAYADVTFSF
SP       50 VSVAQPSRNR KNFKVQIKLQ NPTACTRDA. ...CDPSVTR SAFADVTLSF
E. coli     S            K   K    P             P
PRR1     47 ISLRKTSN.. GRYKSTLKLV VPVVQSQTVN .GIVTPVVVR TSYVTVDFDY
PP7      42 ASLRQNGAK. TAYRVNLKLD QADVVDCSTS VCGELPKVRY TQVWSHDVTI
All         S            +   K                   P

                AAAAA AAAAAAAAAA A   BBBBBB BB
MS2      93 PIFATNSDCE LIVKAMQGLL KD.GNPIPSA IAANSGIY.
fr       93 PVFATNDDCA LIVKALQGTF KT.GNPIATA IAANSGIY.
GA       93 PIFAATDDVT VISKSLAGLF KV GNPIAEA ISSQSGFYA
Qβ       97 TQYSTDEERA FVRTELAALL AS..PLLIDA IDQLNPAY.
SP       96 TSYSTDEERA LIRTELAALL AD..PLIVDA IDNLNPAY.
E. coli         -                          A I      Y
PRR1     94 DARSTTKERN NFVGMIADAL KADLMLVHDT IVNLQGVY.
PP7      91 VANSTEASRK SLYDLTKSLV ATSQVEDLVV NLVPLGR..
All
```

Figure II-18. The EF-loop in coat proteins of ssRNA phages is positively-charged for those in *Allolevivirus*
Members of *Levivirus* (the first three: MS2 and fr which are essentially the same virus, while GA is a virus in the same genus) do not have an internal-facing positively charged loop, whereas members of *Allolevivirus* (Qβ and SP) have

positive residues on their internal-facing loop. PP7 and PRR1 are also ssRNA phages but are not always placed within *Levivirus* or *Allolevivirus*, depending on the authors. Adapted with permission from (Golmohammadi, Fridborg et al. 1996).

The SRNRK loop in Qβ could play a role in the assembly of the virus. It is thought for all ssRNA phages that the coat protein is recruited to the gRNA based off of structures akin to their 'operator,' a stem-loop at the start of the replicase (Rumnieks and Tars 2018). The SRNRK loop comes in contact with the gRNA all throughout Qβ, but this feature is more explainable due to its location within the coat protein than due to the positive charges in the loop.

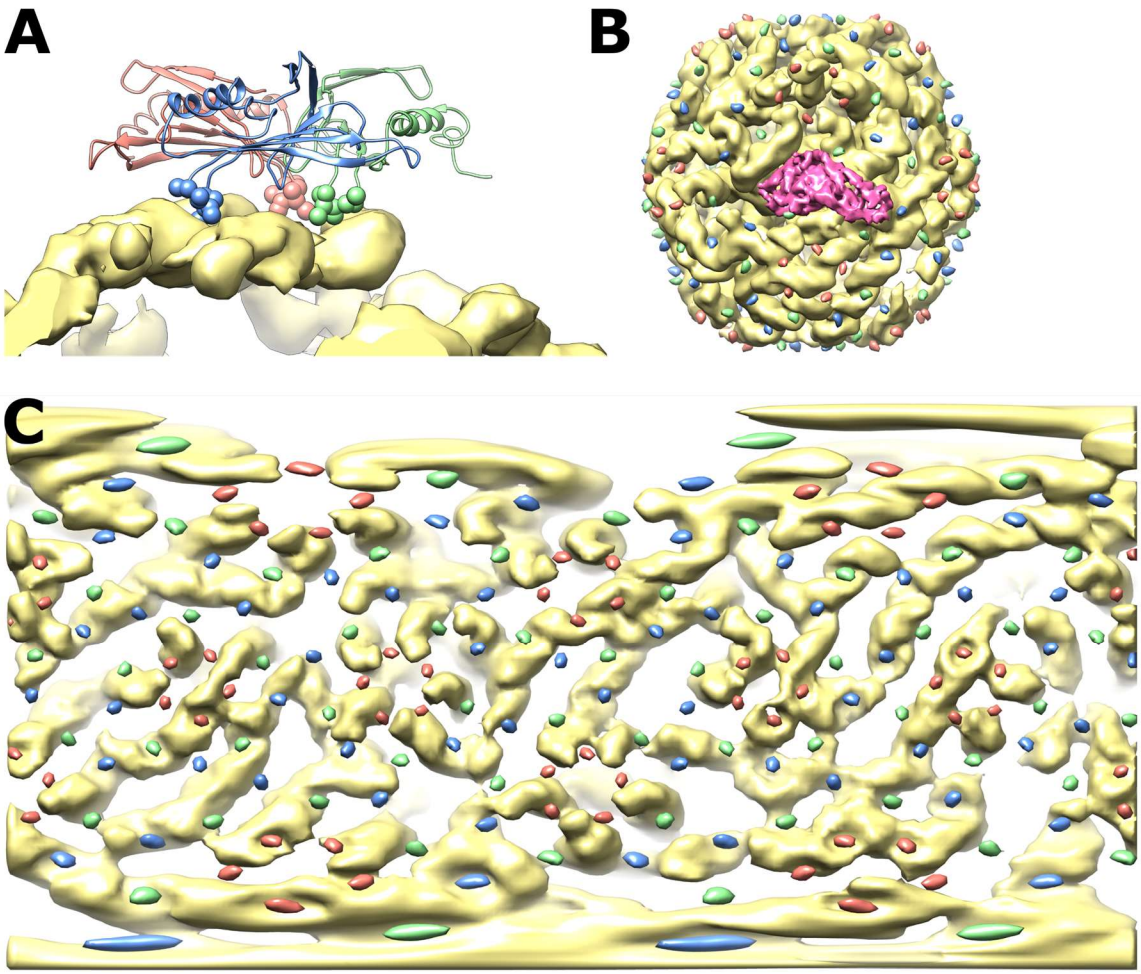

Figure II-19. The internal-facing SRNRK loop interacts with the gRNA in every position of the coat protein within the capsid of Qβ
The coat proteins are shown as a ribbon model (in A) are colored, as previously, with red, green, and blue colors representing A, B, and C monomers (respectively). The five amino acids representing the SRNRK loop are represented as bubbles. In panels B and C, the 5 amino acid SRNRK loop is shown, while the remainder of the protein is shown as a ribbon model. Panel B shows the dominant conformation of the gRNA within the Qβ structure, with bubbles shown for the SRNRK loop, while the rest of the density for the coat proteins removed. Panel C is a Mercator Projection of the virus, with the SRNRK loops colored as in A and B, minus the rest of the coat protein.

However, the positively-charged SRNRK loop could bind to the negatively-charged RNA with less specificity than a loop with neutral amino acids. The electron

density for the RNA of Qβ is less stable than that of the related virus MS2. Our structures of Qβ only used ~30% of the particles, whereas that from MS2 used close to 100% of the particles (Gorzelnik, Cui et al. 2016, Cui, Gorzelnik et al. 2017, Dai, Li et al. 2017).

In an effort to identify the specific binding interactions of the Qβ coat protein with the gRNA, the asymmetric structure was examined for the operator. The operator is an RNA stem-loop at the start of the replicase to which the coat protein has a very high affinity toward (van Duin and Tsareva 2006). For Qβ, this structure was crystallized fairly recently (Figure II-20).

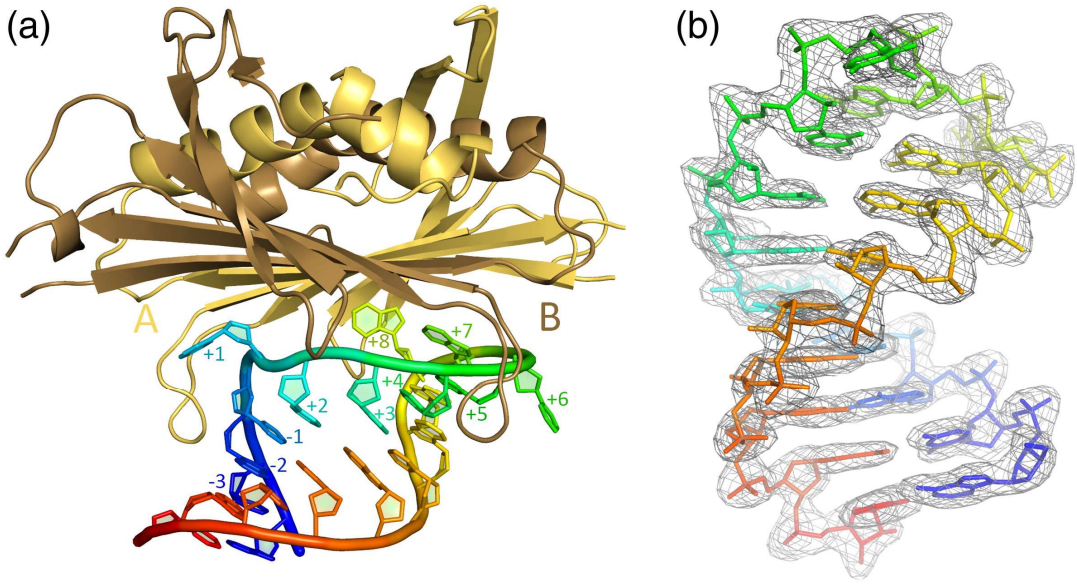

Figure II-20. Coat-Operator complex for Qβ
The coat-operator complex for Qβ is shown panel A, with the coat protein dimer in brown and the RNA of the operator rainbow colored. Panel B shows the electron density of RNA colored as in A. Figure adapted with permission from Rumnieks and Tars, 2014.

To determine the location of this complex in the asymmetric structure, the crystal structure was docked into the electron density. As it turns out there are thirty places where the operator-coat protein crystal structure can be docked into the electron density perfectly (Figure II-21). There are another twenty plus locations where the coat protein dimers bind to a gRNA stem-loop, but the loops have an orientation that is an opposite-handedness of the crystal structure. In the other cases the gRNA comes in contact in the middle of RNA helices or on stem-loops that form junctions.

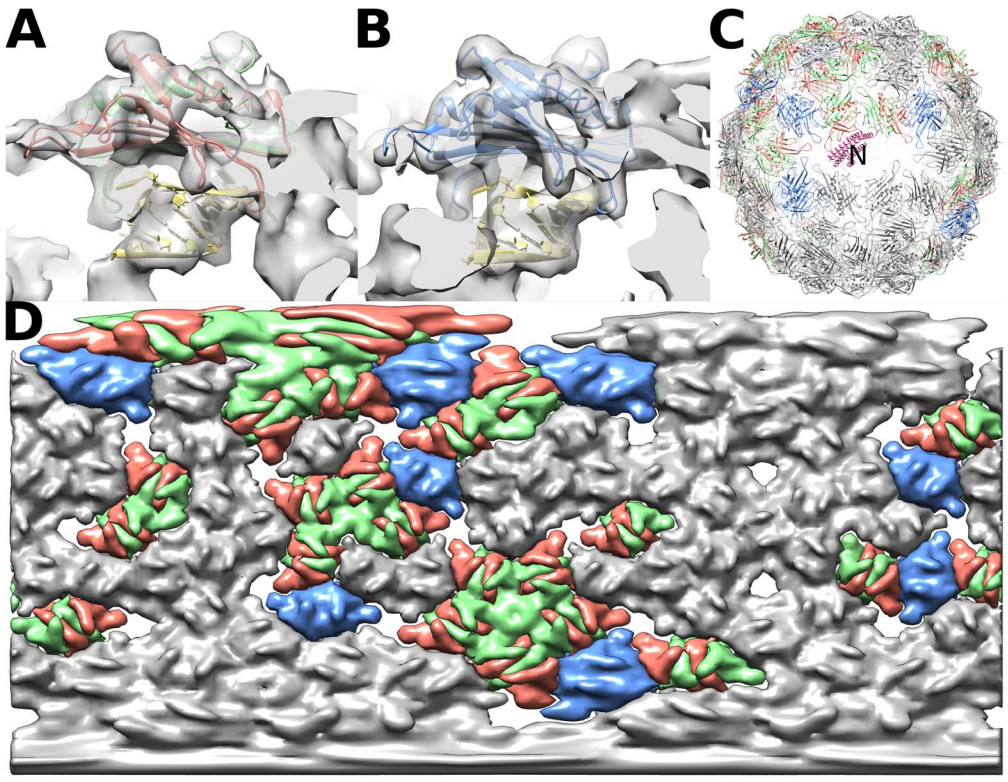

Figure II-21. Positions within the capsid where the crystal structure of a coat protein dimer bound to the operator can be fit into the electron density
The crystal structure of the operator stem-loop bound to the coat protein dimer (from Rumnieks and Tars 2014) is docked into the cryoEM electron density. A)

An A/B dimer with the ribbon model from the crystal structure colored as red and green, and the RNA colored in yellow. B) The crystal structure fit into a C/C dimer, with the ribbon model colored as blue. C) Ribbon models of the coat protein dimers and the maturation protein (pink) that make up the entire capsid. The coat protein dimers which do not interact with operator-like hairpins are colored in gray. D) A Mercator projection of the electron density for the capsid, with the coat protein dimers which interact with operator-like stem-loops colored according to their conformation. As in Panel C, the coat proteins which do not interact with operator-like stem-loops are colored in gray. The electron density for the maturation protein is excluded in this panel.

Why are there so many operator-like binding events? Genetic studies for Qβ and MS2 have identified the minimum sequences for the operator (Lim, Spingola et al. 1996). For Qβ this is an RNA stem, with a 3nt loop, and an adenosine in the third position of the loop. In order to identify potential operator-like sequences in the genome, which might fold into a structure similar to the operator, the genome was scanned for the minimum sequence of the operator (Figure II-22). There are at least twenty-six such sites throughout the genome, but not all are feasible (as some are too close to one another or have other interactions). This is a small subset of the potential operators, and a complete analysis of the RNA secondary structure within the capsid will be needed to determine the specificity of the coat proteins for the RNA.

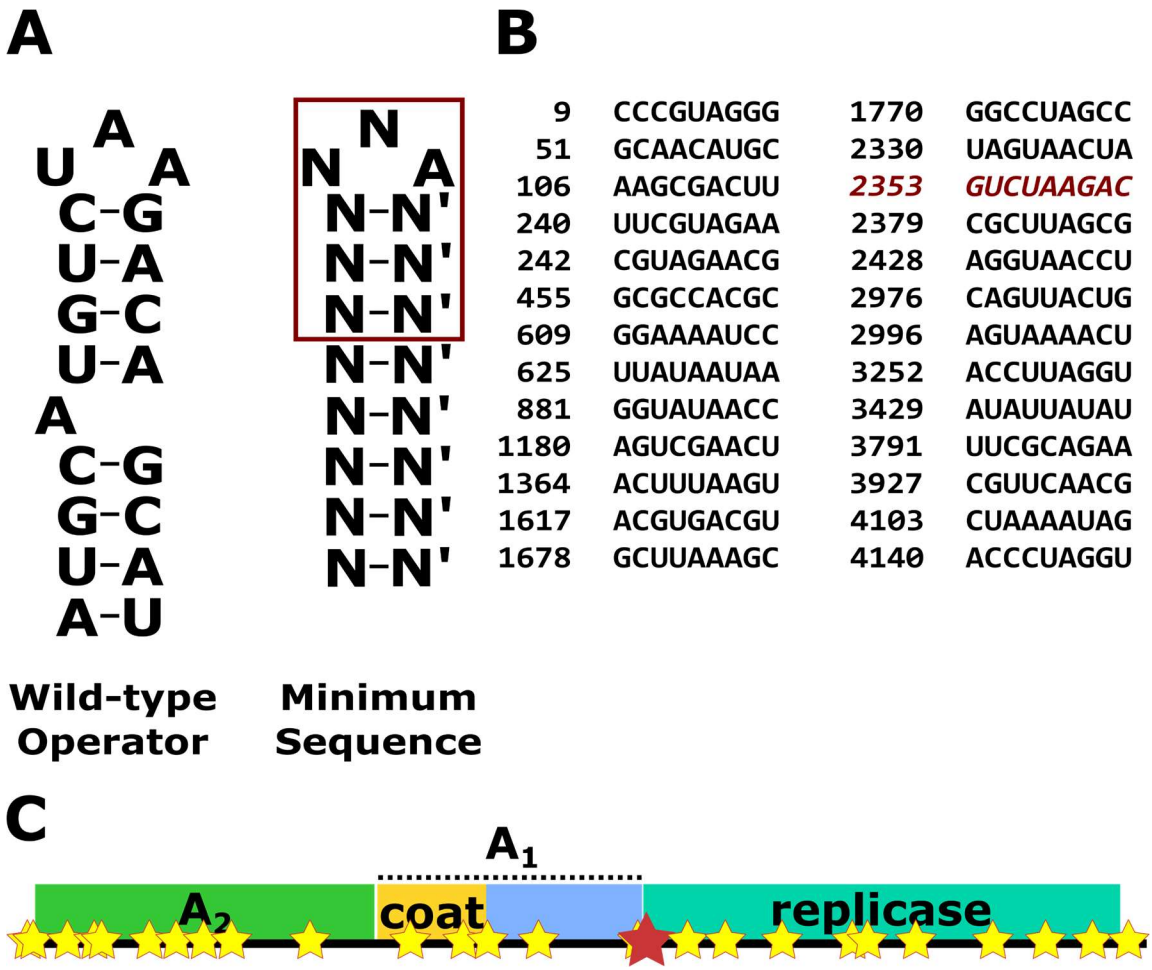

Figure II-22. Potential 'operator-like' sequences
A) The wild-type operator, at the start of the replicase, is shown in the left of panel A, while the minimum consensus sequence is shown on the right of the panel. B) Taking 3nt stem with a 3nt loop with an adenosine in the third position (shown as a maroon box in panel A), and scanning the genome of Qβ, there are at least 26 potential 'operator-like' sequences, with the true operator shown in maroon italics. C) The 'operator-like' sequences seen in panel B are seen on the genome of Qβ in yellow stars, with the true operator seen as a maroon star.

2.4. Discussion

Unlike dsDNA phages, which pump their DNA into a preformed capsid, ssRNA phages assemble their capsids around the gRNA (Gelbart and Knobler 2009, Stockley,

White et al. 2016). Inherently, this allows for a certain level of promiscuity but which RNA is the correct RNA? How do you select for it?

The overall structure of the capsid is still about the same between the crystal structure, the symmetric cryoEM reconstruction, and the asymmetric reconstruction. What changed was the appearance of the maturation protein. Never before had the maturation protein of Qβ been visualized. Our reconstruction of Qβ came out shortly after an asymmetric reconstruction of MS2, which also showed the maturation protein (Koning, Gomez-Blanco et al. 2016). Shortly thereafter another group published a lower resolution asymmetric reconstruction of MS2 (Zhong, Carratala et al. 2016). A few months later a much higher resolution structure of MS2 was produced, which showed the maturation protein and better resolution of the gRNA (Dai, Li et al. 2017). It appears the low-hanging fruit was grasped by multiple parties, with one pushing further to enhance the story.

The structures of Qβ and MS2 are relatively similar. The major differences within the coat protein were subject to a thorough review by Golmohammadi in his paper on the crystal structure of the Qβ capsid (Golmohammadi, Fridborg et al. 1996). The biggest structural difference between the two phages is the presence of a few copies of the readthrough protein A_1, a minor capsid component. A_1 is required for infection via an unknown mechanism (Hofstetter, Monstein et al. 1974). A_1 does not appear to have a favored location within the assembled phage, as such, it was not found within our

reconstruction. Since the N-terminal domain of A_1 is the coat protein, A_1 could be located anywhere within the capsid. The absence of electron density for A_1 within our reconstruction is due to the flexible linker between the C-terminal domain and the coat protein domain.

Overall, the maturation proteins for Qβ and MS2 are remarkably similar, as would be expected since they share the same roles of binding to the gRNA as well as the F-pilus. They both have an α-region which interacts with the gRNA and the β-region, which binds to the F-pilus. Both bind the gRNA in a similar fashion, through a large amount of positive charges in the α-region. The β-region of the maturation protein of Qβ is also used to bind to the lysis target, MurA, and so is probably less specialized for binding the F-pilus. In competition experiments, whether planned or by accident, MS2 will overtake Qβ when the two are used to infect *E. coli*.

Capsid assembly can be triggered *in vitro*, without viral RNA by increasing the concentration of coat proteins to a certain level (Rolfsson, Toropova et al. 2010). Assembly can be initiated at a lower level with phage RNA added in or just adding in RNA stem-loops with the operator for that particular phage (Beckett and Uhlenbeck 1988, Beckett, Wu et al. 1988). The role of the operator has historically been regarded as being for translational control of the replicase (Weissmann 1974). Once there are sufficient levels of phage RNA (and thus replicase) that the molar concentration of coat proteins reaches a certain level, then there is also a large amount of the lysis protein, so

the phage should stop replicating and start packaging. The operators are known for many well studied ssRNA phages, i.e. those identified before Krishnamurthy's 'hyperexpansion' of the ssRNA phage viriome (Krishnamurthy et al., 2016). Some have proposed the operator as a site of nucleation for the capsid, since the operator was the highest affinity stem-loop within the phage RNA (Rolfsson, Middleton et al. 2016, Garmann, Goldfain et al. 2019). However, if the operator is deleted the capsid is still able to assemble around the viral RNA (Pickett and Peabody 1993). This is presumably because there are many operator-like sequences within the genome.

CHAPTER III

KNOWLEDGE GAINED FROM THE Qβ STRUCTURE ENABLED DETERMINATION OF INFECTION REQUIREMENTS[2]

3.1. Introduction

Single-stranded RNA phages have highly branched genomes, for multiple reasons: to protect against host RNases (Pasloske, Walkerpeach et al. 1998); to prevent replication dead ends when dsRNA is formed if the product anneals to the template, which the replicase holoenzyme cannot replicate (Tomita, Ichihashi et al. 2015); for assembly as the coat proteins bind to RNA stem-loops with better affinity than in the middle of RNA helices (Kelly, Grosberg et al. 2016); as well as being more stable in the capsid than unpaired bases or than an RNA of all helices (Beren, Dreesens et al. 2017). Many groups have found that ssRNA viruses are more compact than regular mRNAs, owing to the requirements for self-assembly and packaging in a tight shell (Yoffe, Prinsen et al. 2008).

[2] Some of the work comes from 'Structures of Qβ virions, virus-like particles, and the Qβ–MurA complex reveal internal coat proteins and the mechanism of host lysis' by Zhicheng Cui, Karl V. Gorzelnik, Jeng-Yih Chang, Carrie Langlais, Joanita Jakana, Ry Young, and Junjie Zhang *Proceedings of the National Academy of Sciences of the United States of America* 2017 114:44 11697–11702. Figures are republished with permission.

3.2. Materials and Methods

3.2.1. Bacterial Strains and Media.

To increase the resolution of the structure of Qβ, the virus was purified as in Chapter 2. Mutants of Qβ were made using pBRT7QB, a fully sequenced plasmid graciously supplied by Sean Leonard, University of Texas. Other cDNA plasmids of Qβ were also supplied by Charlotte Knudsen, University of Åarhus; Nori Ichihashi, Osaka University; Alain Waffo, Alabama State University; René Olsthoorn, Leiden University; David Mills, by way of Carl Dobkin, as Mills' lab had shut down.

E. coli was grown with at least 1:4 aeration (500ml in a 2L flask, or 3ml in a 15ml tube) at 37°C with >200 r.p.m. shaking, in Luria Broth (10g/L tryptone, 10g/L NaCl, 5g/L yeast extract). When appropriate antibiotics were added: 100µg/ml ampicillin, 50µg/ml kanamycin, 33µg/ml chloramphenicol, or 5µg/ml tetracycline. As in Chapter II, the type of tetracycline mattered, so tetracycline hydrate was used. cDNA plasmids with Qβ were propagated in *E. coli* DH10B (F^-) to ensure reinfection did not occur.

3.2.2. Qβ 3' Mutations

Standard molecular biology techniques were used. The Qβ cDNA plasmid used for the R1 and U1 mutations was pBRT7QB, although the primers used could work with most any cDNA plasmid derived from the Anc(P1) sequence (Kashiwagi, Sugawara et al. 2014). Primers used are listed in Table III-1.

Table III-1. Primers used for R1 and U1 mutations

#234	R1 Stem-loop deletion F.	GGGGTCTTCCAGGGCACGtaaTGGGAGGGC GCCAATATGG
#235	R1 Stem-loop deletion F.	CCATATTGGCGCCCTCCCAttaCGTGCCCTG GAAGACCCC
#238	R1 Scrambled F.	CTTCCAGGGCACGAAGGTaGCaagcttgCAtG AaGCtTgAggTGGGAGGGCGCCAATATG
#239	R1 Scrambled R.	CATATTGGCGCCCTCCCAccTcAaGCtTCaTG caagcttGCtACCTTCGTGCCCTGGAAG
#240	U1 loop alteration F.	CTTACGAGTGAGAGGGGGTTGCCCTCTCT CCTCC
#241	U1 loop alteration R.	GGAGGAGAGAGGGCAACCCCCTCTCACTC GTAAG
#242	U1 loop deletion F.	CACAATTACTCTTACGAGTCTCCGGGGGA TCCACTAG
#243	U1 loop deletion R.	CTAGTGGATCCCCCGGAGACTCGTAAGAG TAATTGTG

The cDNA template used was pBRT7QB, with a temperature gradient for the annealing temperature. Extension times ranged from 15s/kb to 1min/kb. DMSO was added to 4-10% of the total reaction volume. Template plasmids were degraded using DpnI, and the PCR product gel purified. Gel purified PCR products were transformed into DH10B.

3.2.3. Western Blots for A_2

The 3' mutants, plus wild-type, were grown in *E. coli* DH10B (F^-), as was a negative control, DH10B with an empty pBR322 plasmid. 5ml cultures were inoculated from a single colony and grown with 100µg/ml ampicillin. The cultures were grown overnight, then diluted to $OD_{600} = 0.5$, with the OD measured again after dilution. The samples were spun down, then the supernatant was incubated on ice for 30min before the

TCA precipitation. Samples were TCA precipitated by adding ice cold TCA to 10% final volume, and acetone to 50% final volume. The sample was spun down at 18,000 x *g* for 15min at 4°C and the supernatant aspirated off. The pellet was washed twice with ice cold acetone, then resuspended in sample loading buffer, boiled for 15min, then loaded onto a 4-20% Tris-Tricine gel. The gels were run for ~1hr, then transferred to a membrane using a semidry apparatus and blocked for at least half an hour with 2% milk in TBS with 0.1% Tween 20. The buffer was replaced with fresh buffer which had an α-A_2, a gift from the Young lab, derived against a synthetic peptide PKLPRGLRFGA (from the N-terminal end of A_2) (Reed, Langlais et al. 2013). The antibody was used at a 1:1,000 dilution. Several dilutions were tried, but this showed the best results. After applying the antibody overnight the membranes were washed 5 times for at least five minutes each time with TBST. The sample was then incubated with a secondary antibody (1:3,000 dilution of goat-anti-rabbit, kindly provided by the Gohil lab).

3.3. Results and Discussion

After increasing the overall number of micrographs of Qβ, from 712 used in the initial asymmetric reconstruction, to 2,370, the overall number of particles was correspondingly increased from 51,815 to 248,445 (Gorzelnik et al., 2016; Cui et al., 2017). The number of particles able to be used after refinement increased from 12,975 to 76,843 (Gorzelnik et al., 2016; Cui et al., 2017). The overall resolution increased from 7 Å to 4.7 Å (Figure III-1) (Gorzelnik et al., 2016; Cui et al., 2017).

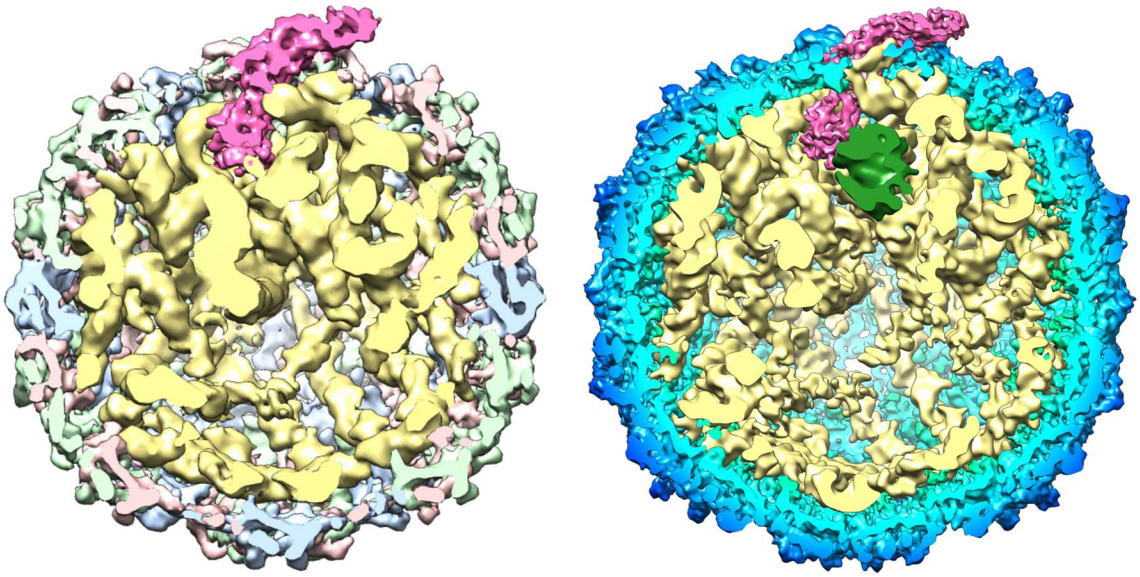

Figure III-1. Comparison of Qβ with more or less data
The published structure of Qβ (left) had 712 micrographs and reached 7 Å resolution. After collecting more data, 2,370 micrographs, enabled increasing the resolution of the structure to 4.7 Å. The figures are colored as per their respective papers. The coat proteins on the left colored as per their conformer type, A, B, and C are light red, light green, and light blue, respectively. The coat proteins on the right are colored radially, with the region in the interior colored a lighter blue than the exterior portion. The gRNA is colored the yellow for both, with the maturation protein colored in pink. The green density in the improved structure is protein density that does not belong to the maturation protein.

The cryoEM density of the virion is similar to that of the previous Qβ structure. The maturation protein, which was first visualized in the asymmetric structure then subsequently crystallized by another group, appears the same in the β-region (Cui et al., 2017; Rumnieks and Tars 2017). However, at the base of the α-region we had mistakenly colored electron density as belonging to the maturation protein in the initial paper on the asymmetric structure (Gorzelnik et al., 2016). This mistake is understandable, inasmuch as the electron density was clearly protein and not RNA.

With increased resolution, and the crystal structure of A_2, the extra electron density was determined not to be part of the maturation protein but also not RNA (Figure III-2).

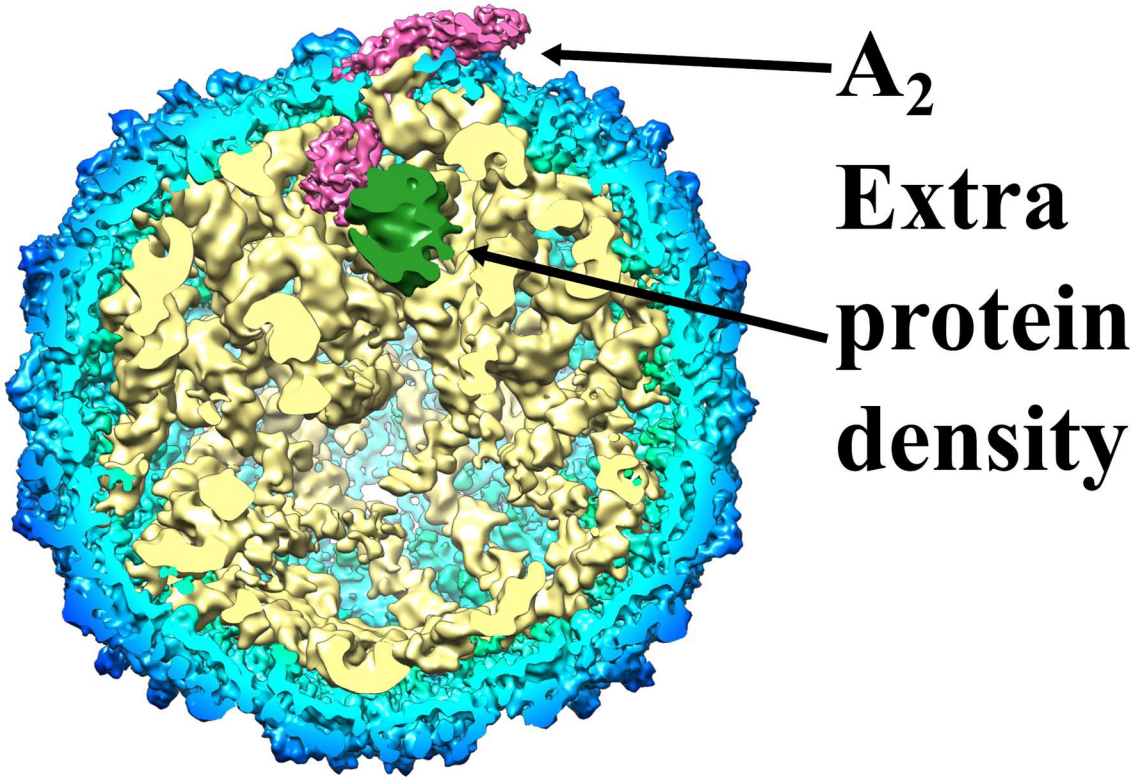

Figure III-2. Increased resolution of Qβ reveals extra protein density by the maturation protein
Docking the then newly solved crystal structure of A_2 into the increased resolution structure of Qβ revealed that the maturation protein has extra protein density proximal to the β-region. The electron density did not have characteristics of RNA.

The extra protein density could have been a host protein from *E. coli*, but it was present in the dominant classes, which all had the maturation protein and made up about 30% of the overall particles (Cui et al., 2017). Since the copy number was high enough, it could not be an *E. coli* protein, as previous studies of purified Qβ showed that there are

just three protein components in the purified virus (Takamatsu and Iso 1982). These proteins are: the coat protein, which predominates the other proteins; A_1, the elongated coat protein, which is at levels of 3-15 copies per virion; and the maturation protein, present at about one copy per virus (Weber and Konigsberg 1975). With that knowledge in hand, we docked in the crystal structures of each of these proteins. The only crystal structure which fit the electron density was that of a coat protein dimer (Figure III-3).

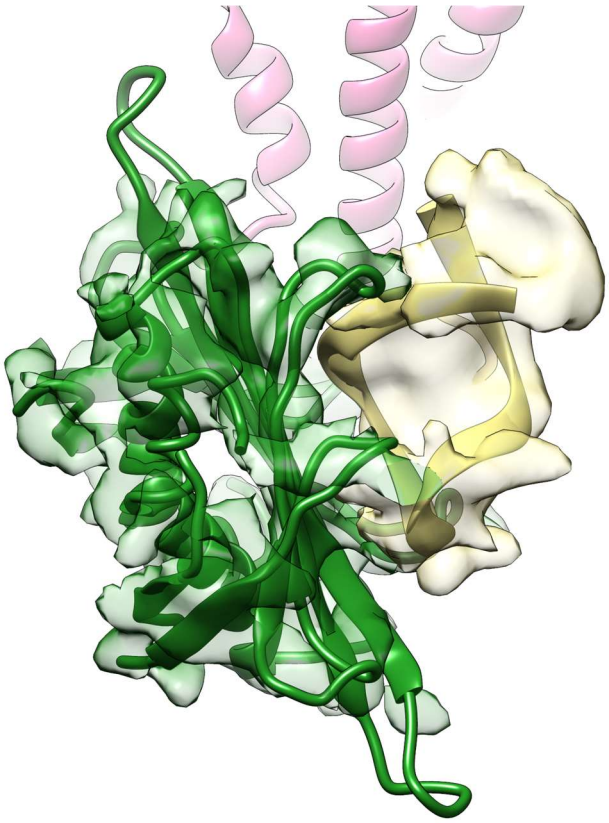

Figure III-3. The extra protein density belongs to an internally sequestered coat protein dimer
The electron density of the region of Qβ near the base of the maturation protein, which is protein density, colored in a transparent green. This region very closely fits the crystal structure of a coat protein dimer. The electron density belonging to gRNA is colored in a transparent yellow, with an RNA stem-loop modeled in.

The region belonging to the maturation protein, A_2, is shown with a ribbon model from the crystal structure of the maturation protein, fit into where the electron density was before masking it out.

That this sequestered coat protein dimer is there in the 30% of particles which have good density for A_2 shows that this is not a mistake of packaging, but rather a conserved trait of Qβ. To confirm this hypothesis more work would be needed. One of my labmates, Jeng-Yih Chang, undertook the task of modeling the gRNA into the electron density of the improved asymmetric structure. While the resolution of the gRNA is not uniform, the electron density was strong enough to trace the backbone. With the knowledge of the secondary structure of Qβ, from two papers and a dissertation (Beekwilder et al., 1995; Beekwilder et al., 1996; Beekwilder dissertation) he was able to model in the secondary structure of the entire Qβ RNA into the electron density. What Jeng-Yih found was that the maturation protein came in contact with the gRNA at the terminal stem-loop, designated by Beekwilder et al., as U1 (Figure III-4) (Beekwilder et al., 1995).

As Jeng-Yih modeled in the entire genome, he also determined where the true operator interacts with the capsid, as well as several other interactions which are distinct in Qβ, such as the long-distance interactions identified previously (Klovins, Berzins et al. 1998, Klovins and van Duin 1999). These observations will be discussed in Jeng-Yih's dissertation. The interactions of the 3' end of the gRNA will be discussed here, as they were biochemically validated.

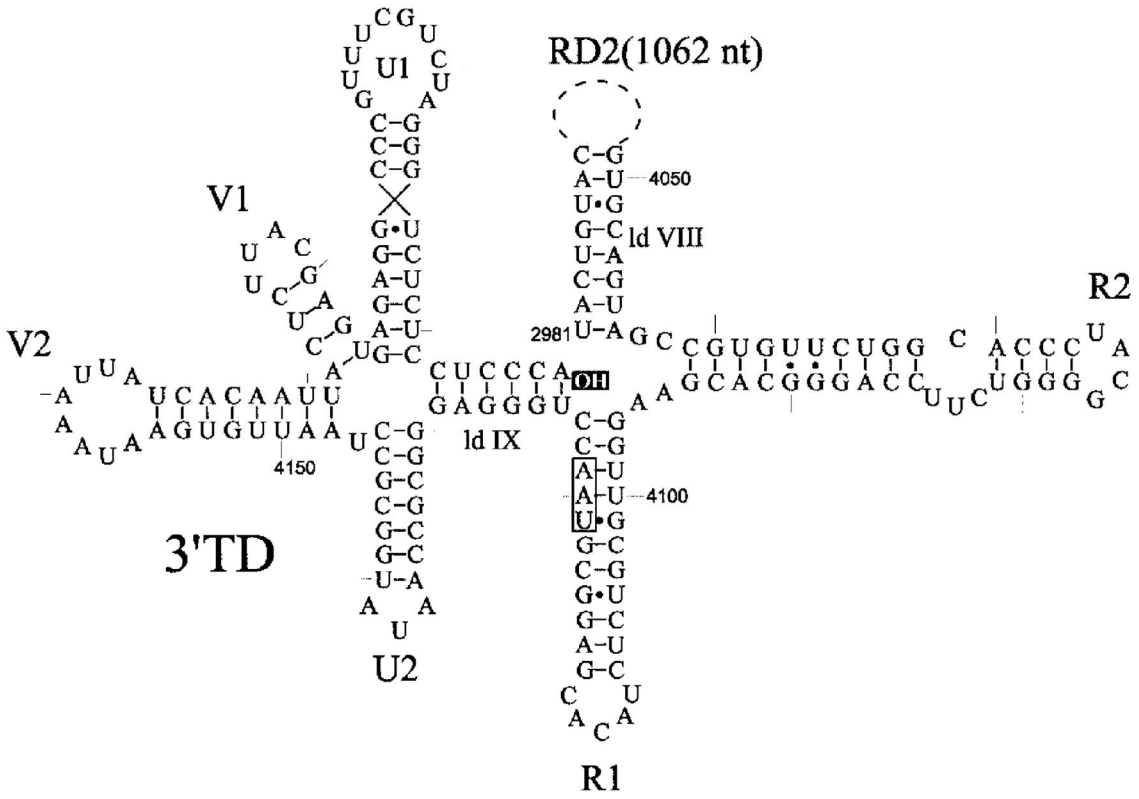

Figure III-4. RNA secondary structure of the 3' end of Qβ
The 3' RNA secondary structure. Loops are labeled according to the convention established by Beekwilder (Beekwilder, Nieuwenhuizen et al. 1995). From the center, going clockwise, the start of the sequence at nucleotide 2981 is within the replicase, with ld VIII standing for long-distance interaction 8. RD2, is replicase domain 2. R2 and R1 are the last stem-loops within the replicase, the stop codon is boxed. Then there is long-distance interaction 9, which interacts with the terminal 6 nucleotides. U2, V2, V1, and U1 are untranslated. Figure is adapted with permission from Klovins and van Duin (Klovins and van Duin 1999).

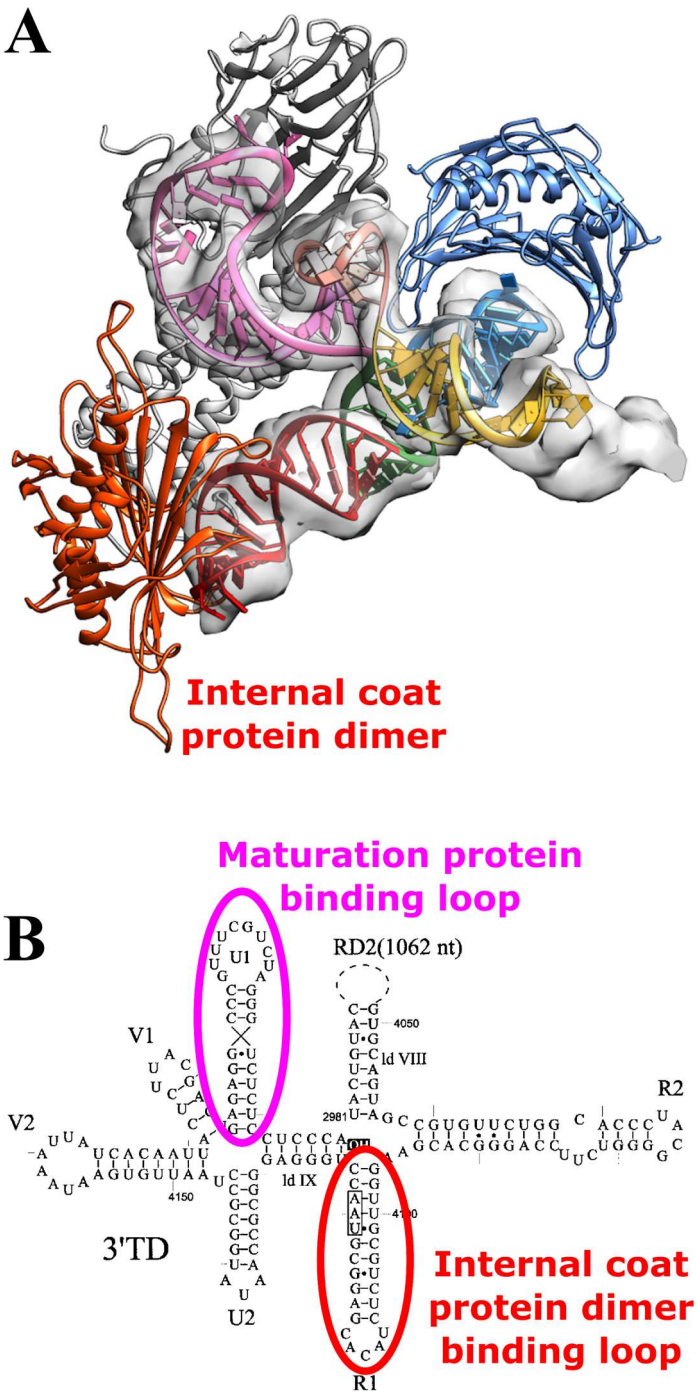

Figure III-5. Electron density showing of the 3' end of the Qβ gRNA where it interacts with the maturation protein and internal coat protein dimer
Panel A is courtesy of Jeng-Yih Chang, who modeled in the entire genome of Qβ into the electron density inside the capsid. The 3' end is in proximity to the maturation protein and the internal coat protein dimer. The maturation protein is

shown as a grayed ribbon model of the protein backbone. The loop with which the maturation protein interacts is colored in magenta. A coat protein dimer which is part of the capsid is shown as a blue ribbon model, with the gRNA stem-loop that it interacts with (U2) is also colored in blue. The gRNA stem-loop R1 is colored in red, as is a ribbon model of the internal coat protein dimer. The remaining colored RNA is as follows: tan is V1; green is ld IX (long-distance interaction 9); and yellow is V2. Panel B is adapted with permission from Klovins and van Duin (Klovins and van Duin 1999).

Jeng-Yih also identified the stem-loop which comes in contact with the internal coat protein dimer, designated R1 by Beekwilder and colleagues (Beekwilder et al., 199?). R1 is termed as such because it is the last stem loop in the 3' end of the gene for the replicase (Beekwilder et al., 199?). The stem-loop numbers increase going toward the 5' position. In order to validate the gRNA model for both the maturation protein binding stem-loop and the internal coat protein binding stem-loop, mutations were made. As the R1 loop is within the coding sequence of the replicase, one could argue that a knockout of the loop, eliminating the terminal seven amino acids, would alter the function of the replicase. The terminal seven amino acids of the Qβ replicase are not conserved in MS2 or other ssRNA phages (Kidmose et al., 2010), and have never had a lethal mutation found within them, with lethal defined as abolishing replicase activity. Therefore, two classes of mutants were made for U1 and R1, either complete knockouts of the loops, or alterations of the loops such that the secondary structure was different than predicted by Beekwilder (Figure III-6). U1 is within the 3' untranslated region, but even a modest alteration might affect infectivity.

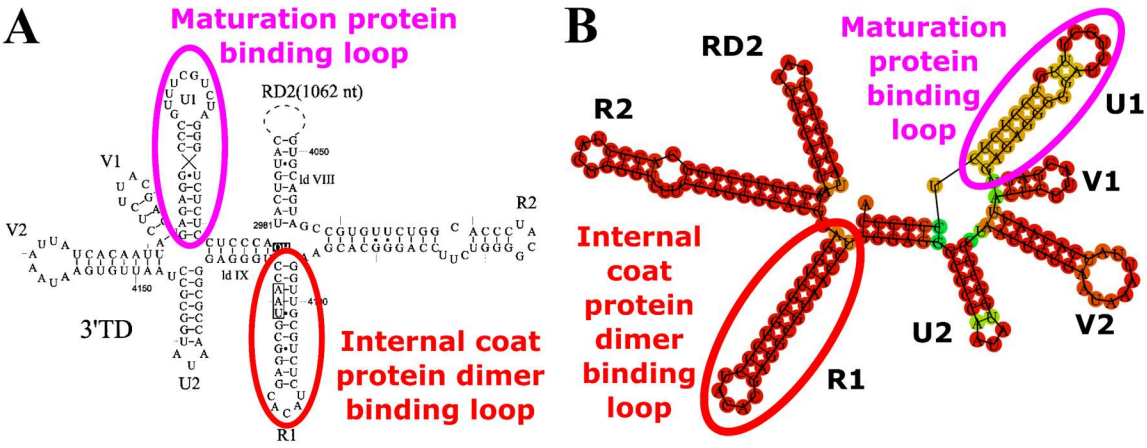

Figure III-6. Secondary structure of the 3' end of the Qβ genome
Panel A is as in Figure III-5, shown side-by-side with Panel B to readjust the reader's eyes. Panel B, and subsequent figures, are made with RNAfold, a tool for RNA secondary structure prediction. Similar software was used in Panel A, but the authors also used a phylogenetic comparison with related phages to determine the gRNA secondary structure (Beekwilder, Nieuwenhuizen et al. 1995, Klovins and van Duin 1999). The maturation protein and internal coat protein dimer binding loops are circled for both, and colored as in Figure III-5. Panel A is adapted with permission from Klovins and van Duin (Klovins and van Duin 1999).

Due to the high mutation rate of ssRNA phages, and thus their ability to deleterious overcome mutations within a relatively short amount of time (Olsthoorn and van Duin 1996), these mutations were made in a cDNA plasmid within a non-permissive host, *E. coli* DH10B (F^-). The choice of plasmid was thought to be important, with the desire for a higher titer producing plasmid in order to see if there was even a relatively minor effect. Plasmids were requested from all over the country and the world, with many laboratories agreeing to share materials. Titers from overnight cultures were determined (Table III-2). As an interesting side-note, Qβ was the first organism to be put completely on a plasmid (Taniguchi, Palmieri et al. 1978).

Table III-2. Qβ cDNA plasmids obtained
Various Qβ cDNA plasmids were obtained from different sources in order to modify the genome.

Plasmid:	Titer:	Source:	Notes:
pQBm100	~10^{10}-10^{11} PFU/ml	Young lab, Texas A&M	Stored in ER2738 (an F^+ $RecA^+$ strain), sequence unstable with genome repeats. No suitable for cloning, great for purifications. Produces WT phage.
pSKQB	~10^{6}-10^{7} PFU/ml	Nori Ichihashi Osaka University	Used for initial RNA delivery experiments (Chapter IV).
pQB7	~10^{7}-10^{8} PFU/ml	Charlotte Knudsen University of Åarhus	Hard to transform.
pQB7	~10^{7}-10^{8} PFU/ml	Alain Waffo Alabama State University	Hard to transform.
pQB8	~10^{7}-10^{8} PFU/ml	Alain Waffo Alabama State University	Negative sense. Unstable, hard to transform.
pQBm100	~10^{8}-10^{9} PFU/ml	René Olsthoorn Leiden University	Not sequenced.
pBR322QB	~10^{8} PFU/ml	David Mills by way of Carl Dobkin NY Staten Island Institute	Not sequenced.
pBRT7QB	~10^{9}-10^{10} PFU/ml	Sean Leonard University of Texas, Austin	Sequenced, stable. Originally from Weber, who worked with C. Weissman.

After trying out eight plasmids, the decision was made to work with pBRT7QB, from Sean Leonard at the University of Texas, Austin. This plasmid produced approximately the highest titer of phage, with the advantage of being fully sequenced (Table III-2). Mutations were made via Quikchange, and confirmed via sequencing of the 3' region. Both removing R1 and altering the stem-loop of R1 to change the

secondary structure (Figure III-7), while keeping the coding sequence the same, resulted in the phage being unable to infect. It could be argued that the removal of R1 caused the replicase to be inactive, but this would not be the case for the mutant which still codes for the intact replicase.

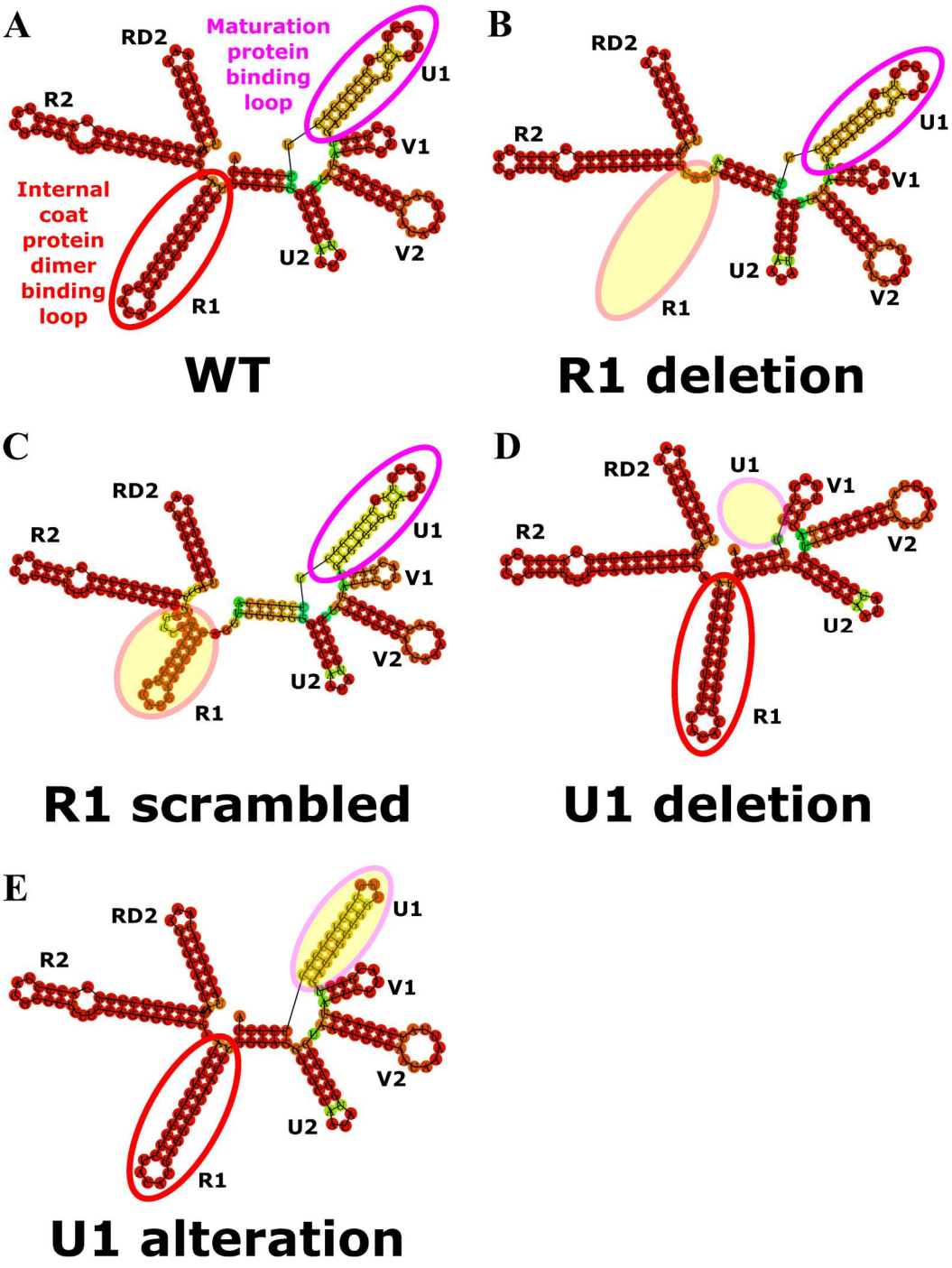

Figure III-7. RNA secondary structure of 3' mutants
The secondary structure of the 3' end of the wild-type (WT) gRNA is shown in Panel A. The domains are marked as per Beekwilder's notation (Beekwilder, Nieuwenhuizen et al. 1995). The maturation protein and internal coat protein binding loops are circled in pink and red, respectively. B) The secondary

structure of the same region is shown, after R1 was deleted to enable a UAA stop codon. C) R1 was mutated such that it maintained coding for the amino acids within the C-terminal end of the replicase, while disrupting the secondary structure of R1. D) A deletion of U1, the maturation protein binding loop. E) an alteration of U1 such that the loop was only 3nt. All mutations are highlighted in yellow.

As for the U1 mutants, the U1 alteration was slightly different than the R1 alteration, in that U1 does not code for any proteins, so the loop at the end of the RNA stem was modified to change it from a predicted six nucleotide loop to a three nucleotide loop to see if the maturation protein binding was specific for the loop or the stem. The loop as also deleted, to see if a more severe mutation would disrupt binding (Figure III-7).

Surprisingly, mutations within R1 and U1 were lethal. Instead of a drop in titer, the titer was abolished (Figure III-8). This was immediately surprising. Based on our hypotheses, the loss of infectivity due to the U1 mutations are understandable. If the maturation protein is unable to bind to the gRNA in the right place there are several scenarios that could prevent infection: the maturation protein might not be surface exposed after capsid assembly; the maturation protein might not be bound to the gRNA strong enough to be able to get the gRNA pulled in during the process of infection; the gRNA might not make it out of the capsid and into the cell before RNases cleave some or part of it; or the maturation protein might not even assemble into the capsid, let alone bind the gRNA if it does.

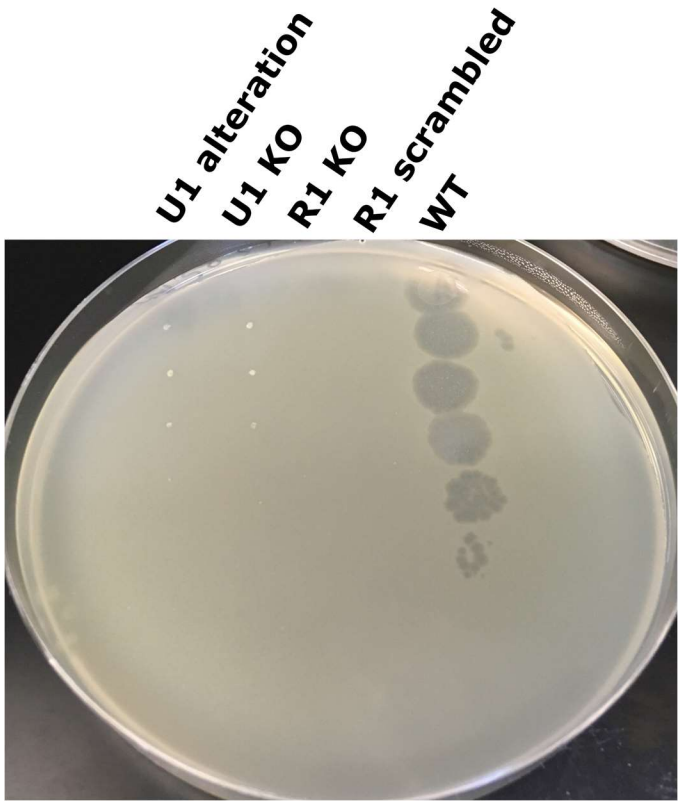

Figure III-8. Alterations of the RNA secondary structure abolish infectivity
Plaque assay of serial dilutions from an overnight culture (grown in F^- cells) to determine if the mutants were able to make infectious particles.

As far as the R1 mutations go, the R1 deletion could have had an effect on the replicase activity. Mutations in amino acids encoded by R1 have never been described as lethal (Kidmose, Vasiliev et al. 2010), so this should not have abolished plaquing altogether. As both the R1 knockout and the R1 alteration both abolished plaquing, the stem-loop itself might have an outsized role, potentially in packaging. R1 and U1 are proximal to each other, and the maturation protein binds one while the 'internal' coat protein dimer binds the other, the two stem-loops could require each other for correct folding. I hypothesize that mutations in either U1 or R1 will prevent the gRNA from

being in a conformation such that the maturation protein can bind correctly during assembly of the virus.

The viruses are clearly not infectious, but are they made? The mutations were made on a cDNA plasmid containing Qβ, which normally yields about 10^9-10^{10} PFU/ml just from leaky expression after growing overnight. To eliminate the possibility that there was a secondary mutation, which prevented some component of the virus to be made, the entire genomes for the R1 and U1 mutants were sequenced. There were no secondary site mutations. The mutants should still have produced normal titers unless the specific mutations had an effect. Incidentally, when screening new mutants isolated after transforming the Quikchange, there were plenty of other mutations which had no effect on the ability to plaque. Sequenced mutations occurred in the regions around R1 and U1, with either single base changes or large insertions or deletions. As these mutants were able to plaque this rules out the possibility that secondary mutations also negatively affect the ability to form infectious phages. Most of the mutations can probably be attributed to reduced fidelity of the DNA polymerase under high DMSO concentrations.

In order to determine why the mutants were no longer infectious it is important to see if they still produce the maturation protein. The maturation protein binds to the gRNA as well as the host receptor. If the maturation protein is no longer produced, then the phage cannot bind the cell. In order to do this, overnight cultures were spun down

(as described in the Materials and Methods), the supernatants TCA precipitated, run on a SDS-PAGE, then transferred to a membrane and immunoblotted with an antibody against A_2 (Figure III-9). The rationale for determining whether there was A_2 in the supernatants of cultures was that if A_2 is expressed the cells will lyse eventually, even if A_2 is not assembled into an infectious virion. The cultures were grown in the same method as was used to determine if the mutants titered. The supernatant of DH10B (F^-), which was the strain carrying the plasmid, was also used to ensure that the antibody against the maturation protein was specific.

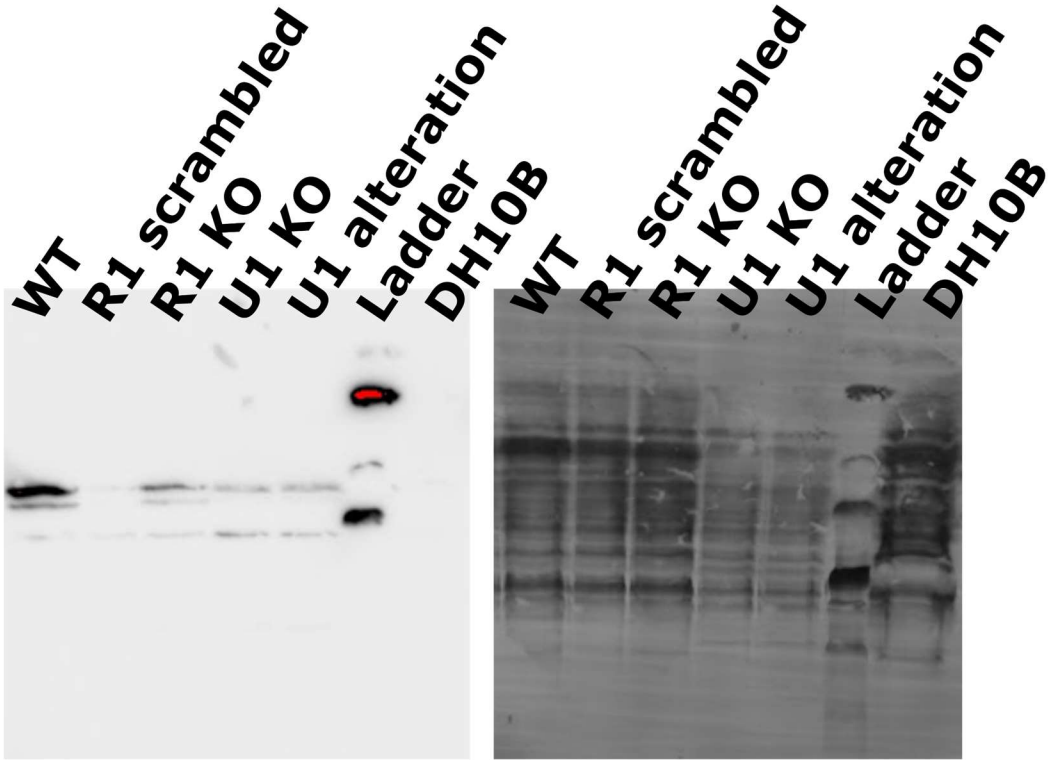

Figure III-9. Western blot of whole cell lysates to show MP production
The strains carrying the Qβ cDNA plasmid with wild-type, mutations, or no plasmid, were grown at overnight at 37°C then diluted to OD_{600} = 0.5, the OD measured again, the cultures spun down, then TCA precipitated, and loaded onto

an SDS-PAGE. The gels were transferred to a membrane and blotted with an antibody against A_2. The membrane was then stained, to show total protein loaded and transferred. DH10B is a negative control, the *E. coli* (F^-) strain which the others were transformed into.

The maturation protein was detected in wild-type, and each of the mutants, but not in the host strain alone. That the maturation protein is seen at different levels is merely due to the differences in TCA precipitation and loading. There is a very little amount of maturation protein in each sample, and the result is repeatable but with levels differing slightly from experiment to experiment. The maturation protein is present in just one copy per capsid, versus 180 copies of the coat protein for Qβ (178 for most other ssRNA phages). The phages are also only present at a level of ~10^9 – 10^{10} PFU/ml from an overnight culture of bacteria, which is about OD_{600} = ~5, or ~10^9 cells/ml. The majority of proteins in the supernatant would be from *E. coli,* and thus even a tiny change in the total protein loaded can affect the levels of maturation protein seen by Western blot. However, that the experiment is reproducible shows that the maturation protein is being made across the samples.

The top band (across the samples) in the Western blot is presumably the intact maturation protein, with the other lower bands being degraded maturation protein. This can happen in a variety of ways, but is probably attributable to the fact that the maturation protein did not evolve to be stable by itself, rather it evolved to be inserted into the capsid, and thus protected from proteases. The Western blot of the supernatants was total protein, which means that the maturation protein detected could be either free-

floating or within the virion, either surface-exposed or sequestered within the capsid. To determine if the maturation protein produced in the mutants and wild-type was actually assembled into the particle, the virions were purified with a CsCl density gradient and analyzed via Western blot (Figure III-10).

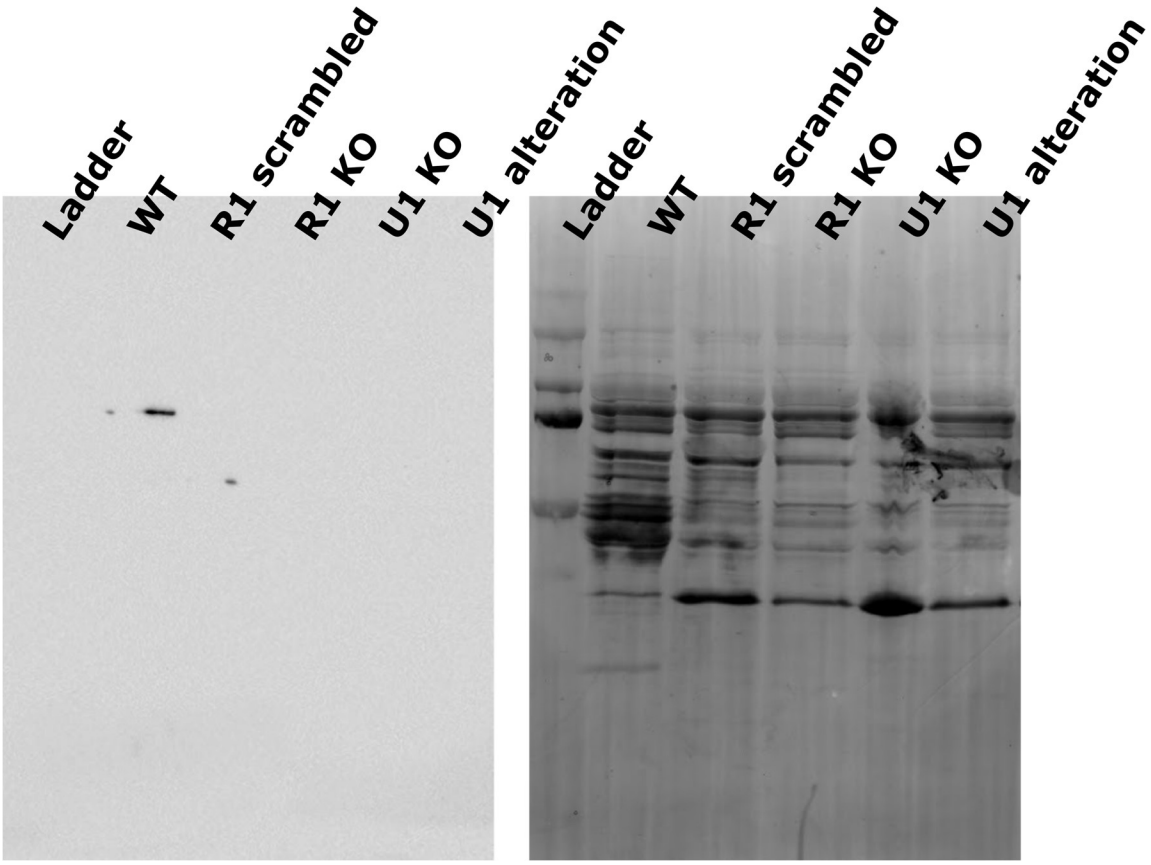

Figure III-10. Western blot of CsCl purified Qβ and mutants
Western blot of CsCl purified samples (left). The membrane was stained after performing the Western blot (right). It should be noted that the phages were produced off of a plasmid, rather than from an infection, and thus the titer is lower, even for wild-type.

The maturation protein was only seen in wild-type Qβ, there may be some degraded maturation protein within the R1 scrambled mutant (the mutant which had the

coding sequence of the replicase intact but the nucleotides altered such that the gRNA secondary structure was altered). The protein bands seen in the stained membrane can be from soluble proteins from throughout the CsCl gradient, or could be from ribosomes being purified with the phage capsid band. Ribosomes are about the same size as ssRNA phages and also have a large protein component. That there is only intact maturation protein in the wild-type band indicates that the maturation protein is not able to successfully incorporate into either of the mutant types.

It is easy to understand how mutating the RNA stem-loop to which the maturation protein binds could abolish infectivity, as without the maturation protein there can be no host binding. How is it that a mutation in the end of the replicase could abolish maturation protein binding? For that we must think about how the gRNA is folded and presented such that the maturation protein can bind. The biggest class of Qβ particles, when doing a 3D classification, have a sequestered coat protein dimer binding R1. This coat protein dimer must somehow stabilize the 3' end of the gRNA such that it is in a conformation to which the maturation protein can bind. There is little evidence, and less agreement, on how capsid assembly takes place. Does that maturation protein bind first, or do coat proteins? At which point in time does the maturation protein bind?

What this work shows is that the maturation protein is unable to bind without an RNA stem-loop which has the sequestered coat protein dimer. It does not indicate order of binding but could be interpreted as the maturation protein binds after the internal coat

protein. The fact that Qβ needs the internal coat protein dimer fits in with some of our previous results, inasmuch as the gRNA of Qβ is not universally in one conformation. After we collected more micrographs of Qβ, Zhicheng Cui did a focused classification around the maturation protein and the gRNA of Qβ and found that only ~30% of particles had the maturation protein (Figure III-11). These particles all had the extra protein density belonging to a coat protein dimer. The other ~70% of particles still had electron density within the capsid for RNA, but it was not able to be classified into groups.

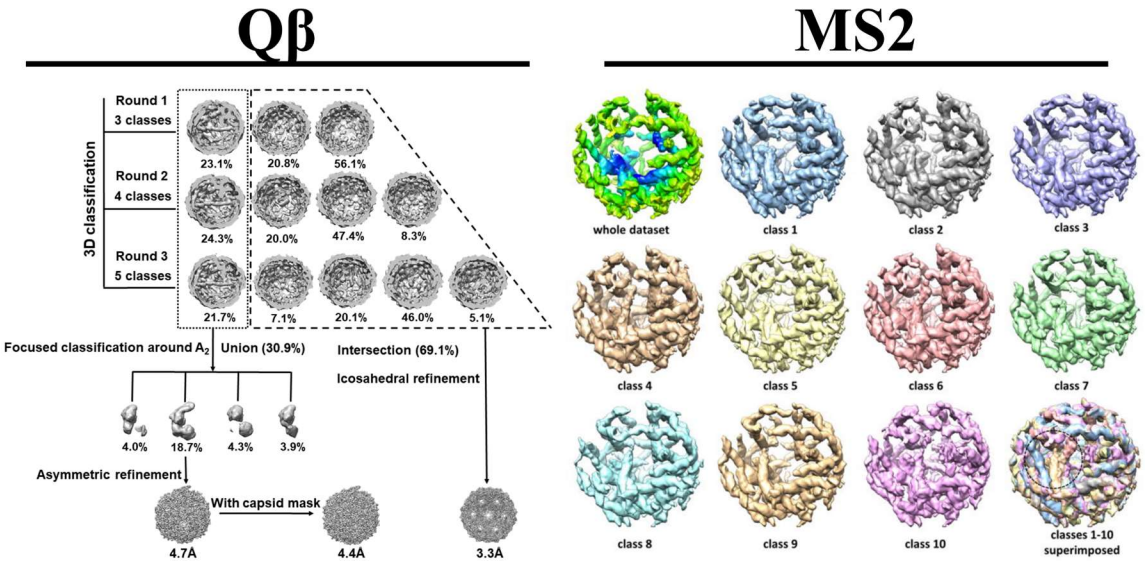

Figure III-11. Electron density for Qβ and MS2 reveal differences in gRNA
The gRNA of Qβ is only classified into the dominant conformation 30% of the time, whereas the gRNA of MS2 is in the dominant conformation >95% of the time, with only slight variations. Left panel (Qβ) made by Zhicheng Cui for the 2017 *PNAS*, right panel (MS2) adapted from Dai et al., 2017, with permission.

That only 30% of particles had the gRNA in a confirmation which enabled classification is probably due to random packaging of the RNA. If the gRNA is folded

in such a way that the maturation protein is unable to bind, then the virus could form a perfect icosahedral particle, with 90 coat protein dimers in the capsid. Even with a full-length gRNA that is in one conformation, this would hinder 3D classification, as there would not be a break in the symmetry of the capsid to anchor the gRNA. As it is, the gRNA of Qβ is complex, as there are several reported long-distance interactions. We do not know if the internal coat protein dimer is required for these long-distance interactions to occur. In MS2 there is no internal coat protein dimer, as well as no long-distance interactions, and thus the folding of the gRNA is much simpler. The authors who solved the high-resolution structure of MS2 stated that >95% of MS2 particles had gRNA which fell into one of 10 classes, all closely related (Figure III-11, right panel, notice the color variation in the top left panel within the MS2 dataset to see the variations in the whole dataset).

Part of the difference in gRNA folding, and the lack of an internal coat protein dimer in MS2, can be attributed to the sizes of the UTRs in the two viruses: 5' UTR of Qβ is 59nt, with the 3' UTR is 93nt; for MS2 the 5' UTR is 129nt and the 3' UTR is 171nt. These differences are magnified if you compare them in terms of genome size, Qβ has UTRs of 152nt out of a 4,217nt genome (3.6%), whereas MS2 devotes 300nt out of a 3,569nt genome (8.4%). There has not been a consensus on why MS2 devotes more genome space to untranslated regions, but folding of the gRNA into a conformation that can successfully pack 95% of the time versus 30% of the time could be a good hypothesis.

The bulk of the difference between the translated regions of Qβ and MS2 is the extended coat protein A_1, which is caused by a readthrough of the stop codon of the coat protein in Qβ. The maturation protein, regular coat protein, and replicase of Qβ are only slightly bigger than their counterparts in MS2 (421 vs. 394, 134 vs. 131, and 590 vs. 546, each is the amino acid difference for Qβ vs. MS2 for the MP, coat, and replicase, respectively). The extended coat protein is evolutionarily conserved in *Allolevivirus,* and is required for infection. The extended coat protein is 330 amino acids, 2.5 times the size of the coat protein. The N-terminal end of A_1 is that of the coat protein, so A_1 could theoretically be located anywhere within the capsid if it forms heterodimers with the coat protein. It is not known whether the coat protein can form heterodimers with A_1, logically it should, but A_1 might only be able to form a homodimer. There is one area within our reconstruction which could possibly have been A_1, at the base of the maturation protein there was electron density outside of the capsid (Figure III-12).

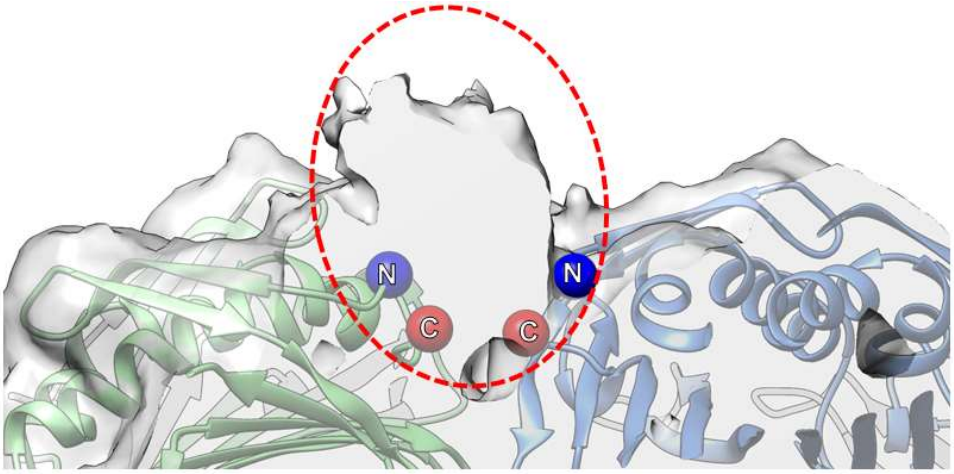

Figure III-12. Extra electron density seen in the asymmetric Qβ structure
The electron density is shaded in grey. Ribbon models of the coat proteins fit into the electron density are colored, the N- and C-termini of each coat protein are also colored. The extra electron density is circled.

That there was electron density seen here is probably due to A_1 being stabilized by the maturation protein. If A_1 was anywhere else within the capsid, with nothing to stabilize it, the electron density would be averaged out during refinement. While A_1 is required for infection, it might affect assembly of the capsid. Two mutants of A_1 were made: one where the leaky stop codon was replaced with a better stop codon, resulting in a 'coat only' mutant; while the other had the coat protein stop codon replaced with a tryptophan (what the leaky stop codon is read as in A_1), with a corresponding A_1 only mutant. Neither of these phages were infectious. This confirms biochemical experiments done in the early 1970s, in which the components of Qβ were isolated from each other and added back in with individual elements missing, to see what was required for infection (Hofstetter, Monstein et al. 1974).

CHAPTER IV

ALTERNATIVE CAPSID ASSEMBLIES ENABLE USE OF SSRNA PHAGE COAT PROTEINS TO DELIVER FOREIGN RNAS

4.1. Introduction

There is an extensive literature the purification of capsids based on the coat proteins of ssRNA phages. However, this is always done from a starting point where the coat protein is overexpressed, either from an infection or from overexpression off of a plasmid (Figure IV-1). The challenge is to purify the phage from similarly sized macromolecules, when it is not overexpressed. The ssRNA phages for *E. coli* cannot be overexpressed, at least not in their entirety, as they include lysis proteins which will kill *E. coli*. The only way to have them on a cDNA plasmid is to have either zero or very low levels of expression. If there is any expression, the replicase will amplify whatever RNA is made, to the point where it would appear to be an infection. Eventually the phage will lyse the cell. There is the potential that growth at lower temperatures would allow the phage to be induced, but this has never been explored. All that is known is that Qβ does not grow as well at lower temperatures, or higher temperatures. In any event, it is difficult to purify non-infectious mutants, such as the R1 and U1 mutants which have a functional lysis protein but cannot infect.

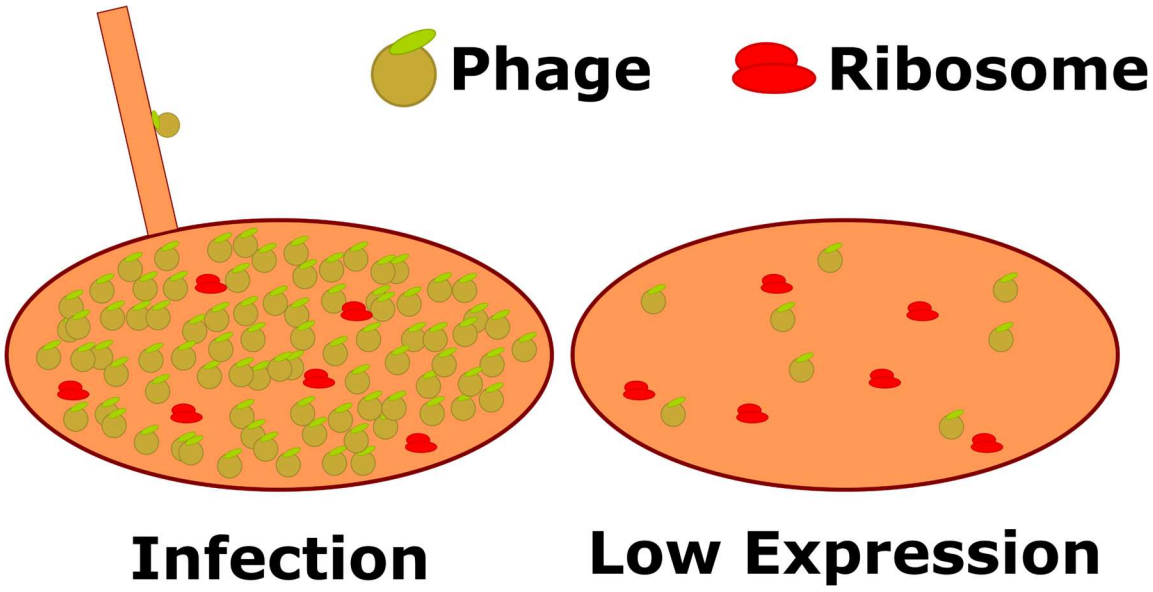

Figure IV-1. Differences in phage number during an infection versus low level expression
The ribosomes are the only similarly sized macromolecule in lab strains of *E. coli* used for infection. When there are low levels of expression the phage might not even outnumber the ribosomes. Purifying phage away from ribosomes is simpler under expression levels produced during an infection, as any contaminating ribosomes are at much lower levels than the phage. At low level expression the ribosomes might be at a similar level to the phage, so any contamination in the phage purification is much more dramatic.

4.2. Materials and Methods

4.2.1. Bacterial Strains and Growth Conditions

The new purification for wild-type Qβ, and the R1 and U1 3' mutants, had the cDNA plasmids grown in *E. coli* DH10B. Wild-type Qβ had a titer of ~10^8 PFU/ml from an overnight of the plasmid in DH10B, ~1,000x less than an infection. BL21(DE3) had been tried, but the cells were lysed by A_2 with even leaky induction from the T7 promoter. Using BL21(DE3) pLysS and BL21(DE3) with pZA32_MurAA (from *B. subtilis*) did not increase the yields of wild-type, and so while they were tried with the

mutants, the final purification steps did not use these cell lines. The pZA32_MurAA was a kind gift of Karthik Chamakura, from the Young lab. Appropriate antibiotics were used as in Chapter III. The mutants (and wild-type) were let to grow overnight at 37°C with shaking, as in Chapter III. For induction and purification of the Qβ-VLPs with either eGFP or kanamycin RNA inserted into the genome, plasmids were initially cloned into DH10B, then were produced in BL21(DE3) with pLysS, and coat/A_1 on pET28.

4.2.2. Purification Methods

The final purification of alternative capsids is as follows: single colonies of *E. coli* DH10B with pBRT7QB, either wild-type or any of the mutants, were picked for overnight growth at 37°C with aeration of 1:4 and appropriate antibiotics (ampicillin for wild-type and the 3' mutants; ampicillin, kanamycin, and chloramphenicol for the VLPs to deliver eGFP or kanamycin RNA which were grown in BL21). The cultures were inoculated 1:100 into 500ml, again with aeration of 1:4, then grown overnight at 37°C with shaking at 200r.p.m. Cells were spun down at 7,000 x *g* for 30min. The supernatant saved and stored at 4°C. The cell pellets were either frozen or lysed right away. When lysing, the cell pellets were resuspended in 125mM NaH_2PO_4 pH 8.0, with RNase and DNase added to 10U/ml. Cell pellets were typically resuspended in buffer at a level of 20ml/L of growth media. Cells were lysed in a French Press at 25,000 p.s.i., multiple times, then cell debris was removed by spinning at 30,000 x *g* for 30min. Soluble cell lysates were then added to the supernatants, at 4°C.

Supernatants were then adjusted to 500mM NaCl, with 5M NaCl, and had powdered PEG6000 added to 10% w/v. This was further incubated at 4°C for at least an hour, then spun down in a 4°C prechilled centrifuge at 8,000 x *g* for 30min. The supernatant decanted and the pellet resuspended in ~10ml 125mM NaH_2PO_4 pH 8.0 per liter. This sample had additional RNase and DNase added to 10U/ml and was incubated at 37°C for at least two hours. The samples were then chloroform extracted to remove the PEG. The samples had equal volumes of chloroform added, were vortexed, then spun down at 40,000 x *g* for 30min. The aqueous layer (top) was removed and chloroform extracted again, until the PEG interface between the chloroform and aqueous layers was removed, about five chloroform extractions. The chloroform extractions were necessary, because the PEG associates with the particles and alters the separation on size exclusion chromatography.

After removing the PEG and chloroform, the samples were concentrated using a 10kDa MWCO centrifugal concentrator, until the volume was ~3-4ml. This was to ensure that the amount loaded on the S500 column was less than $1/100^{th}$ of the total bed volume. The sample was filtered through a 0.22µm filter, then loaded onto the S500 column. The buffer used was the same as what the samples were resuspended in, 125mM NaH_2PO_4 pH 8.0. The fractions which contained infectious phage (for wild-type) were concentrated using the same MWCO centrifugal filters. The same fractions were used for the 3' mutants.

After establishing a purification protocol for low yield particles, the protocol was used to purify high-yields of particles from an infection of wild-type Qβ (~10^{11} pfu/ml, or 1,000x more than off of the cDNA plasmid). This was to see if the purification method altered the particles purified. The expression level did not alter the amount of differentially sized particles.

4.2.3. Cloning for the GFP and Kanamycin Resistance VLPs

Standard molecular biology techniques were used to clone eGFP and the kanamycin resistance gene into the Qβ cDNA plasmids, initially into pSKQB then into pBRT7QB, as pBRT7QB gave a 100x greater yield for wild-type Qβ. The sequence of eGFP was amplified from peGFP-N1, a kind gift of Baoyu Zhao of the Pingwei Li lab. The kanamycin resistance sequence was amplified off of pET28. The pSKQB and pBRT7QB plasmids were digested with AflII, which cuts inside of A_1, to linearize them, then were amplified with new restriction sites on either end of A_1. Primers used to amplify the plasmids were interchangeable between pSKQB and pBRT7QB, and are listed in Table IV-1.

Table IV-1. Primers used for modifying Qβ cDNA to make a delivery VLP

#146	eGFP with SphI cutsite SerLeu F.	ATTTGAGCATGCTTAAGGGCGAGGAGCTGTTCACC
#147	eGFP with SphI cutsite SerLeu R.	TTACTACCATGGACAGCTCGTCCATGCCGAGAGTGATCC
#148	pSKQB NcoI cutsite F.	AGCTGTCCATGGTAGTAACTAAGGATGAAATGCATGTCTAAGACAGC
#149	pSKQB SphI cutsite R.	CTTAAGGCATGCTCAAATTGACCCAAAGTTTCAACGC
#202	pSKQB linear Kan F.	ACTCTTCCTTTTTCAATATTATTGAAGC
#203	pSKQB linear Kan R.	GTAACTGTCAGACCAAGTTTACTC
#204	Kan cassette F.	Tcctttgatcttttctacgggg
#205	Kan cassette R.	acttttcggggaaatgtgc

4.3. Results and Discussion

In order to purify the mutant phages, those with mutations in either R1 or U1 which left the phage non-infectious, a multitude of approaches were tried. The most used method to purify ssRNA bacteriophages was not effective, as seen by the background protein bands after a purification. This method, concentrating the phages via ammonium sulfate precipitation, then separating the viruses with a cesium chloride density gradient, does not separate out the phage sufficiently from ribosomes. Ribosomes and ssRNA phages are very close in density, as they have roughly the same amount of RNA while ssRNA phages have a slightly larger amount of protein (Kurland 1960). When the ssRNA phages outnumber ribosomes 10:1 or more, as they do after a normal infection, the phages are easy to purify. At least when measuring purity via a protein gel, since the ribosomal proteins will be outnumbered by the coat proteins by multiple orders of magnitude, as there are ~180 copies of the coat protein per phage while most ribosomal proteins are present at only one copy per. When trying to purify

the phage at a lower level, say equal numbers of phage to ribosomes or less, then the ribosomal proteins will show up on protein gels.

Without being able to infect the R1 and U1 mutants were not able to express the replicase and overexpress coat proteins as would be normal. As a result, the phage were only able to be produced off of a plasmid, a non-native case for production of virus particles. Even wild-type Qβ is hard to purify when expressed off of a plasmid and it cannot infect other hosts (as seen in the stained membranes of Western blots from cesium chloride purified bands, Figure III-10). In order to purify the R1 and U1 mutants away from ribosomes, a new method of purification had to be established.

Several different purification methods were tried. Several patents have been issued for purifying ssRNA phage VLPs produced by overexpressing the coat proteins. The patents mostly used ion exchange chromatography, so these were the first purifications tried. Ion exchange chromatography with strong and weak anion exchangers (DEAE, Q-sepharose, and CIM-monolith AEX CIMultus) were tried with a variety of buffers to purify wild-type Qβ, without satisfactory results. The phage bound the columns but not tightly and were released gradually, not in a single peak. These tests were done with supernatants, dialyzed ammonium sulfate precipitations, and cesium chloride purified samples. Without being able to get a high titer out of an individual peak for wild-type, there was no hope for isolating mutants produced at a much lower

level. A new method of purification for ssRNA phage was required, in order to be able to purify the non-infectious mutants away from ribosomes.

The first step is to concentrate the phage (or later on phage mutants) from the supernatant into a quantity which is realistic to put onto an FPLC column (or a cesium chloride gradient, if possible). When MS2 was first being crystallized in the late 1970s, a variety of conditions were tested to concentrate the phage: different types of polyethylene glycol (PEG), different concentrations of PEG, different concentrations of ammonium sulfate, and using different temperatures (Min Jou, Raeymaekers et al. 1979). MS2 precipitates at 0.9M ammonium sulfate, roughly 120g/L (Min Jou, Raeymaekers et al. 1979), or less than half the amount typically used to precipitate ssRNA phages (280g/L). This experiment was repeated with Qβ. Qβ precipitates out at roughly the same level of ammonium sulfate as MS2 (using 25g intervals, from 0-350g/L). At 280g/L there are many more proteins precipitated, and thus higher levels of non-phage contamination. PEG precipitations, while used extensively for dsDNA phages, for some reason are not typically used for ssRNA phages. The two methods are roughly equivalent, with excellent precipitation of phage particles for both ammonium sulfate and PEG. The one problem with PEG is removing it, as the interaction with column matrices will be affected, either by binding or the size of the particles. PEG is relatively easy to remove in principle, just extracting with equal volumes of chloroform, but in practice a lot of the phage particles are often lost at the interface between the

organic and aqueous layers. That being said, PEG precipitation is preferable, as there are fewer contaminating proteins.

PEG precipitated wild-type phage was applied onto multiple columns, as before, with the best result being a size exclusion column. Size exclusion columns tried were a prepacked Superdex 200; a prepacked Superose 6; a handpacked 30cm x 1.5cm Superdex 300; a handpacked 100cm x 1.5cm Sephacryl 500; and a prepacked 26/60 Sephacryl 500-HR. The best results were from the S500 columns, with separation from smaller macromolecules (Figure IV-2). After screening with wild-type Qβ and doing plaque assays, coat/A_1 VLPs were used, as well as the R1 and U1 mutants, which are unable to plaque. To determine which fractions held the VLPs, gels were run. The same sample was loaded onto each column to eliminate the possibility of artifacts from different purifications. In the case of coat/A_1, which is shown in Figure IV-2 and Figure IV-3, the VLPs were overexpressed, cells spun down and lysed, the insoluble debris removed by spinning down at 18,000 x *g*, then the supernatant from the lysed cells added back into the supernatant from the culture. The whole sample was cooled to 4°C, with PEG6000 added to 10% and 5M NaCl added to a final concentration of 0.5M. The sample was spun down at 15,000 x *g*, then the supernatant decanted. The pellet was resuspended in 125mM NaH_2PO_4 pH 8.0, then extracted with an equal volume of chloroform. After each extraction the samples were vortexed then spun down at 45,000 x *g* for 30 minutes. The organic layer (the bottom) and the interface between the organic and aqueous layers (the PEG) were discarded and the aqueous layer removed for another

round of extractions. This was repeated for five chloroform extractions. The aqueous layer from the final extraction had RNase added to 10U/ml and was incubated at 37°C overnight. By incubating the samples in a low salt buffer without divalent cations, the contaminating ribosomes are less stable, thus more susceptible to RNases. Incubating the samples overnight with RNase, enabled degradation of the ribosomes which further facilitated separation based on size.

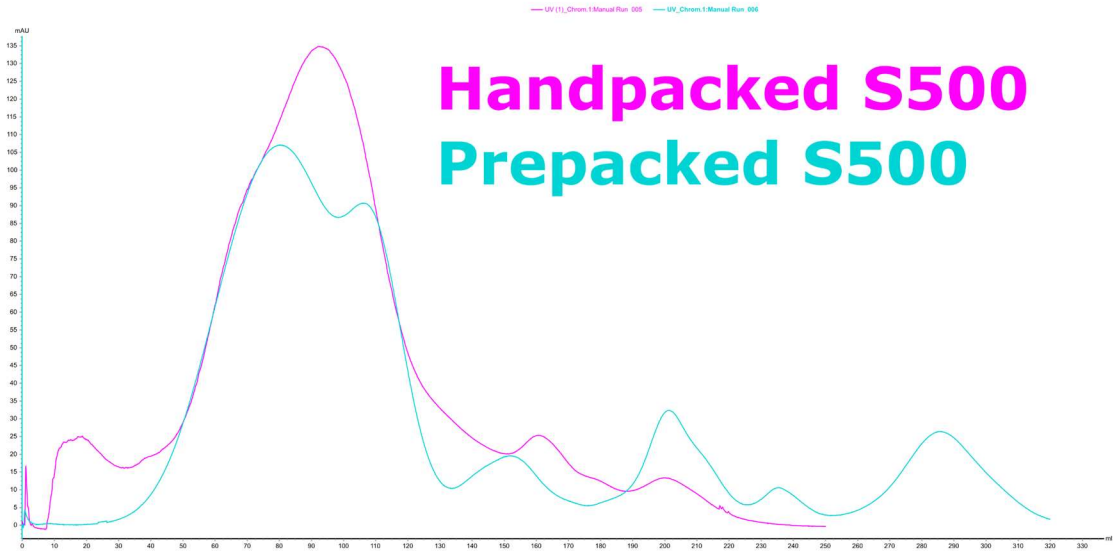

Figure IV-2. Purification of VLPs over S500
Purification of coat/A_1 VLPs overexpressed in BL21(DE3), purified via PEG precipitation, then loaded onto either a handpacked S500 column, or a prepacked S500. The sample was split and the same amount loaded onto each column.

An intact ribosome is roughly 25nm, slightly smaller than wild-type Qβ at 27nm. By degrading the ribosome into the large and small subunits, as well as cleaving exposed RNA with RNases, the samples were better able to be separated. In the case of the

prepacked S500 there were distinct peaks for the phage capsids and the ribosomes (compare lanes 2 and 3 with lanes 5 and 6 Figure IV-3).

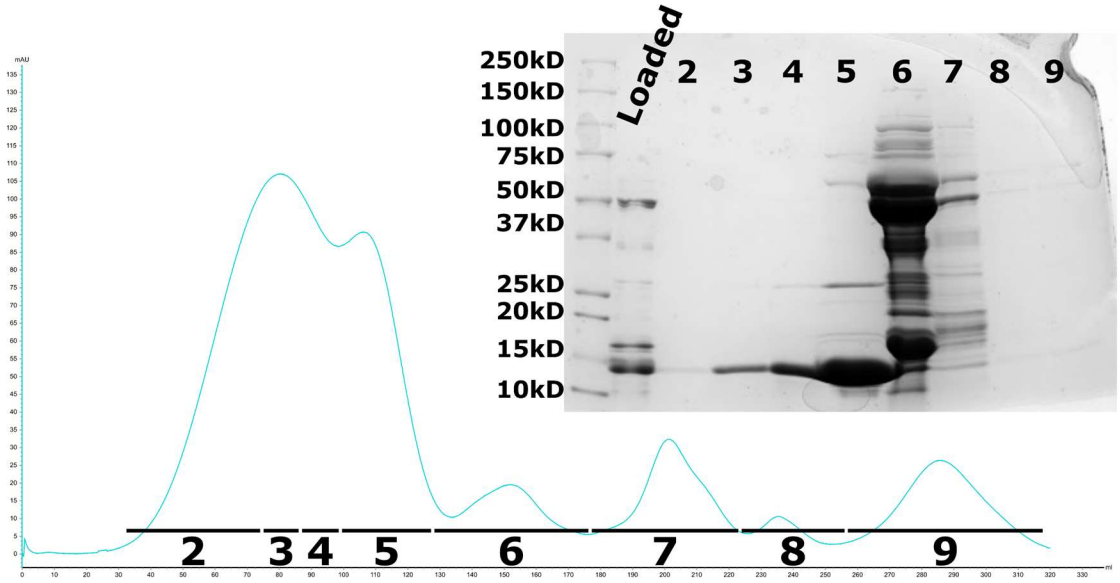

Figure IV-3. Coat/A_1 VLPs loaded onto an S500 to separate phage capsids from ribosomes
Virus-like particles from overexpressing coat/A_1 were semi-purified and concentrated using PEG precipitation, as described in the Materials and Methods, then loaded onto a prepacked S500. The X-axis indicates volume and the Y-axis is absorbance at 280nm, seen in arbitrary units.

After determining which fractions contained coat proteins, the samples were pooled and concentrated with a centrifugal concentrator. While the concentrators are labelled as 10kDa, 30kDa, 50kDa, etc… in reality about half of the titer of a sample can go through a 50kDa filter, even though Qβ is about 4MDa. Even a 30kDa filter will not stop all plaque forming units from going through (K. Chamakura, personal communication; and data not shown). This is even mentioned on the websites of certain companies, if you look hard enough. This only applies to centrifugal concentrators,

when buffer exchanging with dialysis tubing there is no centrifugal force applied, so the MWCO of the tubing more accurately reflects the true cutoff. The samples were concentrated from ~100ml to ~2ml with a 10kDa MWCO concentrator, then filtered and loaded onto an S200 column. Loading onto a S200 was to ensure that any proteins associated with the phage capsid but not integrated into it were separated away from it. The phages or VLPs come off in the void volume of the column, while smaller proteins came off later.

The purifications of R1 and U1 mutants were repeated enough times to show homogeneity on a protein gel, and relative purity when doing mass spectrometry of the samples (Larry Dangott and the Protein Chemistry lab at Texas A&M University, data not shown). The coat protein is the dominant protein in the purifications, which led to questions once the samples were imaged. This was because the particles produced by R1 and U1 mutants are non-homogenous *T*=3 capsids (Figure IV-4). There are many capsids which are smaller than the 27nm *T*=3 wild-type Qβ virion.

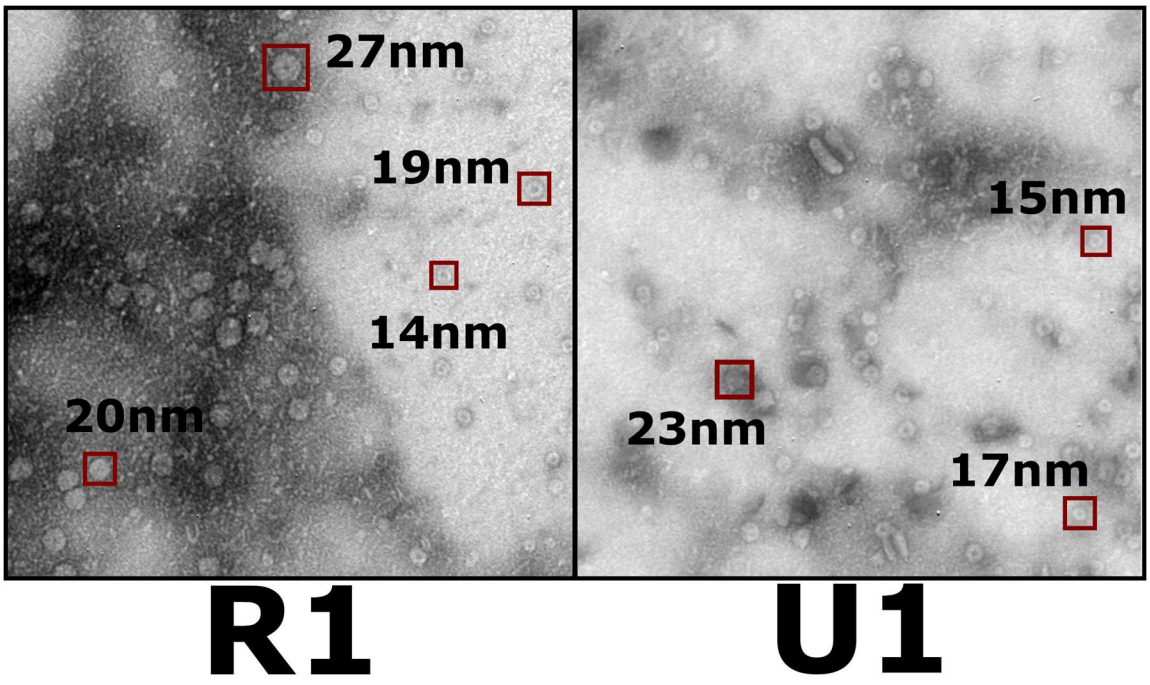

Figure IV-4. Negative stain of R1 and U1 mutants shows varying capsid sizes After developing a purification procedure to remove contaminating ribosomes, or other large macromolecular complexes, the R1 and U1 mutants showed varying capsid sizes. The R1 mutant purified for this experiment was the R1 loop scrambled to maintain the same amino acid sequence but have an alternative secondary structure. The U1 mutant was the U1 loop alteration, such that the stem loop had a predicted 3nt stem-loop, as opposed to the 6nt stem-loop.

What was the source of this homogeneity? The first scenario contemplated was that the particles lacked the maturation protein (as seen in Figure III-10). For thirty-five years it was assumed that the maturation protein (of MS2) bound two places on the gRNA, constricting movement and allowing the coat proteins time to bind and condense the gRNA to a size which fit into the capsid (Shiba and Suzuki 1981, Stockley, Ranson et al. 2013). This paradigm was thought to extend to all ssRNA phages, and helped establish theories on viral assembly (Toropova, Basnak et al. 2008, Twarock, Bingham et al. 2018). However, the high-resolution structure of MS2 showed that *in virio* the

maturation protein of MS2 only interacted with the gRNA in one location on the genome (Dai, Li et al. 2017). However, the structural studies of MS2 (and Qβ or any other ssRNA phage for that matter) all used density gradient purified phage. This biases the structural studies, inasmuch as during the purification process there is a selection for a certain protein to RNA ratio, with the whole procedure measured based on which fractions are infectious.

While there were smaller capsids for the R1 and U1 mutants, the virions were produced from much lower levels coat protein relative to an infection. Even for wild-type viruses produced off of the plasmid, the PFU/ml was at least an order of magnitude lower than that produced during an infection. To determine whether the lower concentration of coat proteins made a difference, or if the smaller capsids of the virions were just finally seen because the purification technique did not bias the results, the coat protein and A_1 were overexpressed off of a plasmid. The coat protein was cloned into both pET28a (under a T7 promoter for expression in BL21) and pZA12 (under a *lac* promoter, for expression in strains without the T7 polymerase). There were no discernable differences in expression between the two plasmids in BL21(DE3) or when comparing BL21(DE3) with ER2738 for pZA12 (not shown). After purification with PEG precipitation, extraction with chloroform, RNase treatment, the running over S500 and S200 size exclusion columns, there were non-homogenous capsids for coat/A_1, (Figure IV-5).

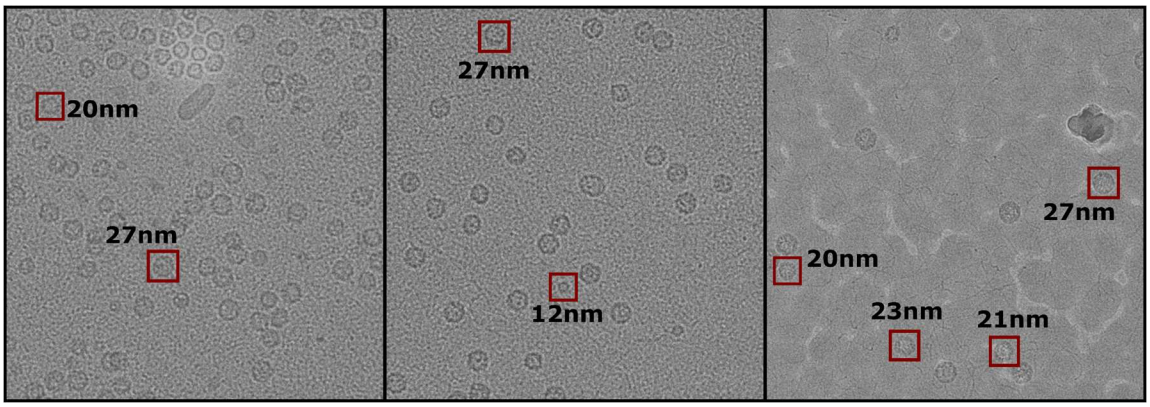

Figure IV-5. Coat/A_1 VLPs also show varying sizes when purified via gel filtration
When purified via gel filtration on an S500 then concentrated and loaded on an S200, the coat/A_1 VLPs show varying capsid sizes when examined with cryoEM.

By inadvertently removing the bias of selecting for infectivity and having to purify the capsids of a virus produced at much lower levels, away from similarly sized ribosomes which might be hidden if the phage were produced at a higher rate, the purification procedure for the R1 and U1 mutants might have revealed what is really happening in a cell during capsid assembly. The levels of coat proteins are not homogenous in the entirety of the cytoplasm, so there are local concentrations of coat proteins which are higher or lower, depending on any number of factors. This will almost certainly play a factor in assembly. The coat proteins of ssRNA phages have been shown to assemble on non-self RNAs (Witherell, Gott et al. 1991, Pickett and Peabody 1993), which would lead to different assembly pathways than around a wild-type RNA.

While the coat protein / A_1 sequence used was that of wild-type Qβ, and so should have the same local secondary structure as the native gRNA, the RNA is only a fraction of the whole genome, 990nt out of 4,217nt. To further validate the hypothesis that the coat proteins can assemble into alternative capsids, regardless of whether the virions produced are infectious, my lab mate, Jeng-Yih Chang went back through the old data which Zhicheng Cui and I had collected. What Jeng-Yih found was that there were alternative capsid assemblies in wild-type Qβ, even after cesium chloride purification (Table IV-2).

Table IV-2. Phages and VLPs formed from Qβ coat proteins form capsids other than *T*=3
Data collected by Karl Gorzelnik, Zhicheng Cui, and Jeng-Yih Chang, from samples purified by Karl Gorzelnik. Data processed by Jeng-Yih Chang.

	***T*=3**	***T*=4**	**Prolate**	**Oblate**	**Total**
WT Qβ (CsCl)	247,552 **(97.0%)**	1,201 **(0.5%)**	4,513 **(1.8%)**	1,868 **(0.7%)**	255,134
WT Qβ with MurA (CsCl)	150,798 **(98.4%)**	424 **(0.3%)**	1,441 **(0.9%)**	541 **(0.4%)**	153,195
WT Qβ (gel filtration)	51,507 **(76.1%)**	0	200 **(0.3%)**	15,954 **(23.6%)**	67,661
Coat/A_1 (gel filtration)	7,689 **(57.1%)**	0	53 **(0.4%)**	5,724 **(42.5%)**	13,466

There were abnormal particles found in cesium chloride purified Qβ, from datasets of just wild-type and those of the wild-type virus incubated with MurA (the lysis target of A_2). Anecdotally, there were many more smaller capsids in the R1 and U1 mutants than wild-type. Jeng-Yih examined the old datasets and found that there were

small particles for wild-type Qβ, just in an extremely low amount. Was this a result of a biased purification or the expression conditions for R1 and U1?

Since the R1 and U1 mutants are unable to infect, they are produced solely from the cDNA plasmid. With a functional A_2 protein encoded on the cDNA, the genome is unable to be overexpressed, as most any expression level will lyse the cells. Expression off of a cDNA plasmid will lead an RNA, of whatever length, which will eventually lead to a gRNA of the appropriate length. While an RNA produced from a leaky promoter might be too long to fit into a capsid, over time the host RNases will chew off ends which are not structured enough to refuse them, as viral RNAs have been evolved to. As the gRNA is also mRNA, the maturation protein, coat, and replicase will be expressed, with the replicase copying the gRNAs to the point that enough maturation protein is produced to lyse the cell.

In order to determine if the R1 and U1 mutants produced *T*=3 particles at a lower level, or if biased purifications led to the alternative particles just not being counted for wild-type Qβ, the new purification procedure was used for Qβ produced from an infection and coat/A_1 overexpressed from a plasmid (pET28 in BL21). When purifying Qβ and Qβ-VLPs using the size-based method, as opposed to the cesium chloride density purification, there were more smaller particles for both wild-type Qβ and the Qβ-VLPs (Table IV-2).

The alternative sized Qβ particles are not unprecedented. When the MS2 coat protein is overexpressed there were *T*=4 particles observed (de Martin Garrido, Crone et al. 2019). The same is true for the PP7 coat protein (Zhao, Kopylov et al. 2019), although the PP7 coat protein was modified to express foreign antigens. This is the first instance of alternative capsid assemblies resulting from an ssRNA phage infection. There are at least four types of particles produced by the coat proteins of Qβ. The dominant capsid morphology is the wild-type *T*=3, which has 180 copies of the coat protein with infectious virions having the symmetry disrupted by the maturation protein and an internal coat protein dimer, while virus-like particles have perfect symmetry. The *T*=4 capsids seen for the MS2 and PP7 coat proteins is also made by Qβ, these particles have 240 coat proteins, and thus have a larger volume to encapsidate RNA. The particles do not have an identifiable maturation protein, but due to class averaging it could be anywhere and be removed from the final structure, so it is hard to say that they do not have the maturation protein. Two unique classes of particles are the prolate particles which are composed of 150 coat proteins and have 5-fold symmetry at the dominant axis and the 'small' particles which are composed of 132 coat proteins (Figure IV-6).

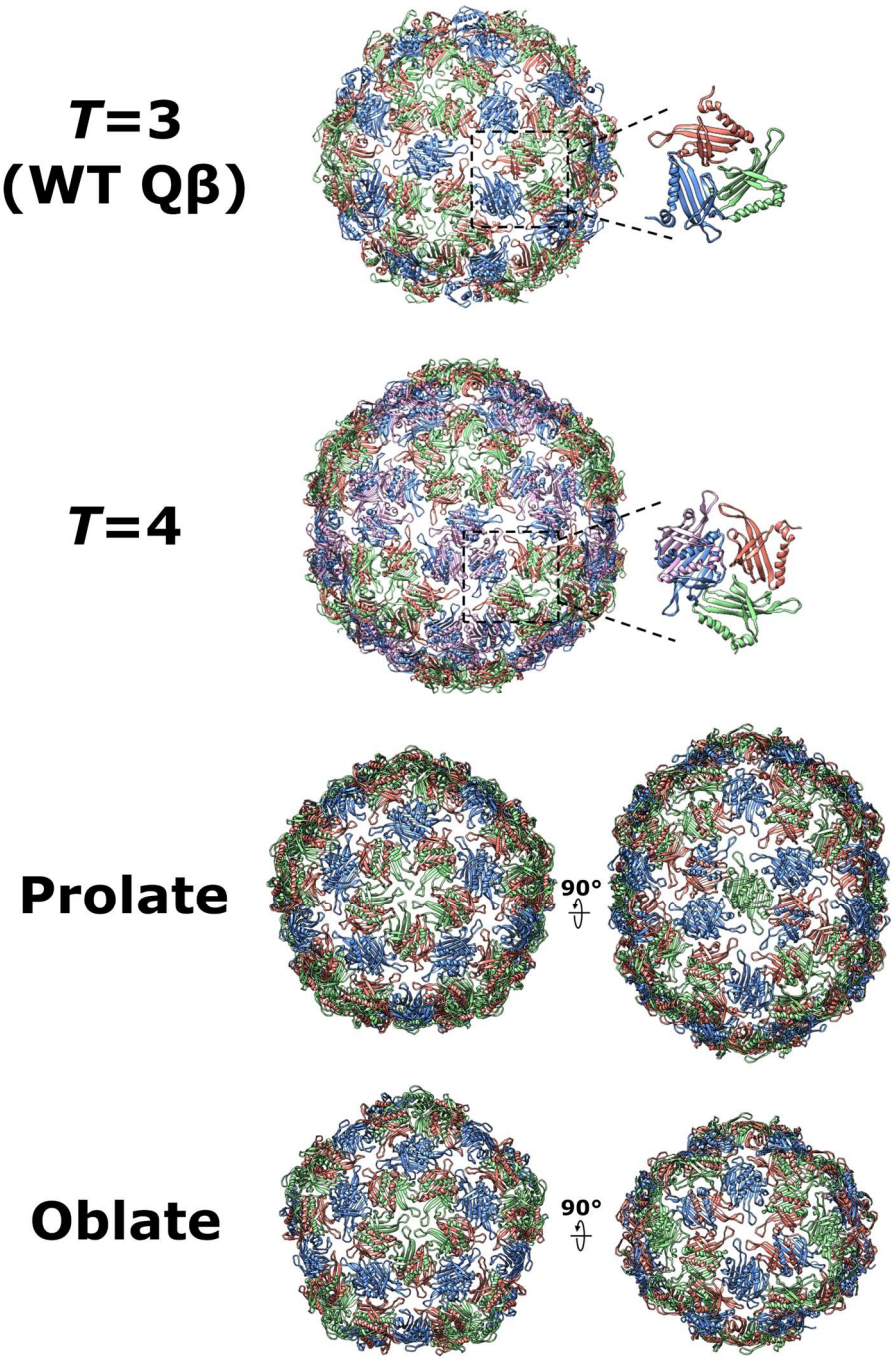

Figure IV-6. Alternative Qβ capsid morphologies
Qβ can form canonical *T*=3 capsids, as well as elongated *T*=4 capsids, as have been reported for other ssRNA phage coat proteins. The additional prolate and small capsids are unique, so far, to Qβ. The coat proteins are colored as their conformer would be in the wild-type *T*=3 capsid. The data collection, processing, and analysis done by Jeng-Yih Chang, from particles purified by Karl Gorzelnik.

That these new particles have never been detected is probably not due to the unique features of Qβ, but rather a bias in purification and structural analyses done by any group, ourselves included. It is not hard to understand why someone would not chase something they cannot catch. When doing a purification most groups would follow after activity, rather than a protein band on a gel. Even when analyzing data, some small amounts may be hidden, as is the case with the *T*=4 capsids. When viewed by any angle other than seeing the capsid at its widest, the virus would appear to be a *T*=3 capsid, and might be incorporated into those structures. As a result, the numbers for the *T*=4 capsid are most certainly smaller than the amount produced, but are hidden.

Knowing that the coat proteins are rather promiscuous, assembling around RNAs which are not full-length, wild-type gRNAs, I hypothesized that virus-like particles could be used to deliver an RNA of interest to *E. coli*. Looking at the genome of Qβ, there are the fewest 'operator-like' sequences in the coat protein. This was probably evolved to prevent reduced translation of the coat protein, which is required at the highest levels of all ssRNA phage proteins (Weber and Konigsberg 1975). With a lower number of 'probable' coat binding sites, at least compared to other regions of the genome, the coat/A_1 was replaced with eGFP (Figure IV-7). The coat/A_1 ribosome binding site was left intact, for a high level of eGFP expression once the viral gRNA entered the cell. The VLPs can still be assembled with the true-operator and plenty of

other operator-like sequences spread throughout the genome, as well as the maturation protein binding RNA stem-loops at the end of the genome.

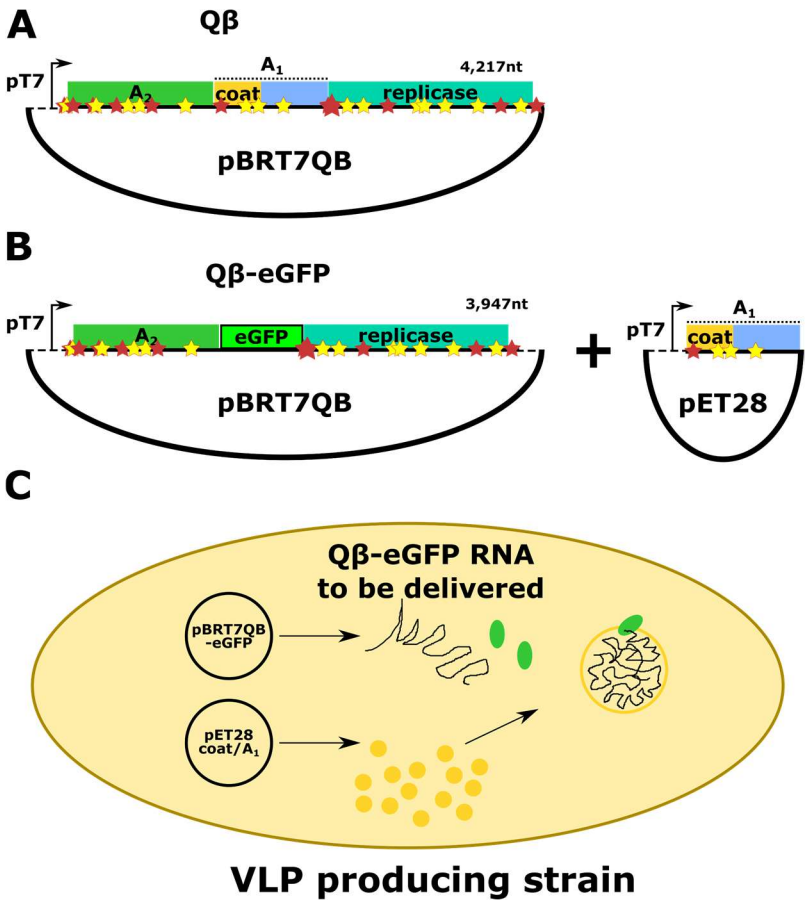

Figure IV-7. Schematic for the production and delivery of eGFP RNA within the Qβ gRNA
Qβ on a cDNA plasmid (A) was modified to replace the coat protein and A1 with eGFP (B). The coat protein and A1 were supplied in trans on a separate plasmid. When producing the eGFP-VLPs (C), the plasmids were in a host with the T7 polymerase, in order to overexpress both the Qβ-eGFP gRNA and the coat protein / A_1.

Once the VLPs are assembled, and purified as before, they can be used to infect male *E. coli*, in this case, the same host as a regular Qβ infection, ER2738. When imaging the infection, it is apparent that at least some VLPs are infectious, as the host is

not normally green, but there are green cells able to be seen (Figure IV-8). It is hard to determine the efficiency of infection, as these VLPs are not able to form new phage on their own, so a purification cannot be titered.

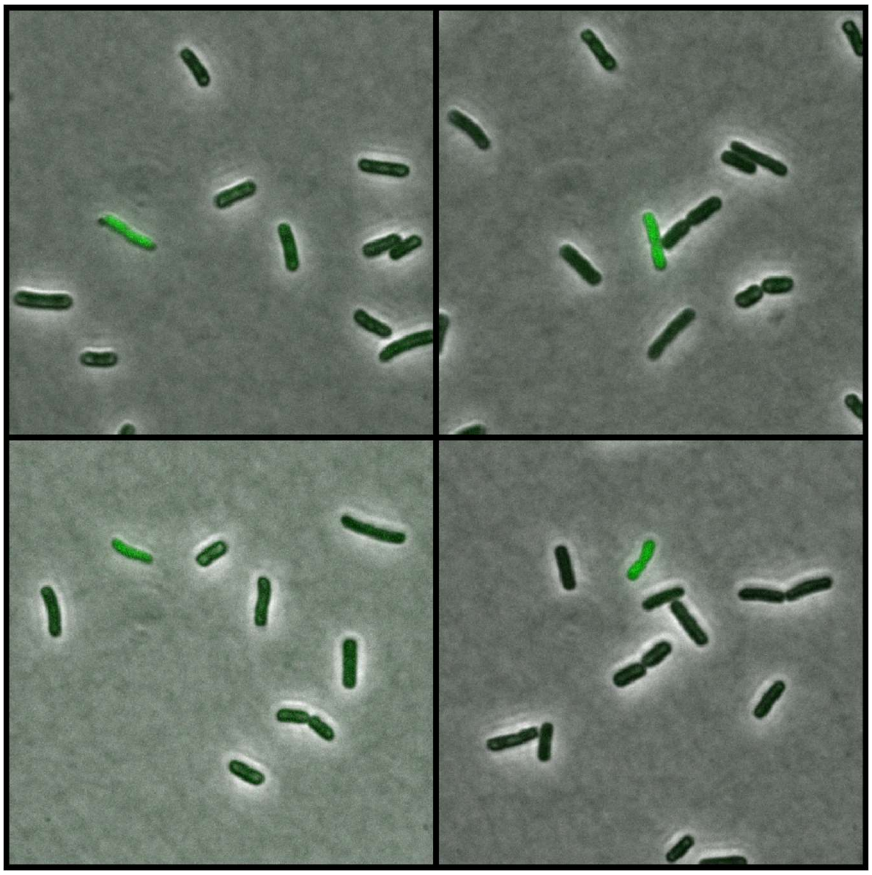

Figure IV-8. Qβ-eGFP VLPs delivering eGFP RNA
Representative images of Qβ-eGFP VLPs delivering eGFP RNA into *E. coli*. ER2738 (F^+) was infected with VLPs which contained eGFP in place of coat/A_1.

The Qβ-eGFP VLPs, which lack a coding sequence for the coat/A_1, were also titered onto male *E. coli* which had coat/A_1 on an inducible plasmid (pZA12), with varying levels of IPTG added into the plates. None of the VLPs titered with any level of

IPTG, presumably because with any IPTG the coat protein will be in a high enough concentration to repress the replicase. We also do not know what the effect of replacing a third of the genome will have on packaging the new virus. What this does show is that some VLPs are able to be made, and some of those VLPs are able to package RNA encoding eGFP. If the RNA was exposed to the environment it would be degraded by RNases and would not be able to be taken up into the cell without binding the maturation protein.

In an effort to determine the packaging efficiency, other than a traditional titer, a kanamycin resistance gene was cloned out of pET28 and put into the Qβ genome in place of the coat/A_1 (Figure IV-9). The rationale behind this was that one could back calculate the number of colonies from the amount used for an infection to determine the number of viable infectious particles. This worked better in theory than in practice. A liquid culture in exponential growth (OD_{600} = ~0.4) was taken and infected with a Qβ-Kan^R VLP preparation, let recover for half an hour, then either had kanamycin added to the liquid culture or was spun down and spread onto kanamycin plates. The samples were then grown overnight and the OD_{600} measured or colonies counted. The Qβ-Kan^R VLPs were able to give some resistance to kanamycin, as seen in panels B and C of Figure IV-9.

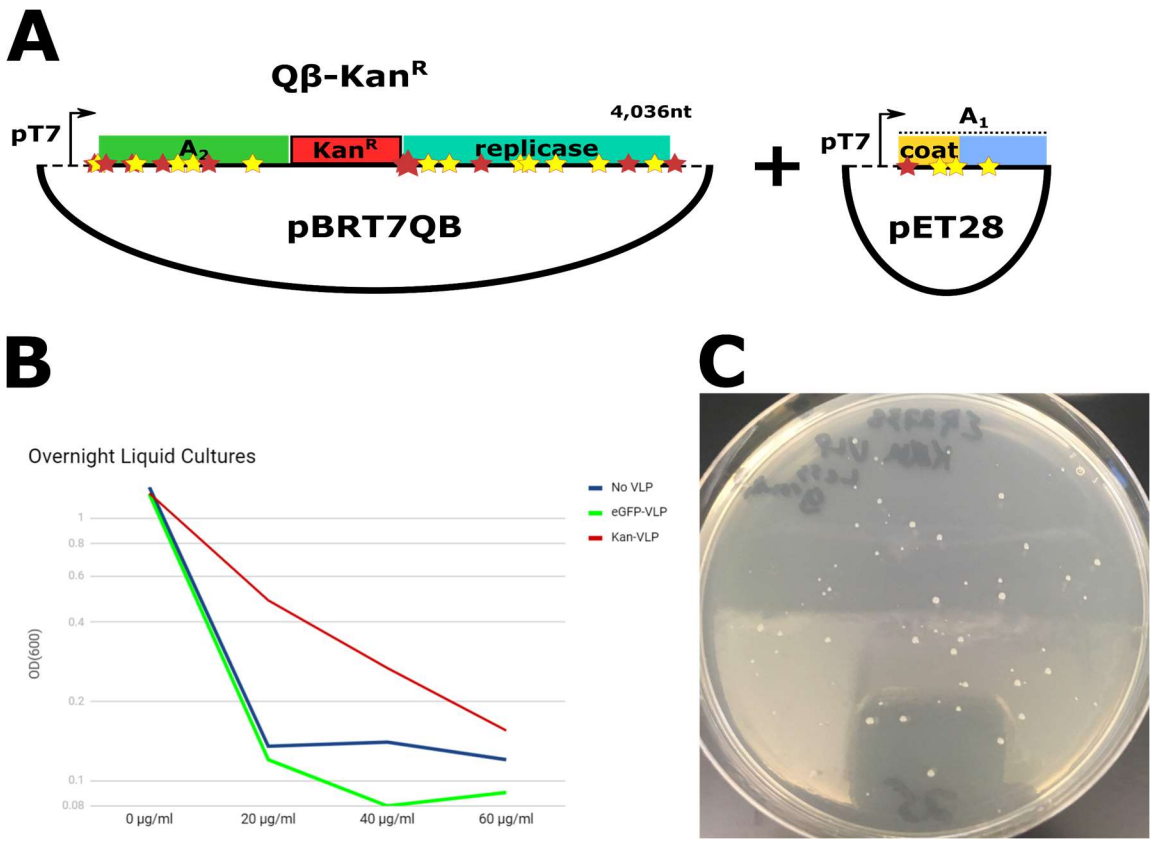

Figure IV-9. Kanamycin resistance RNA delivered into *E. coli*
Schematic for where the kanamycin resistance was inserted into the Qβ cDNA (A) and how the coat protein / A_1 was overexpressed to assemble into a VLP. B) The kan^R-VLP was added to liquid cultures of *E. coli* with varying amounts of kanamycin (described in the materials and methods) and grown overnight. The same cultures were spread onto kanamycin plates (C).

This was repeated at least three times for each experiment. However, the colonies cannot be struck from these plates onto fresh kanamycin plates, and the liquid culture growth cannot be passaged into a new tube. The most likely explanation is that the Qβ-Kan^R RNA is not stable. There are two possibilities for this: host RNases degrade the RNA or uncontrolled expression of the replicase leads to cell death. Either is likely. Inducing expression of the MS2 replicase will eventually kill the cell (Figure IV-10),

even without the lysis protein, as the replicase will alter the RNA levels of important genes, throwing off homeostasis in the cell. The other possibility is that the RNA is degraded by host RNases. The RNA in ssRNA phages is highly branched, for several reasons, to better package inside a capsid, to protect from RNases, or to prevent ssRNA from coming back on itself to form dsRNA which prevents replication by the RdRp (Hohn 1969, Tomita, Ichihashi et al. 2015, Beren, Dreesens et al. 2017).

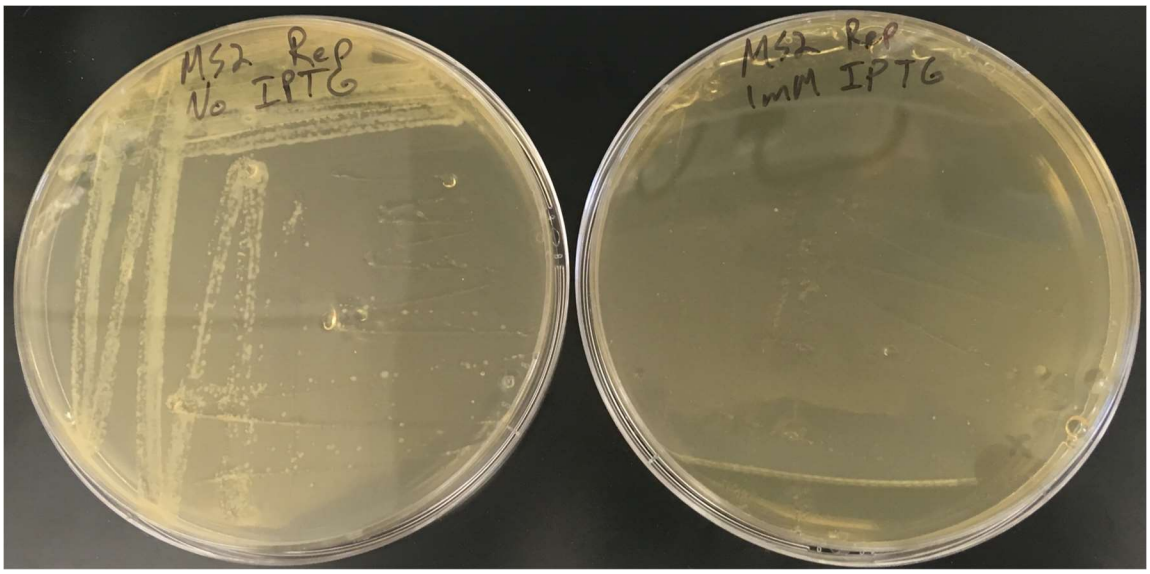

Figure IV-10. MS2 replicase expression is lethal
The MS2 replicase cloned into pET28 is lethal when expressed. At 1mM IPTG you can get induction in an exponentially growing culture, but if it is struck out for individual colonies the cells cannot grow. BL21(DE3) with the MS2 replicase in pET28 was grown in an overnight, then spotted onto plates with or without IPTG, and struck for single colonies. The strain was a gift from Jirapat Thongchol.

In any event, that the Qβ-Kan^R RNA infections were not able to support resistance to kanamycin for long is good, inasmuch as the cells did not evolve resistance

to kanamycin. Transient gene expression is actually the best reason to use RNA as a delivery tool, at least as far as regulatory agencies are concerned, because there is no risk of integration. With that in mind, the Qβ-eGFP VLPs were used to see if the VLP could deliver eGFP RNA to eukaryotic cells (Figure IV-11). What was found was unexpected, as a eukaryote should not be able recognize a bacterial ribosome binding site. However, the whole cells lit up green.

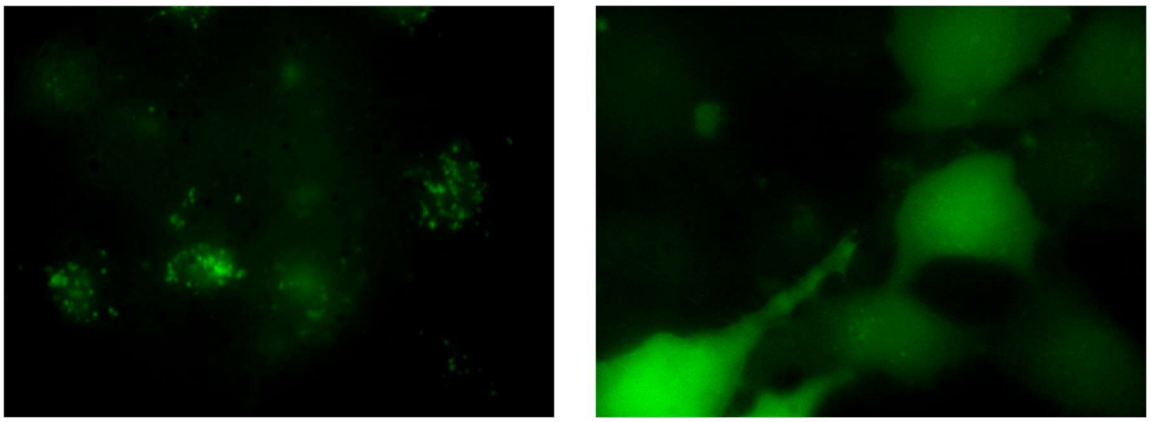

Figure IV-11. eGFP-VLPs used to transfect HeLa cells
The eGFP-VLPs used for bacterial infections were used to infect mammalian cells. Done by Aaron Jacobson of the Pellois lab, with VLPs purified by Karl Gorzelnik.

This was unexpected, as the Qβ-eGFP VLPs were purified via size-exclusion over an S500 column, with the capsids isolated from the same fractions as wild-type (infectious) Qβ. An emission spectrum of the eGFP-VLP vs. wild-type Qβ shows that there is fluorescence from the eGFP-VLPs but not so much for wild-type.

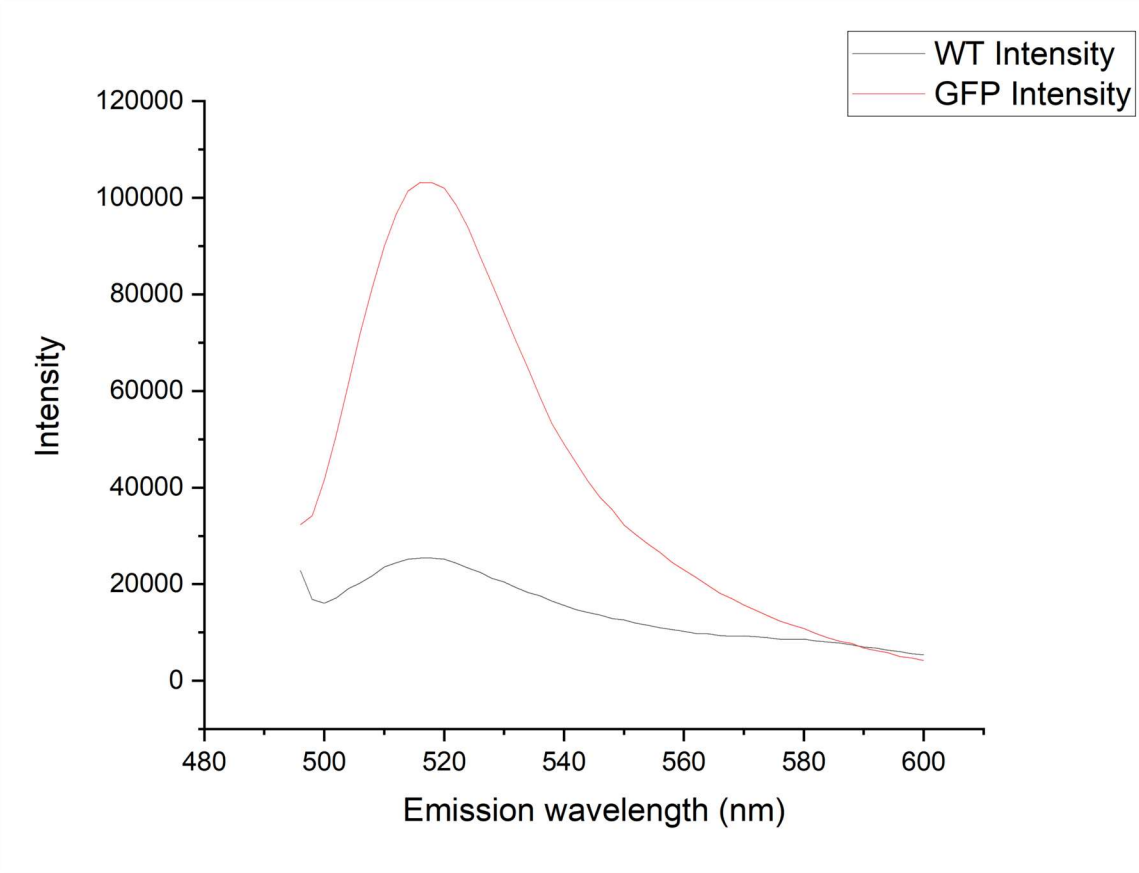

Figure IV-12. Wavelength scan of WT Qβ and the eGFP-VLP
Emission spectra for wild-type Qβ and Qβ-eGFP VLPs. Normalized (diluted) before putting into a spectrophotometer for equivalent levels of protein. Excitation wavelength is 488nm. Measurements done by Aaron Jacobson.

Since the GFP eluted with the VLPs, not in the void volume fraction or later after the VLPs, in all likelihood the GFP is either within the VLP or associated with it. If it was aggregated together it would either come earlier (the void volume) or later (smaller particles). It is possible that the GFP is in capsids that do not have any RNA, or have smaller RNAs but not a full-length RNA. In any event, it does not matter for infection of bacterial cells, as the fluorescence seen inside of a cell cannot come from the cell

taking up intact GFP. Determining the efficiency of packaging foreign RNAs into VLPs is a priority for any delivery applications.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Single-stranded RNA phages are not as simple as we think, or even make them out to be. Chapter II showed how complex they are when packaged, while Chapter III showed how easy it is to disturb the infectivity of these organisms. Chapter IV then showed how promiscuous, and thus hardy, they actually are, while also showing they are a lot weirder than we thought. The structure of Qβ was interesting, four years ago. What might have a higher impact is the more recent work, showing that not only does Qβ also form *T*=4 particles, which had only been shown for MS2 and PP7 when overexpressing their coat proteins, but that they also form smaller particles. That these particles are actually a large percentage of total coat proteins, *during an infection with wild-type phage*, as opposed to the *T*=4 particles or prolate, means that researchers have been biasing their structures of ssRNA phages, and perhaps many more viruses, simply by the purification techniques. Since the ssRNA phages all, to our knowledge, package their gRNA based off of their operator sequences, and alternative capsid morphologies have been seen for two other phages, it is in all likelihood the case that the other ssRNA phages also form smaller capsids. This will certainly need to be tested.

Within these particles there is still RNA density. What is this RNA? The larger particles could encapsidate the 4,217nt of gRNA and then some, but the smaller particles? In all likelihood they are packaging host RNAs. The mechanism for this is

easy enough to comprehend: the coat proteins bind to the operator at a high level, but as was shown in Chapter II, they also bind to many 'operator-like' stem loops. Presumably, the higher the ratio of 'operator-like' stem-loops to that of the rest of the RNA, the greater the likelihood of packaging. It could also be that the interior-facing SRNRK loop of Qβ (Figure II-18), which is positively charged, can promiscuously bind RNAs. Mechanistically, either or both of these packaging methods could explain faulty packaging of non-viral RNAs, but what are they? Most likely they are host RNAs (Figure V-1). Unlike many dsDNA phages, which can degrade the host chromosomal DNA, ssRNA phages lack the coding capacity to specialize proteins to do so. Therefore, the host RNA polymerases will keep transcribing genes as would be normal. Even if an ssRNA phage is able to produce ~50,000 particles, that would still be slightly less than the number of ribosomes per cell (Bremer and Dennis 2008). Ribosomes are slightly smaller than ssRNA phages, so it is within the realm of possibility that ribosomal RNAs could be packaged within a virus-like particle.

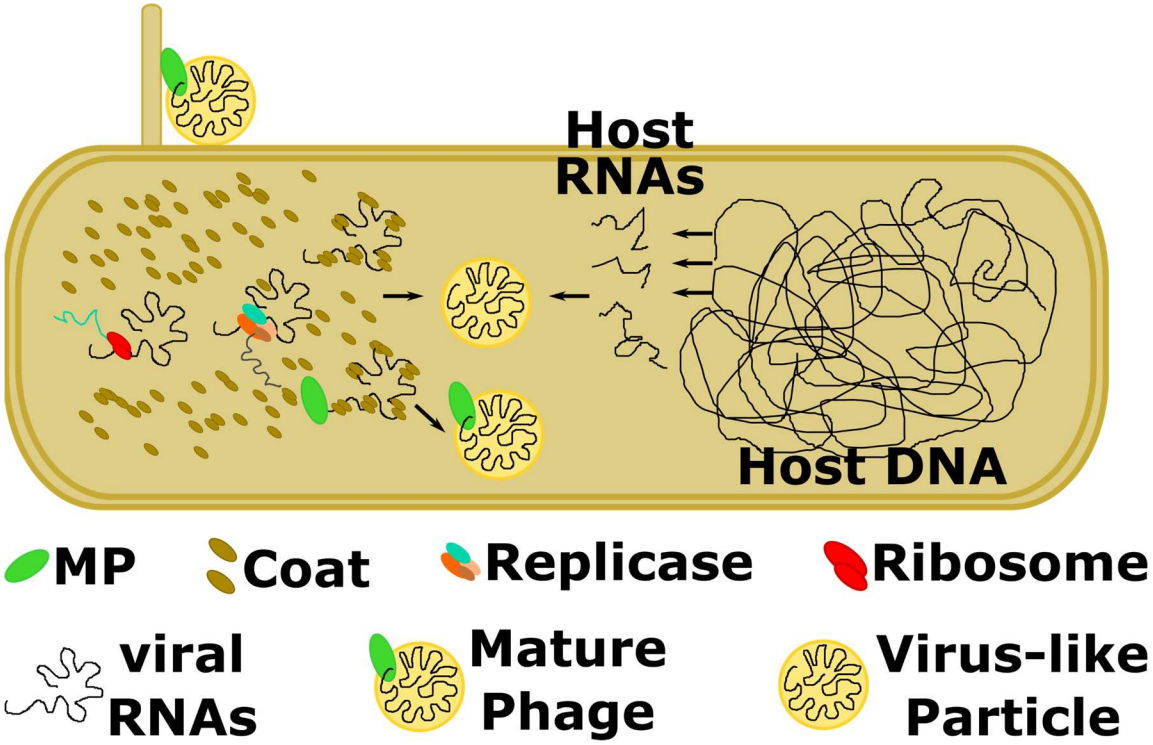

Figure V-1. Cellular dynamics during VLP assembly
The phage coat proteins are in a dynamic cellular environment with host RNAs as well, which could lead to virus-like particle (VLP) assembly with host RNAs surrounded by phage coat proteins. VLPs could be made with phage or host RNAs.

Once infection has occurred, the viral gRNA is in a dynamic cytoplasm which contains all the host RNAs, proteins, DNA, etc… The rules for packaging have yet to be established. There are certainly host RNAs with 'operator-like' motifs. All of the research on packaging signals for ssRNA phages has been with MS2. For MS2 to assemble into a *T*=3 VLP, there are packaging signals which indicate regions of the viral gRNA with a high-affinity for the coat protein (Rolfsson, Middleton et al. 2016). However, the studies with MS2 were done with the idea in mind that MS2 always forms *T*=3 capsids and *in vivo* exclusively uses the viral gRNA as a template. It was recently found that MS2 can form *T*=4 capsids when the coat protein is overexpressed (de Martin

Garrido, Crone et al. 2019). Just as this study found that both wild-type Qβ from an infection and the Qβ coat protein overexpressed can form *T*=4 capsids (Table IV-2). The reason that the MS2 coat protein was only found to form *T*=4 capsids in a low percentage in that study, and not form the oblate or prolate particles observed for Qβ, is the same as for the CsCl purifications of Qβ. The purification protocols used density as a means to separate out the particles, either a glycerol gradient in the case of the MS2 VLPs or CsCl in the case of Qβ produced from an infection. Presumably if the data collected from the gel filtration is scaled up in size, there will be more *T*=4 particles for Qβ, as the oblate and prolate particles increased significantly (Table IV-2).

It is interesting that Qβ was able to form the alternative capsid morphologies during an infection, as this has never been reported before for any ssRNA phage. Why the *T*=4, prolate, and oblate particles were assembled presumably goes back to packaging specificity. The rules for packaging are not fully understood. If packaging is based off of 'operator-like' stem loops, as proposed following the determination of the structure of Qβ in Chapter II, then any RNA with such a stem-loop could be packaged. While RNAs have been tagged with the operators of various ssRNA phages in order to assemble VLPs with the coat proteins of those phages, the efficiency of packaging is unknown. ssRNA phages have a much higher level of secondary structure than most bacterial RNAs. It is conceivable that the RNAs in the *T*=4, prolate, and oblate particles are mispackaged host RNAs. There certainly could be host RNAs in the canonical *T*=3

particles, with the phage coat proteins just driven toward the correct assembly through protein:protein contacts.

My model for the likelihood of packaging host RNAs, in any conformation of capsid, is that the more 'operator-like' stem-loops the more likely that RNA will be packaged (Figure V-2). As the coat proteins need to make protein:protein contacts, the likelihood of assembly will go up if the 'operator-like' structures are closer together, to create an area suitable for capsid assembly intermediates. These intermediates could be the 'faces' of the icosahedral capsid, where the faces are composed of pentamers or hexamers of coat protein dimers (Takamatsu and Iso 1982). The non-canonical capsids could be host RNAs which were able to make assembly intermediates, whose gaps were filled in by coat proteins binding. Even wild-type Qβ gRNA in the *T*=3 infectious capsid only has a handful of 'operator-like' stem-loops touching the capsid (as seen in Figure II-21). As the maturation protein replaces a coat protein dimer in a *T*=3 particle, there are 89 coat protein dimers in the capsid of a mature virion. Of those 89 coat protein dimers, 31 contact 'operator-like' stem-loops, while another 26 interact with stem-loops of the alternative 'handedness'. It is possible that the coat proteins bind the 'operator-like' stem-loops with a higher affinity, and the more of these stem-loops in an RNA, the more likely that particle will assemble into a virus-like particle, of whatever the different conformations. Indeed, the coat proteins could bind host RNAs then diffuse off of the RNA if there is not enough subsequent binding of other coat proteins onto that RNA.

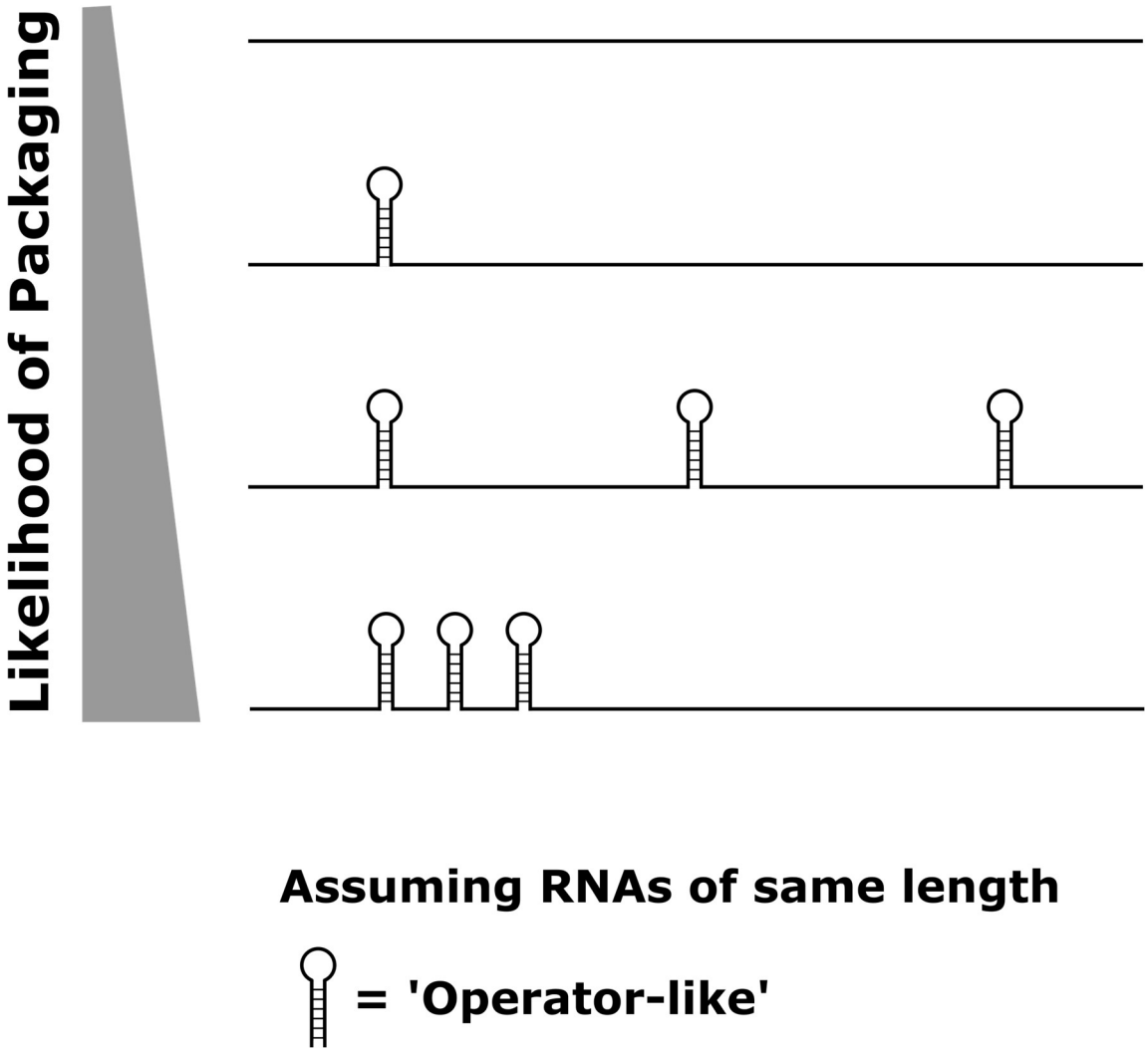

Figure V-2. Likelihood of packaging non-viral RNAs
Based on the non-canonical *T*=4, prolate, and oblate shapes that can be produced during a wild-type infection of Qβ, there could be foreign RNAs in the virus-like particles produced. Especially in the oblate particles which are smaller than the canonical *T*=3 capsids. Assuming there are four RNAs of the same length, the likelihood of packaging is increased if there is an operator-like sequence in the RNA (second line). The more operator-like sequences in an RNA (third line) the more likely that the RNA will be packaged. If there are multiple operator-like sequences in close proximity (fourth line) then they could locally facilitate a nucleation event.

A more comprehensive analysis of the purified capsids will also be something worth undertaking. So far, phage and VLPs can be purified away from the ribosomes, but are there smaller capsids that were missed by excluding fractions containing contaminating ribosomes. If so, are the RNAs within the smaller capsids different than those of the larger, *T*=4 and prolate capsids? What about the 70% of Qβ particles which were not able to be used in the asymmetric reconstruction? While these were still *T*=3 VLPs, the fact that they did not have the maturation protein could mean that they do not have the correct RNA. What were the RNAs in those particles? Capsids based off of ssRNA phages have been used to protect a large number of RNAs, but no one has looked at how many host RNAs are also being protected. This could cause contamination issues down the line, particularly if primers used have homology to any *E. coli* genes. If the VLPs are used for delivering genes into eukaryotes then there is the potential that a gene for a bacterial toxin could also be encoded.

Delivery of RNA (or other cargos) to eukaryotes is the most likely use of ssRNA phage coat proteins. However, as shown in Chapter IV, the delivery of RNA to bacteria is a possibility. Previously, the only way to deliver RNA to bacteria was to electroporate it. This not practical for anything but laboratory studies. Even in the lab RNases are everywhere, and thus the efficiency will not be great. By using RNA protected by a capsid you not only ensure that the RNA remains intact, but also that only the desired RNAs (without contaminating DNA) is delivered. The reason that the gene for the coat protein/A_1 was replaced with foreign RNAs in the delivery experiments in Chapter IV

was because it had fewer operator-like sequences. This was probably evolved, to reduce coat protein binding, and thus increase the translation levels. After all, the coat protein is the highest produced ssRNA phage protein.

In addition to the possibility that host RNAs are packaged into capsids, the RNA within particles that do not fit into the dominant classification (for Qβ) could also be negative-sense gRNA. While these would not be infectious particles, the gRNA is inherently highly-branched. The negative-sense gRNA is just as highly branched as the positive-sense gRNA, and there are several 'operator-like' sequences in it as well. A brief examination of RNAfold secondary structure prediction shows at least five 'operator-like' stem-loops which have an adenosine in the third position of a three-codon loop (Figure V-3). Since there are more stem-loops which bind to the gRNA in the same conformation as the operator, the number of high-affinity stem-loops cannot be strictly limited to those with an adenosine in the third position of a three-codon loop. Even so, looking for stem-loops which are 'operator-like' is a good enough approximation to say that the negative-sense gRNA could be encapsidated.

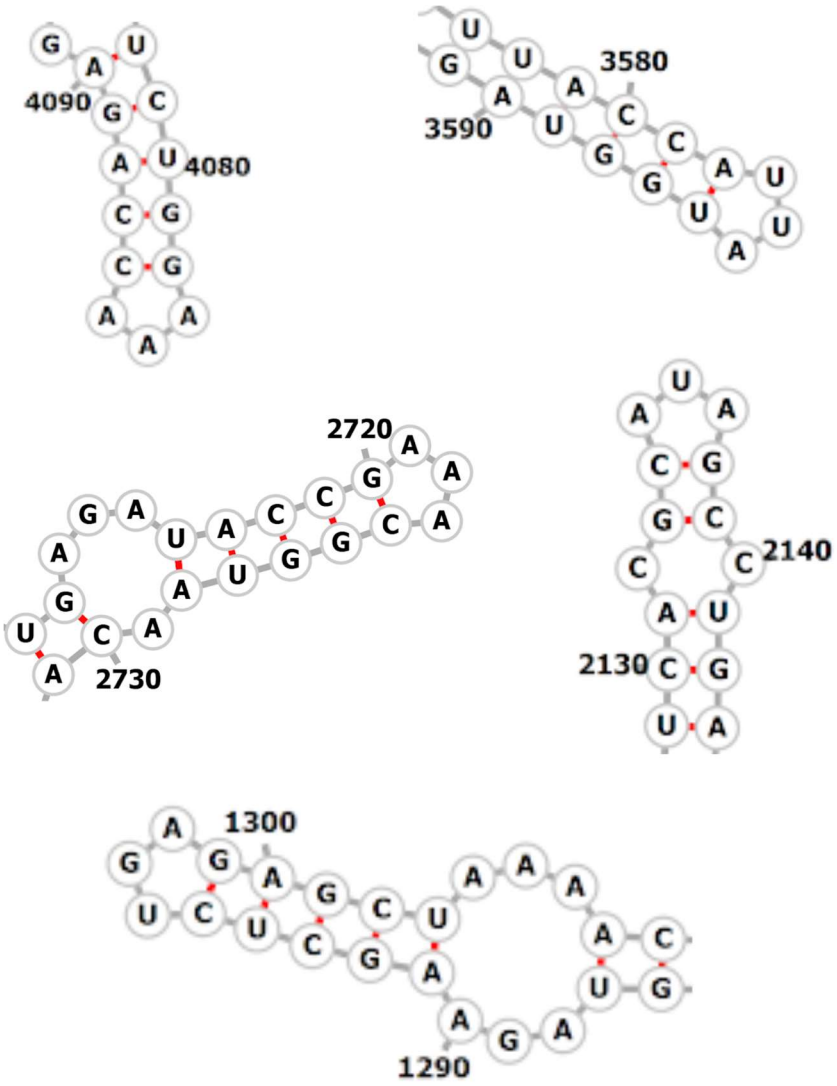

Figure V-3. 'Operator-like' stem-loops in negative-sense Qβ gRNA
When running RNAfold on the negative-sense of the Qβ gRNA, there are at least five 'operator-like' sequences. The nucleotide numbers correspond to the reverse-complement of the gRNA, so the stem-loop in the top left would be in the RNA of the maturation protein, if it was in the positive-sense, while the bottom would be in the replicase.

While the final dataset released from the high-resolution structure of MS2 structure had >95% of the particles coming together (Dai, Li et al. 2017), a previous study had at least 30% of particles not having the maturation protein (Toropova,

Stockley et al. 2011). The number of particles lacking the maturation protein in the previous study is more akin to the numbers that were found for Qβ, although we only had ~30% with the MP (Cui, Gorzelnik et al. 2017). The low percentage of particles able to be used may be due to the fact that Qβ has more long-distance interactions, which are required for infectivity (Beekwilder, Nieuwenhuizen et al. 1996, Klovins, Berzins et al. 1998, Klovins and van Duin 1999). The only long-range interaction MS2 has is the UPS-SD upstream of the A protein, but this is only 100nt in distance, whereas Qβ has many that are very long distance interactions (Groeneveld, Thimon et al. 1995). It is possible that long-range interactions which are not folded correctly prevent the maturation protein from binding. After all, in Chapters III and IV, there is evidence that the maturation protein binding the gRNA in Qβ is very sensitive to RNA secondary structure.

Another possible explanation for why there is non-genomic RNA within the capsid is that the replicases can copy almost any RNA at least once. After one round of replication, 'illegitimate' templates, i.e. those not suited to replication, will be removed from the pool of RNAs able to be copied by the replicase when the minus-strand anneals to the plus-strand. Those RNAs with enough secondary structure to prevent this from happening will keep being replicated (Yumura, Yamamoto et al. 2017). Incidentally, RNAs with higher levels of secondary structure will probably also be packaged at a higher rate, as is seen in the frequency of stem-loops interacting with the capsid in Chapter II. Qβ has been shown to have a unique RNA fraction produced during

infection, the ‘6S’ RNA (Trown and Meyer 1973). This has been shown to be parasitic RNA, derived from host RNAs, which is able to be replicated by the replicase (Avota, Berzins et al. 1998). It would be interesting if an infection by Qβ or other ssRNA phages could support a satellite virus, that is an RNA which can bind the maturation protein and thus be taken up into another cell. If the RNA is replicated but cannot reinfect, it will die out after one infection round, if it could reinfect along with a host, then it could by definition be a satellite virus. Such binding of the maturation protein is required for a satellite ssRNA to be a satellite virus of an ssRNA phage, otherwise the RNA would either be left inside the phage capsid as the pilus machinery pulls the maturation protein / vRNA complex inside the host cell or left in the lysed remains of the host cell and would not be propagated in the next generation.

It is also possible that some of the stem-loops which bind the capsid *in virio* do not play a role in packaging the gRNA, but are merely present in the RNA for the sake of replication, as the known secondary structures of phages from *Leviviridae* are highly branched. Not all the proposed packing signals in the gRNA may bind to the capsid proteins *in vitro* and *in vivo*, leading to a model that is less accurate.

During the dual process of replication and packaging the conformations of the ssRNA secondary structure may vary at different stages, depending on the concentration of coat proteins relative to the replicase or even the ribosome. While this is the canon of

ssRNA phage packaging, only ~10% of viruses are actually viable. What explains the difference between assembled viruses and infectious particles?

Packaging RNA based on recognition of stem-loops is a tricky business. The consensus sequence restricts the RNAs which could theoretically be packaged but is not stringent for just viral RNA. The operator for Qβ is merely a four base pair stem with a three nucleotide loop, that has an adenosine in the third position. This is found in many RNAs. The Qβ operator/coat protein interaction has even been commercialized to protect RNAs of interest from RNases, most often for diagnostic laboratories. If a single operator is placed on an RNA of interest the RNA will be encapsidated by ssRNA phage coat proteins, forming virus-like particles. Just as the capsids protect the phage ssRNA from RNases, the coat dimers which surround the designated RNA shield the diagnostic (or other) RNAs. This phenomenon leads to the question of how many phage particles actually encapsidate viral RNA.

If all it takes is a single operator or operator-like sequence, then many host RNAs could theoretically be encapsidated. Our current model for virion morphogenesis is that multiple operator-like sequences bind coat proteins to facilitate particle formation. While one operator-like sequence is sufficient to enable a coat protein dimer to bind RNA and recruit other coat protein dimers to assemble a virus-like particle, the rate of assembly should increase if more operator-like sequences are proximal to each other (Figure V-4). There may be many reasons why only ~10% of viral particles are

infectious. The maturation protein might be sequestered away inside the virus. The encapsidated particle might not contain quite enough coat protein dimers to keep RNases away from the gRNA. The virus might be mispackaged in such a way that the gRNA cannot be fully released. Or simply, the maturation protein might not bind to the gRNA before it is encapsidated.

The R1 and U1 mutations show that the maturation protein, especially in Qβ, is highly specific to the gRNA. In the future, the specificity will be examined more deeply. Do single nucleotide changes decrease binding or abolish affinity? The mutations were originally designed to validate a model of the gRNA within the asymmetric structure. What the 3' mutations revealed may facilitate the swapping of maturation proteins between phage, even the targeting of new bacteria. The RNA delivery experiments in Chapter IV show, it is possible to swap cargo, so why shouldn't it be possible to swap targets?

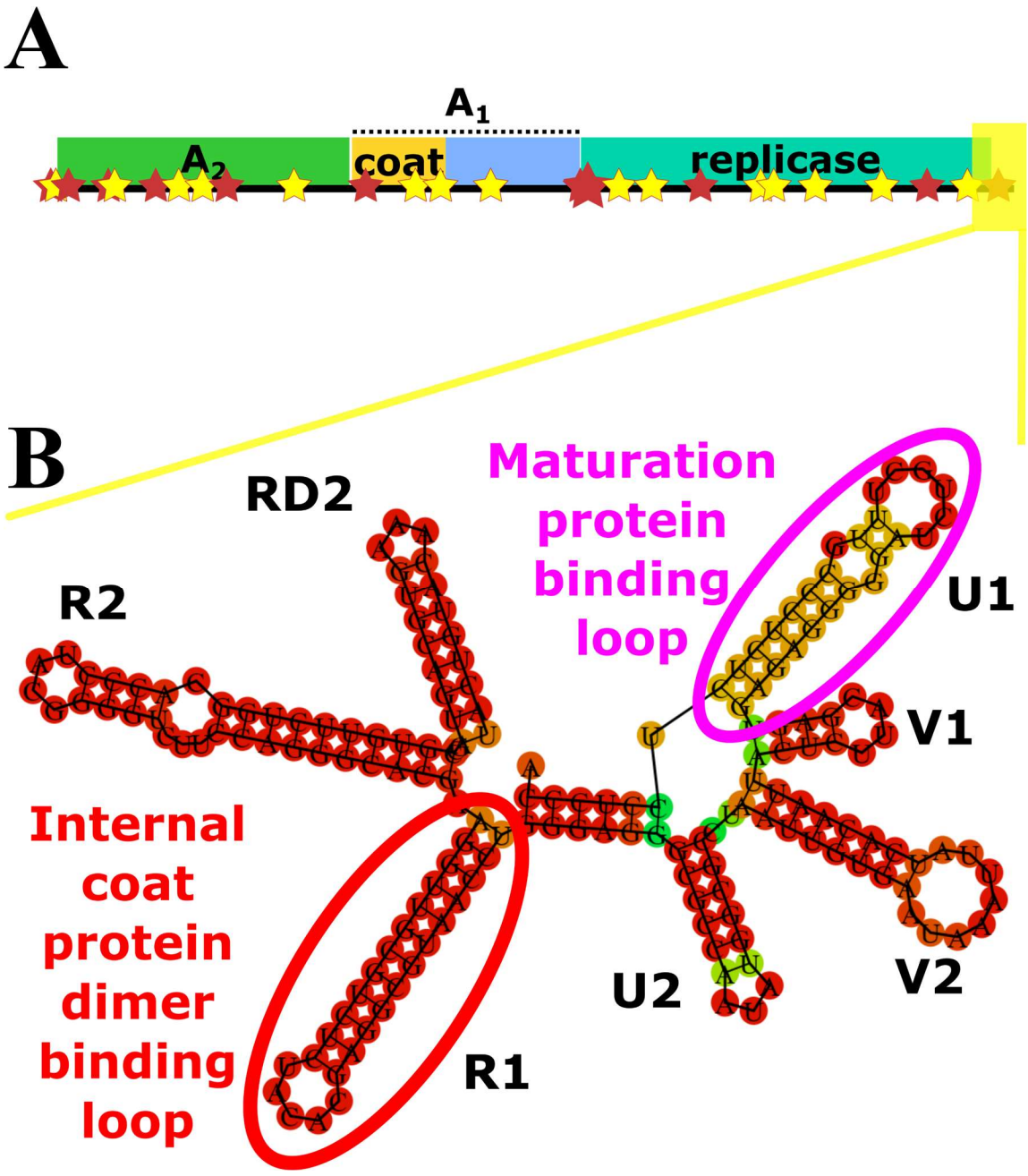

Figure V-4. Requirements for assembly and delivery
The more operator-like sequences an RNA has (A) the more likely it will be encapsidated with the maturation protein. Delivery to bacteria, via retractile pili, depends on the maturation protein being able to bind an RNA, which for A_2 and Qβ means that you need the R1 and U1 stem-loops (B).

One immediate trial which can be attempted from the work is the delivery of fluorescent RNA. That foreign RNAs can be delivered to bacteria has been shown in this work (Chapter IV). In the process of completing this work, a series of dBroccoli aptamers, which can fluoresce when bound to a small molecule, were cloned into the cDNA plasmid of Qβ (Figure V-5). This construct could be co-expressed with the coat protein/A_1 and the maturation protein to produce an infectious virus-like particle with fluorescent RNA. When this particle will be used to infect F^+ *E. coli* the point where the RNA leaves the capsid will be able to be determined. The current model is that the gRNA does not leave the capsid until the phage is at the base of the pilus. This is due to work showing that the phage RNA is not released when the phage are applied to sheared pili.

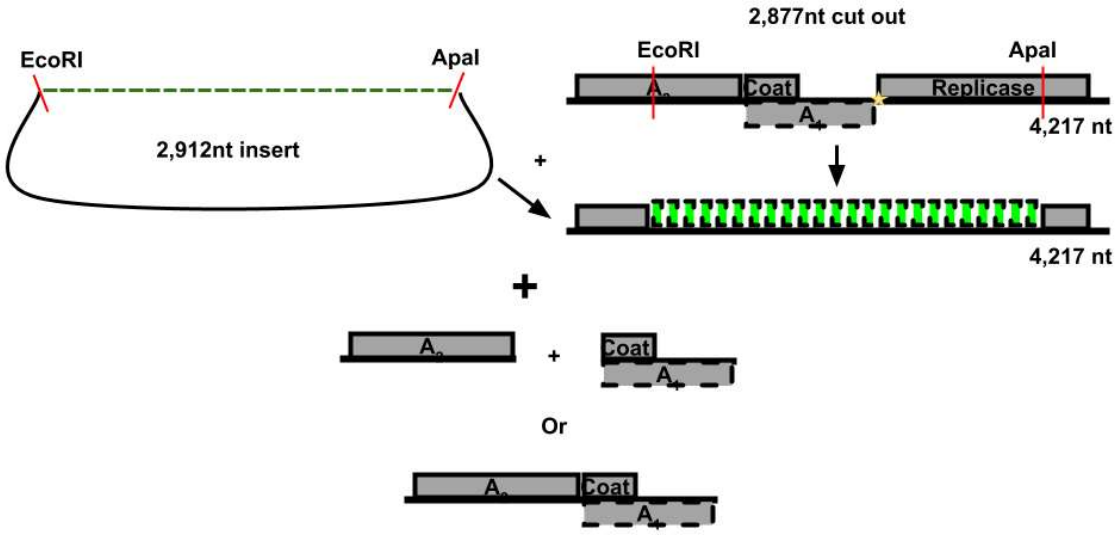

Figure V-5. dBroccoli cloned into pBRT7QB, with potential expression systems
24 dBroccoli aptamers were cut out of pCDNA3_mCherry (a kind gift of Xing Li and Sam Jaffrey) and ligated into pBRT7QB, using traditional restriction-ligation techniques. With coat/A_1 and the maturation protein (A_2) expressed on separate

plasmids this should assemble into an infectious VLP. The modified pBRT7QB should be able to transcribe gRNA with 24x dBroccoli aptamers which can be packaged into an infectious VLP, as per the requirements established in Figure V-4.

What we have learned from this work is that the coat proteins assemble around RNA. They can form non-canonical shapes, probably when they bind non-native RNAs, but this will have to be explored more. Presumably, the more 'operator-like' sequences there are, the better the RNA will be encapsidated. While there was a low efficiency of delivery for the GFP and kanamycin resistance RNAs, it was probably higher than it would have been without the other 3,000nt of Qβ on either side. Finally, in order to deliver a cargo, an ssRNA phage (or VLP) must have the maturation protein incorporated in it. Ideally, any cargo delivered would also have RNA for the replicase delivered as well, in order to increase expression levels once inside the new host.

REFERENCES

- Achtman, M., N. Willetts and A. J. Clark (1971). "Beginning a genetic analysis of conjugational transfer determined by the F factor in Escherichia coli by isolation and characterization of transfer-deficient mutants." J Bacteriol **106**(2): 529-538.

- Adrian, M., J. Dubochet, J. Lepault and A. W. McDowall (1984). "Cryo-electron microscopy of viruses." Nature **308**(5954): 32-36.

- Ahlquist, P., A. O. Noueiry, W. M. Lee, D. B. Kushner and B. T. Dye (2003). "Host factors in positive-strand RNA virus genome replication." J Virol **77**(15): 8181-8186.

- Aksyuk, A. A. and M. G. Rossmann (2011). "Bacteriophage assembly." Viruses **3**(3): 172-203.

- Alfson, K. J., L. E. Avena, M. W. Beadles, H. Staples, J. W. Nunneley, A. Ticer, E. J. Dick, Jr., M. A. Owston, C. Reed, J. L. Patterson, R. Carrion, Jr. and A. Griffiths (2015). "Particle-to-PFU ratio of Ebola virus influences disease course and survival in cynomolgus macaques." J Virol **89**(13): 6773-6781.

- Ando, A., K. Furuse and I. Watanabe (1979). "Propagation of ribonucleic acid coliphages in gnotobiotic mice." Appl Environ Microbiol **37**(6): 1157-1165.

- Arribas, M., L. Cabanillas, K. Kubota and E. Lazaro (2016). "Impact of increased mutagenesis on adaptation to high temperature in bacteriophage Qbeta." Virology **497**: 163-170.

- Ashcroft, A. E., H. Lago, J. M. Macedo, W. T. Horn, N. J. Stonehouse and P. G. Stockley (2005). "Engineering thermal stability in RNA phage capsids via disulphide bonds." J Nanosci Nanotechnol **5**(12): 2034-2041.

- August, J. T., S. Cooper, L. Shapiro and N. D. Zinder (1963). "RNA Phage Induced RNA Polymerase." Cold Spring Harbor Symposia on Quantitative Biology **28**(0): 95-97.

- Avota, E., V. Berzins, E. Grens, Y. Vishnevsky, R. Luce and C. K. Biebricher (1998). "The natural 6 S RNA found in Q beta-infected cells is derived from host and phage RNA." J Mol Biol **276**(1): 7-17.

- Bai, X. C., I. S. Fernandez, G. McMullan and S. H. Scheres (2013). "Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles." Elife **2**: e00461.

- Baker, T. S. and R. H. Cheng (1996). "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy." J Struct Biol **116**(1): 120-130.

- Banerjee, A. K., U. Rensing and J. T. August (1969). "Replication of RNA viruses." Journal of Molecular Biology **45**(2): 181-193.

- Bansho, Y., T. Furubayashi, N. Ichihashi and T. Yomo (2016). "Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system." Proc Natl Acad Sci U S A **113**(15): 4045-4050.

- Barrera, I., D. Schuppli, J. M. Sogo and H. Weber (1993). "Different mechanisms of recognition of bacteriophage Q beta plus and minus strand RNAs by Q beta replicase." J Mol Biol **232**(2): 512-521.

- Basnak, G., V. L. Morton, O. Rolfsson, N. J. Stonehouse, A. E. Ashcroft and P. G. Stockley (2010). "Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid." J Mol Biol **395**(5): 924-936.

- Bayer, M., R. Eferl, G. Zellnig, K. Teferle, A. Dijkstra, G. Koraimann and G. Hogenauer (1995). "Gene 19 of plasmid R1 is required for both efficient conjugative DNA transfer and bacteriophage R17 infection." J Bacteriol **177**(15): 4279-4288.

- Beckett, D. and O. C. Uhlenbeck (1988). "Ribonucleoprotein complexes of R17 coat protein and a translational operator analog." Journal of Molecular Biology **204**(4): 927-938.

- Beckett, D., H.-N. Wu and O. C. Uhlenbeck (1988). "Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly." Journal of Molecular Biology **204**(4): 939-947.

- Beekwilder, J., R. Nieuwenhuizen, R. Poot and J. v. Duin (1996). "Secondary Structure Model for the First Three Domains of Qβ RNA. Control of A-protein synthesis." Journal of Molecular Biology **256**(1): 8-19.

- Beekwilder, J., R. Nieuwenhuizen, R. Poot and J. van Duin (1996). "Secondary structure model for the first three domains of Q beta RNA. Control of A-protein synthesis." J Mol Biol **256**(1): 8-19.

- Beekwilder, M. J., R. Nieuwenhuizen and J. van Duin (1995). "Secondary Structure Model of the Last Two Domains of Single-stranded RNA Phage Qβ." Journal of Molecular Biology **247**(5): 903-917.

- Belyi, V. A. and M. Muthukumar (2006). "Electrostatic origin of the genome packing in viruses." Proc Natl Acad Sci U S A **103**(46): 17174-17178.

- Bendis, I. and L. Shapiro (1970). "Properties of Caulobacter ribonucleic acid bacteriophage phi Cb5." J Virol **6**(6): 847-854.

- Beren, C., L. L. Dreesens, K. N. Liu, C. M. Knobler and W. M. Gelbart (2017). "The Effect of RNA Secondary Structure on the Self-Assembly of Viral Capsids." Biophys J **113**(2): 339-347.

- Berestowskaya, N. H., V. D. Vasiliev, A. A. Volkov and A. B. Chetverin (1988). "Electron microscopy study of Qβ replicase." FEBS Letters **228**(2): 263-267.

- Berkhout, B., B. F. Schmidt, A. van Strien, J. van Boom, J. van Westrenen and J. van Duin (1987). "Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene." Journal of Molecular Biology **195**(3): 517-524.

- Berkhout, B. and J. van Duin (1985). "Mechanism of translational coupling between coat protein and replicase genes of RNA bacteriophage MS2." Nucleic Acids Res **13**(19): 6955-6967.

- Bernhardt, T. G., I. N. Wang, D. K. Struck and R. Young (2001). "A protein antibiotic in the phage Qbeta virion: diversity in lysis targets." Science **292**(5525): 2326-2329.

- Billeter, M. A., M. Libonati, E. Vinuela and C. Weissmann (1966). "Replication of viral ribonucleic acid. X. Turnover of virus-specific double-stranded ribonucleic acid during replication of phage MS2 in Escherichia coli." J Biol Chem **241**(20): 4750-4757.

- Blumenthal, T. and G. G. Carmichael (1979). "RNA replication: function and structure of Qbeta-replicase." Annu Rev Biochem **48**: 525-548.

- Blumenthal, T., T. A. Landers and K. Weber (1972). "Bacteriophage Q replicase contains the protein biosynthesis elongation factors EF Tu and EF Ts." Proc Natl Acad Sci U S A **69**(5): 1313-1317.

- Blumenthal, T., T. A. Landers and K. Weber (1972). "Bacteriophage Q Replicase Contains the Protein Biosynthesis Elongation Factors EF Tu and EF Ts." Proceedings of the National Academy of Sciences **69**(5): 1313-1317.

- Bollback, J. P. and J. P. Huelsenbeck (2001). "Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family Leviviridae)." J Mol Evol **52**(2): 117-128.

- Borodavka, A., R. Tuma and P. G. Stockley (2012). "Evidence that viral RNAs have evolved for efficient, two-stage packaging." Proc Natl Acad Sci U S A **109**(39): 15769-15774.

- Bottcher, B. and M. Nassal (2018). "Structure of Mutant Hepatitis B Core Protein Capsids with Premature Secretion Phenotype." J Mol Biol **430**(24): 4941-4954.

- Bottcher, B., S. A. Wynne and R. A. Crowther (1997). "Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy." Nature **386**(6620): 88-91.

- Bradley, D. E., J. N. Coetzee, T. Bothma and R. W. Hedges (1981). "Phage t: a group T plasmid-dependent bacteriophage." J Gen Microbiol **126**(2): 397-403.

- Brandariz-Nunez, A., T. Liu, T. Du and A. Evilevitch (2019). "Pressure-driven release of viral genome into a host nucleus is a mechanism leading to herpes infection." Elife **8**.

- Bremer, H. and P. P. Dennis (2008). "Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates." EcoSal Plus **3**(1).

- Brinton, C. C. and H. Beer (1967). "The Interaction of Male Specific Bacteriophage with F Pili." The Molecular Biology of Viruses: 251-289.

- Brinton, C. C., Jr. (1965). "The structure, function, synthesis and genetic control of bacterial pili and a molecular model for DNA and RNA transport in gram negative bacteria." Trans N Y Acad Sci **27**(8): 1003-1054.

- Bundy, B. C. and J. R. Swartz (2011). "Efficient disulfide bond formation in virus-like particles." J Biotechnol **154**(4): 230-239.

- Caldeira, J. C. and D. S. Peabody (2007). "Stability and assembly in vitro of bacteriophage PP7 virus-like particles." J Nanobiotechnology **5**: 10.

- Callanan, J., S. R. Stockdale, A. Shkoporov, L. A. Draper, R. P. Ross and C. Hill (2018). "RNA Phage Biology in a Metagenomic Era." Viruses **10**(7).

- Callanan, J., S. R. Stockdale, A. Shkoporov, L. A. Draper, R. P. Ross and C. Hill (2020). "Expansion of known ssRNA phage genomes: From tens to over a thousand." Science Advances **6**(6).

- Cao, L., R. Zhang, T. Liu, Z. Sun, M. Hu, Y. Sun, L. Cheng, Y. Guo, S. Fu, J. Hu, X. Li, C. Yu, H. Wang, H. Chen, X. Li, E. E. Fry, D. I. Stuart, P. Qian, Z. Lou and Z. Rao (2018). "Seneca Valley virus attachment and uncoating mediated by its receptor anthrax toxin receptor 1." Proc Natl Acad Sci U S A **115**(51): 13087-13092.

- Carey, J., V. Cameron, P. L. de Haseth and O. C. Uhlenbeck (1983). "Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site." Biochemistry **22**(11): 2601-2610.

- Carpenter, J. E., E. P. Henderson and C. Grose (2009). "Enumeration of an extremely high particle-to-PFU ratio for Varicella-zoster virus." J Virol **83**(13): 6917-6921.

- Carroni, M. and H. R. Saibil (2016). "Cryo electron microscopy to determine the structure of macromolecular complexes." Methods **95**: 78-85.

- Caspar, D. L. (1956). "Structure of bushy stunt virus." Nature **177**(4506): 475-476.

- Caspar, D. L. and A. Klug (1962). "Physical principles in the construction of regular viruses." Cold Spring Harb Symp Quant Biol **27**: 1-24.

- Cavalli, L. L., J. Lederberg and E. M. Lederberg (1953). "An infective factor controlling sex compatibility in Bacterium coli." J Gen Microbiol **8**(1): 89-103.

- Chamakura, K. and R. Young (2019). "Phage single-gene lysis: Finding the weak spot in the bacterial cell wall." J Biol Chem **294**(10): 3350-3358.

- Chamakura, K. R., G. B. Edwards and R. Young (2017). "Mutational analysis of the MS2 lysis protein L." Microbiology **163**(7): 961-969.

- Chetverin, A. B. (2018). "Thirty Years of Studies of Qbeta Replicase: What Have We Learned and What Is Yet to Be Learned?" Biochemistry (Mosc) **83**(Suppl 1): S19-S32.

- Cielens, I., V. Ose, I. Petrovskis, A. Strelnikova, R. Renhofa, T. Kozlovska and P. Pumpens (2000). "Mutilation of RNA phage Qβ virus-like particles: from icosahedrons to rods." FEBS Letters **482**(3): 261-264.

- Clarke, M., L. Maddera, R. L. Harris and P. M. Silverman (2008). "F-pili dynamics by live-cell imaging." Proc Natl Acad Sci U S A **105**(46): 17978-17981.

- Coetzee, J. N., D. E. Bradley, J. Fleming, L. du Toit, V. M. Hughes and R. W. Hedges (1985). "Phage pilH alpha: a phage which adsorbs to IncHI and IncHII plasmid-coded pili." J Gen Microbiol **131**(5): 1115-1121.

- Coetzee, J. N., D. E. Bradley and R. W. Hedges (1982). "Phages I alpha and I2-2: IncI plasmid-dependent bacteriophages." J Gen Microbiol **128**(11): 2797-2804.

- Coetzee, J. N., D. E. Bradley, R. W. Hedges, J. Fleming and G. Lecatsas (1983). "Bacteriophage M: an incompatibility group M plasmid-specific phage." J Gen Microbiol **129**(7): 2271-2276.

- Coetzee, J. N., D. E. Bradley, G. Lecatsas, L. du Toit and R. W. Hedges (1985). "Bacteriophage D: an IncD group plasmid-specific phage." J Gen Microbiol **131**(12): 3375-3383.

- Cooper, S. and N. D. Zinder (1963). "The Growth of an RNA Bacteriophage: The Role of Protein Synthesis." Virology **20**: 605-612.

- Costa, T. R. D., A. Ilangovan, M. Ukleja, A. Redzej, J. M. Santini, T. K. Smith, E. H. Egelman and G. Waksman (2016). "Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex." Cell **166**(6): 1436-1444 e1410.

- Crick, F. H. and J. D. Watson (1956). "Structure of small viruses." Nature **177**(4506): 473-475.

- Crowther, R. A., L. A. Amos and J. T. Finch (1975). "Three-dimensional image reconstructions of bacteriophages R17 and f2." Journal of Molecular Biology **98**(3): 631-635.

- Crowther, R. A., L. A. Amos, J. T. Finch, D. J. De Rosier and A. Klug (1970). "Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs." Nature **226**(5244): 421-425.

- Cui, Z., K. V. Gorzelnik, J. Y. Chang, C. Langlais, J. Jakana, R. Young and J. Zhang (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 U S A **114**(44): 11697-11702.

- Dai, W., C. Fu, D. Raytcheva, J. Flanagan, H. A. Khant, X. Liu, R. H. Rochat, C. Haase-Pettingell, J. Piret, S. J. Ludtke, K. Nagayama, M. F. Schmid, J. A. King and W. Chiu (2013). "Visualizing virus assembly intermediates inside marine cyanobacteria." Nature **502**(7473): 707-710.

- Dai, X., Z. Li, M. Lai, S. Shu, Y. Du, Z. H. Zhou and R. Sun (2017). "In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus." Nature **541**(7635): 112-116.

- Danziger, R. E. and W. Paranchych (1970). "Stages in phage R17 infection." Virology **40**(3): 554-564.

- Danziger, R. E. and W. Paranchych (1970). "Stages in phage R17 infection." Virology **40**(3): 547-553.

- Davis, J. E., R. L. Sinsheimer and J. H. Strauss (1961). "Bacteriophage MS2 - Another RNA Phage." Science **134**(348): 1427-&.

- de Martin Garrido, N., M. A. Crone, K. Ramlaul, P. A. Simpson, P. S. Freemont and C. H. S. Aylett (2019). "Bacteriophage MS2 displays unreported capsid variability assembling T = 4 and mixed capsids." Mol Microbiol.

- De Rosier, D. J. and A. Klug (1968). "Reconstruction of three dimensional structures from electron micrographs." Nature **217**(5124): 130-134.

- Dent, K. C., R. Thompson, A. M. Barker, J. A. Hiscox, J. N. Barr, P. G. Stockley and N. A. Ranson (2013). "The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release." Structure **21**(7): 1225-1234.

- Domingo, E., R. A. Flavell and C. Weissmann (1976). "In vitro site-directed mutagenesis: Generation and properties of an infectious extracistronic mutant of bacteriophage Qβ." Gene **1**(1): 3-25.

- Domingo, E., D. Sabo, T. Taniguchi and C. Weissmann (1978). "Nucleotide sequence heterogeneity of an RNA phage population." Cell **13**(4): 735-744.

- Dong, Y., Y. Liu, W. Jiang, T. J. Smith, Z. Xu and M. G. Rossmann (2017). "Antibody-induced uncoating of human rhinovirus B14." Proc Natl Acad Sci U S A **114**(30): 8017-8022.

- Duval, M., A. Korepanov, O. Fuchsbauer, P. Fechter, A. Haller, A. Fabbretti, L. Choulier, R. Micura, B. P. Klaholz, P. Romby, M. Springer and S. Marzi (2013). "Escherichia coli ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation." PLoS Biol **11**(12): e1001731.

- Dykeman, E. C., N. E. Grayson, K. Toropova, N. A. Ranson, P. G. Stockley and R. Twarock (2011). "Simple rules for efficient assembly predict the layout of a packaged viral RNA." J Mol Biol **408**(3): 399-407.

- Dykeman, E. C., P. G. Stockley and R. Twarock (2013). "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.

- Eggen, K. and D. Nathans (1969). "Regulation of protein synthesis directed by coliphage MS2 RNA." Journal of Molecular Biology **39**(2): 293-305.

- Eigen, M., C. K. Biebricher, M. Gebinoga and W. C. Gardiner (1991). "The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage." Biochemistry **30**(46): 11005-11018.

- Engelhardt, D. L. and N. D. Zinder (1964). "Host-dependent mutants of the bacteriophage f2 III. Infective RNA." Virology **23**(4): 582-587.

- Ettayebi, K., S. E. Crawford, K. Murakami, J. R. Broughman, U. Karandikar, V. R. Tenge, F. H. Neill, S. E. Blutt, X. L. Zeng, L. Qu, B. Kou, A. R. Opekun, D. Burrin, D. Y. Graham, S. Ramani, R. L. Atmar and M. K. Estes (2016). "Replication of human noroviruses in stem cell-derived human enteroids." Science **353**(6306): 1387-1393.

- Feary, T. W., E. Fisher, Jr. and T. N. Fisher (1964). "Isolation and Preliminary Characteristics of Three Bacteriophages Associated with a Lysogenic Strain of Pseudomonas Aeruginosa." J Bacteriol **87**: 196-208.

- Feiss, M. and C. E. Catalano (2005). Bacteriophage Lambda Terminase and the Mechanism of Viral DNA Packaging. Viral Genome Packaging Machines: Genetics, Structure, and Mechanism**:** 5-39.

- Fiedler, J. D., C. Higginson, M. L. Hovlid, A. A. Kislukhin, A. Castillejos, F. Manzenrieder, M. G. Campbell, N. R. Voss, C. S. Potter, B. Carragher and M. G. Finn (2012). "Engineered mutations change the structure and stability of a virus-like particle." Biomacromolecules **13**(8): 2339-2348.

- Fiers, W. (1979). Structure and Function of RNA Bacteriophages. Comprehensive Virology Volume 13: Structure and Assembly**:** 69-204.

- Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert and M. Ysebaert (1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene." Nature **260**(5551): 500-507.

- Flavell, R. A., D. L. Sabo, E. F. Bandle and C. Weissmann (1975). "Site-directed mutagenesis: effect of an extracistronic mutation on the in vitro propagation of bacteriophage Qbeta RNA." Proc Natl Acad Sci U S A **72**(1): 367-371.

- Frank, J., W. Goldfarb, D. Eisenberg and T. S. Baker (1978). "Reconstruction of glutamine synthetase using computer averaging." Ultramicroscopy **3**: 283-290.

- Friedman, S. D., F. J. Genthner, J. Gentry, M. D. Sobsey and J. Vinje (2009). "Gene mapping and phylogenetic analysis of the complete genome from 30 single-stranded RNA male-specific coliphages (family Leviviridae)." J Virol **83**(21): 11233-11243.

- Fronzes, R., P. J. Christie and G. Waksman (2009). "The structural biology of type IV secretion systems." Nat Rev Microbiol **7**(10): 703-714.

- Frost, L. S., B. B. Finlay, A. Opgenorth, W. Paranchych and J. S. Lee (1985). "Characterization and sequence analysis of pilin from F-like plasmids." J Bacteriol **164**(3): 1238-1247.

- Frost, L. S. and W. Paranchych (1988). "DNA sequence analysis of point mutations in traA, the F pilin gene, reveal two domains involved in F-specific bacteriophage attachment." Mol Gen Genet **213**(1): 134-139.

- Fukuma, I. and S. S. Cohen (1975). "Polyamines in bacteriophage R17 and its RNA." J Virol **16**(2): 222-227.

- Furuse, K., A. Ando, S. Osawa and I. Watanabe (1981). "Distribution of ribonucleic acid coliphages in raw sewage from treatment plants in Japan." Appl Environ Microbiol **41**(5): 1139-1143.

- Garcia-Villada, L. and J. W. Drake (2013). "Experimental selection reveals a trade-off between fecundity and lifespan in the coliphage Qbeta." Open Biol **3**(6): 130043.

- Garmann, R. F., A. M. Goldfain and V. N. Manoharan (2019). "Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome." Proc Natl Acad Sci U S A **116**(45): 22485-22490.

- Garwes, D., A. Sillero and S. Ochoa (1969). "Virus-specific proteins in Escherichia coli infected with phage Qβ." Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis **186**(1): 166-172.

- Gelbart, W. M. and C. M. Knobler (2009). "Virology. Pressurized viruses." Science **323**(5922): 1682-1683.

- Geraets, J. A., E. C. Dykeman, P. G. Stockley, N. A. Ranson and R. Twarock (2015). "Asymmetric genome organization in an RNA virus revealed via graph-theoretical analysis of tomographic data." PLoS Comput Biol **11**(3): e1004146.

- Golmohammadi, R., K. Fridborg, M. Bundule, K. Valegård and L. Liljas (1996). "The crystal structure of bacteriophage Qβ at 3.5 å resolution." Structure **4**(5): 543-554.

- Golmohammadi, R., K. Valegard, K. Fridborg and L. Liljas (1993). "The refined structure of bacteriophage MS2 at 2.8 A resolution." J Mol Biol **234**(3): 620-639.

- Gorzelnik, K. V., Z. Cui, C. A. Reed, J. Jakana, R. Young and J. Zhang (2016). "Asymmetric cryo-EM structure of the canonical Allolevivirus Qbeta reveals a single maturation protein and the genomic ssRNA in situ." Proc Natl Acad Sci U S A **113**(41): 11519-11524.

- Grange, M., D. Vasishtan and K. Grunewald (2017). "Cellular electron cryo tomography and in situ sub-volume averaging reveal the context of microtubule-based processes." J Struct Biol **197**(2): 181-190.

- Grant, T. and N. Grigorieff (2015). "Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6." Elife **4**: e06980.

- Grigorieff, N. (2013). "Direct detection pays off for electron cryo-microscopy." Elife **2**: e00573.

- Groeneveld, H., K. Thimon and J. van Duin (1995). "Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding?" RNA **1**(1): 79-88.

- Guo, F. and W. Jiang (2014). "Single particle cryo-electron microscopy and 3-D reconstruction of viruses." Methods Mol Biol **1117**: 401-443.

- Gussin, G. N. (1966). "Three complementation groups in bacteriophage R17." Journal of Molecular Biology **21**(3): 435-453.

- Gytz, H., D. Mohr, P. Seweryn, Y. Yoshimura, Z. Kutlubaeva, F. Dolman, B. Chelchessa, A. B. Chetverin, F. A. Mulder, D. E. Brodersen and C. R. Knudsen (2015). "Structural basis for RNA-genome recognition during bacteriophage Qbeta replication." Nucleic Acids Res **43**(22): 10893-10906.

- Harland, J. and S. M. Brown (1998). "HSV Growth, Preparation, and Assay." Methods Mol Med **10**: 1-8.

- Harrison, S. C. (2010). "Virology. Looking inside adenovirus." Science **329**(5995): 1026-1027.

- Harrison, S. C., A. J. Olson, C. E. Schutt, F. K. Winkler and G. Bricogne (1978). "Tomato bushy stunt virus at 2.9 A resolution." Nature **276**(5686): 368-373.

- Haruna, I., K. Nozu, Y. Ohtaka and S. Spiegelman (1963). "An RNA "Replicase" Induced by and Selective for a Viral RNA: Isolation and Properties." Proc Natl Acad Sci U S A **50**: 905-911.

- Haruna, I. and S. Spiegelman (1965). "Autocatalytic synthesis of a viral RNA in vitro." Science **150**(3698): 884-886.

- Hatahet, F., D. Boyd and J. Beckwith (2014). "Disulfide bond formation in prokaryotes: history, diversity and design." Biochim Biophys Acta **1844**(8): 1402-1414.

- Hofstetter, H., H. J. Monstein and C. Weissmann (1974). "The readthrough protein A1 is essential for the formation of viable Qβ particles." Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis **374**(2): 238-251.

- Hohn, T. (1969). "Role of RNA in the assembly process of bacteriophage fr." Journal of Molecular Biology **43**(1): 191-200.

- Holmes, K. C. and R. E. Franklin (1958). "The radial density distribution in some strains of tobacco mosaic virus." Virology **6**(2): 328-336.

- Hu, B., P. Khara and P. J. Christie (2019). "Structural bases for F plasmid conjugation and F pilus biogenesis in Escherichia coli." Proc Natl Acad Sci U S A **116**(28): 14222-14227.

- Hung, P. P. and L. R. Overby (1969). "The reconstitution of infective bacteriophage Q beta." Biochemistry **8**(3): 820-828.

- Inokuchi, Y., A. B. Jacobson, T. Hirose, S. Inayama and A. Hirashima (1988). "Analysis of the complete nucleotide sequence of the group IV RNA coliphage SP." Nucleic Acids Res **16**(13): 6205-6221.

- Inokuchi, Y., R. Takahashi, T. Hirose, S. Inayama, A. B. Jacobson and A. Hirashima (1986). "The complete nucleotide sequence of the group II RNA coliphage GA." J Biochem **99**(4): 1169-1180.

- Jiang, W. and L. Tang (2017). "Atomic cryo-EM structures of viruses." Curr Opin Struct Biol **46**: 122-129.

- Jun, S., D. Ke, K. Debiec, G. Zhao, X. Meng, Z. Ambrose, G. A. Gibson, S. C. Watkins and P. Zhang (2011). "Direct visualization of HIV-1 with correlative live-cell microscopy and cryo-electron tomography." Structure **19**(11): 1573-1581.

- Kamen, R. (1970). "Characterization of the Subunits of Qβ Replicase." Nature **228**(5271): 527-533.

- Kamen, R. (1975). Structure and Function of the Qβ RNA Replicase. RNA Phages. N. D. Zinder. Cold Spring Harbor, Cold Spring Harbor Laboratory Press**:** 203-234.

- Kamen, R., M. Kondo, W. Romer and C. Weissmann (1972). "Reconstitution of Q replicase lacking subunit with protein-synthesis-interference factor i." Eur J Biochem **31**(1): 44-51.

- Kannoly, S., Y. Shao and I. N. Wang (2012). "Rethinking the evolution of single-stranded RNA (ssRNA) bacteriophages based on genomic sequences and characterizations of two R-plasmid-dependent ssRNA phages, C-1 and Hgal1." J Bacteriol **194**(18): 5073-5079.

- Karnik, S. and M. Billeter (1983). "The lysis function of RNA bacteriophage Qbeta is mediated by the maturation (A2) protein." EMBO J. **2**(9): 1521-1526.

- Kashiwagi, A., R. Sugawara, F. Sano Tsushima, T. Kumagai and T. Yomo (2014). "Contribution of silent mutations to thermal adaptation of RNA bacteriophage Qbeta." J Virol **88**(19): 11459-11468.

- Kavita, K., F. de Mets and S. Gottesman (2018). "New aspects of RNA-based regulation by Hfq and its partner sRNAs." Curr Opin Microbiol **42**: 53-61.

- Kelly, J., A. Y. Grosberg and R. Bruinsma (2016). "Sequence Dependence of Viral RNA Encapsidation." J Phys Chem B **120**(26): 6038-6050.

- Kidmose, R. T., N. N. Vasiliev, A. B. Chetverin, G. R. Andersen and C. R. Knudsen (2010). "Structure of the Qbeta replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins." Proc Natl Acad Sci U S A **107**(24): 10884-10889.

- Klovins, J., V. Berzins and J. van Duin (1998). "A long-range interaction in Qbeta RNA that bridges the thousand nucleotides between the M-site and the 3' end is required for replication." RNA **4**(8): 948-957.

- Klovins, J., V. Berzins and J. Van Duin (1998). "A long-range interaction in Qβ RNA that bridges the thousand nucleotides between the M-site and the 3′ end is required for replication." RNA **4**(8): 948-957.

- Klovins, J., G. P. Overbeek, S. H. van den Worm, H. W. Ackermann and J. van Duin (2002). "Nucleotide sequence of a ssRNA phage from Acinetobacter: kinship to coliphages." J Gen Virol **83**(Pt 6): 1523-1533.

- Klovins, J. and J. van Duin (1999). "A long-range pseudoknot in Qbeta RNA is essential for replication." J Mol Biol **294**(4): 875-884.

- Kolakofsky, D. and C. Weissmann (1971). "Possible mechanism for transition of viral RNA from polysome to replication complex." Nat New Biol **231**(19): 42-46.

- Kolakofsky, D. and C. Weissmann (1971). "Qβ replicase as repressor of Qβ RNA-directed protein synthesis." Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis **246**(3): 596-599.

- Kondo, M., R. Gallerani and C. Weissmann (1970). "Subunit Structure of Qβ Replicase." Nature **228**(5271): 525-527.

- Koning, R. I., J. Gomez-Blanco, I. Akopjana, J. Vargas, A. Kazaks, K. Tars, J. M. Carazo and A. J. Koster (2016). "Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ." Nat Commun **7**: 12524.

- Kostyuchenko, V. A., E. X. Lim, S. Zhang, G. Fibriansah, T. S. Ng, J. S. Ooi, J. Shi and S. M. Lok (2016). "Structure of the thermally stable Zika virus." Nature **533**(7603): 425-428.

- Krahn, P. M., R. J. O'Callaghan and W. Paranchych (1972). "Stages in phage R17 infection." Virology **47**(3): 628-637.

- Kramer, F. R. and D. R. Mills (1981). "Secondary structure formation during RNA synthesis." Nucleic Acids Res **9**(19): 5109-5124.

- Krishnamurthy, S. R., A. B. Janowski, G. Zhao, D. Barouch and D. Wang (2016). "Hyperexpansion of RNA Bacteriophage Diversity." PLoS Biol **14**(3): e1002409.

- Krueger, R. G. (1969). "Serological relatedness of the ribonucleic acid-containing coliphages." J Virol **4**(5): 567-573.

- Kuo, C. H., L. Eoyang and J. T. August (1975). Protein Factors Required for the Replication of Phage Qβ RNA In Vitro. RNA Phages. N. D. Zinder. Cold Spring Harbor, Cold Spring Harbor Laboratory Press**:** 259-277.

- Kurland, C. G. (1960). "Molecular characterization of ribonucleic acid from Escherichia coli ribosomes." Journal of Molecular Biology **2**(2): 83-91.

- Kutlubaeva, Z. S., H. V. Chetverina and A. B. Chetverin (2017). "The Contribution of Ribosomal Protein S1 to the Structure and Function of Qβ Replicase." Acta Naturae **9**(4): 24-30.

- Laanto, E., S. Mantynen, L. De Colibus, J. Marjakangas, A. Gillum, D. I. Stuart, J. J. Ravantti, J. T. Huiskonen and L. R. Sundberg (2017). "Virus found in a boreal lake links ssDNA and dsDNA viruses." Proc Natl Acad Sci U S A **114**(31): 8378-8383.

- Landers, T. A., T. Blumenthal and K. Weber (1974). "Function and structure in ribonucleic acid phage Q beta ribonucleic acid replicase. The roles of the different subunits in transcription of synthetic templates." J Biol Chem **249**(18): 5801-5808.

- Lang, S., P. C. Kirchberger, C. J. Gruber, A. Redzej, S. Raffl, G. Zellnig, K. Zangger and E. L. Zechner (2011). "An activation domain of plasmid R1 TraI protein delineates stages of gene transfer initiation." Mol Microbiol **82**(5): 1071-1085.

- Lerner, T. J. and N. D. Zinder (1977). "Discontinuous release of phage f2." Virology **79**(1): 236-238.

- Lim, F., M. Spingola and D. S. Peabody (1996). "The RNA-binding site of bacteriophage Qbeta coat protein." J Biol Chem **271**(50): 31839-31845.

- Liu, Y., J. Sheng, A. L. W. van Vliet, G. Buda, F. J. M. van Kuppeveld and M. G. Rossmann (2018). "Molecular basis for the acid-initiated uncoating of human enterovirus D68." Proc Natl Acad Sci U S A **115**(52): E12209-E12217.

- Lodish, H. F. (1971). "Thermal melting of bacteriophage f2 RNA and initiation of synthesis of the maturation protein." Journal of Molecular Biology **56**(3): 627-632.

- Lodish, H. F., K. Horiuchi and N. D. Zinder (1965). "Mutants of the bacteriophage f2." Virology **27**(2): 139-155.

- Lodish, H. F. and N. D. Zinder (1966). "Replication of the RNA of Bacteriophage f2." Science **152**(3720): 372-377.

- Loeb, T. (1960). "Isolation of a bacteriophage specific for the F plus and Hfr mating types of Escherichia coli K-12." Science **131**(3404): 932-933.

- Loeb, T. and N. D. Zinder (1961). "A bacteriophage containing RNA." Proc Natl Acad Sci U S A **47**: 282-289.

- Low, H. H., F. Gubellini, A. Rivera-Calzada, N. Braun, S. Connery, A. Dujeancourt, F. Lu, A. Redzej, R. Fronzes, E. V. Orlova and G. Waksman (2014). "Structure of a type IV secretion system." Nature **508**(7497): 550-553.

- Lyumkis, D. (2019). "Challenges and opportunities in cryo-EM single-particle analysis." J Biol Chem **294**(13): 5181-5197.

- Manchak, J., K. G. Anthony and L. S. Frost (2002). "Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer." Mol Microbiol **43**(1): 195-205.

- Maneewannakul, S., K. Maneewannakul and K. Ippen-Ihler (1992). "Characterization, localization, and sequence of F transfer region products: the pilus assembly gene product TraW and a new product, TrbI." J Bacteriol **174**(17): 5567-5574.

- Marvin, D. A. and H. Hoffmann-Berling (1963). "Physical and chemical properties of two new small bacteriophages." Nature **197**: 517-518.

- Marvin, D. A. and B. Hohn (1969). "Filamentous bacterial viruses." Bacteriol Rev. **33**(2): 172-209.

- Mekler, P. (1981). Determination of Nucleotide Sequences of the Bacteriophage Qβ Genome: Organization and Evolution of an RNA Virus. Ph.D., Universitat Zurich.

- Meng, R., M. Jiang, Z. Cui, J. Y. Chang, K. Yang, J. Jakana, X. Yu, Z. Wang, B. Hu and J. Zhang (2019). "Structural basis for the adsorption of a single-stranded RNA bacteriophage." Nat Commun **10**(1): 3130.

- Meyer, F., H. Weber and C. Weissmann (1981). "Interactions of Qβ replicase with Qβ RNA." Journal of Molecular Biology **153**(3): 631-660.

- Miki, T., T. Horiuchi and N. S. Willetts (1978). "Identification and characterization of four new tra cistrons on the E. coli K12 sex factor F." Plasmid **1**(3): 316-323.

- Min Jou, W., G. Haegeman, M. Ysebaert and W. Fiers (1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein." Nature **237**(5350): 82-88.

- Min Jou, W., A. Raeymaekers and W. Fiers (1979). "Crystallization of bacteriophage MS2." Eur J Biochem **102**(2): 589-594.

- Miranda, G., D. Schuppli, I. Barrera, C. Hausherr, J. M. Sogo and H. Weber (1997). "Recognition of bacteriophage Qbeta plus strand RNA as a template by Qbeta replicase: role of RNA interactions mediated by ribosomal proteins S1 and host factor." J Mol Biol **267**(5): 1089-1103.

- Miyakawa, K., A. Fukuda, Y. Okada, K. Furuse and I. Watanabe (1976). "Isolation and characterization of RNA phages for Caulobacter crescentus." Virology **73**(2): 461-467.

- Miyake, T., I. Haruna, T. Shiba, Y. H. Itoh, K. Yamane and I. Watanabe (1971). "Grouping of RNA Phages Based on the Template Specificity of Their RNA Replicases." Proceedings of the National Academy of Sciences **68**(9): 2022-2024.

- Moce-Llivina, L., M. Muniesa, H. Pimenta-Vale, F. Lucena and J. Jofre (2003). "Survival of bacterial indicator species and bacteriophages after thermal treatment of sludge and sewage." Appl Environ Microbiol **69**(3): 1452-1456.

- Morton, V. L., E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock and P. G. Stockley (2010). "The impact of viral RNA on assembly pathway selection." J Mol Biol **401**(2): 298-308.

- Moustafa, I. M., V. K. Korboukh, J. J. Arnold, E. D. Smidansky, L. L. Marcotte, D. W. Gohara, X. Yang, M. A. Sanchez-Farran, D. Filman, J. K. Maranas, D. D. Boehr, J. M. Hogle, C. M. Colina and C. E. Cameron (2014). "Structural dynamics as a contributor to error-prone replication by an RNA-dependent RNA polymerase." J Biol Chem **289**(52): 36229-36248.

- Ning, X., D. Nguyen, L. Mentzer, C. Adams, H. Lee, R. Ashley, S. Hafenstein and J. Hu (2011). "Secretion of genome-free hepatitis B virus--single strand blocking model for virion morphogenesis of para-retrovirus." PLoS Pathog **7**(9): e1002255.

- Nishihara, T., H. Morisawa, N. Ohta, J. F. Atkins and Y. Nishimura (2004). "A cryptic lysis gene near the start of the Qbeta replicase gene in the +1 frame." Genes Cells **9**(10): 877-889.

- Novotny, C. P. and P. Fives-Taylor (1974). "Retraction of F pili." J Bacteriol. **117**(3): 1306-1311.

- Nuttall, D., D. Maker and E. Colleran (1987). "A method for the direct isolation of IncH plasmid-dependent bacteriophages." Letters in Applied Microbiology **5**(2): 37-40.

- Olsen, R. H. and D. D. Thomas (1973). "Characteristics and purification of PRR1, an RNA phage specific for the broad host range Pseudomonas R1822 drug resistance plasmid." J Virol **12**(6): 1560-1567.

- Olsthoorn, R. and J. van Duin (2011). Bacteriophages with ssRNA. eLS.

- Olsthoorn, R. C. and J. van Duin (1996). "Random removal of inserts from an RNA genome: selection against single-stranded RNA." J Virol **70**(2): 729-736.

- Osawa, S., K. Furuse, M. S. Choi, A. Ando, T. Sakurai and I. Watanabe (1981). "Distribution of ribonucleic acid coliphages in Korea." Appl Environ Microbiol **41**(4): 909-911.

- Osawa, S., K. Furuse and I. Watanabe (1981). "Distribution of ribonucleic acid coliphages in animals." Appl Environ Microbiol **41**(1): 164-168.

- Overby, L. R., G. H. Barlow, R. H. Doi, M. Jacob and S. Spiegelman (1966). "Comparison of two serologically distinct ribonucleic acid bacteriophages. II. Properties of the nucleic acids and coat proteins." J Bacteriol **92**(3): 739-745.

- Paranchych, W. (1966). "Stages in phage R17 infection: The role of divalent cations." Virology **28**(1): 90-99.

- Paranchych, W. (1975). Attachment, Ejection and Penetration Stages of the RNA Phage Infectious Process. RNA Phages. N. D. Zinder. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press**:** pp. 85–111.

- Paranchych, W., S. K. Ainsworth, A. J. Dick and P. M. Krahn (1971). "Stages in phage R17 infection." Virology **45**(3): 615-628.

- Paranchych, W. and A. F. Graham (1962). "Isolation and properties of an RNA-containing bacteriophage." J Cell Comp Physiol **60**: 199-208.

- Paranchych, W., P. M. Krahn and R. D. Bradley (1970). "Stages in phage R17 infection." Virology **41**(3): 465-473.

- Parent, K. N., J. R. Schrad and G. Cingolani (2018). "Breaking Symmetry in Viral Icosahedral Capsids as Seen through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy." Viruses **10**(2).

- Pasloske, B. L., C. R. Walkerpeach, R. D. Obermoeller, M. Winkler and D. B. DuBois (1998). "Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards." J Clin Microbiol **36**(12): 3590-3594.

- Passmore, L. A. and C. J. Russo (2016). "Specimen Preparation for High-Resolution Cryo-EM." Methods Enzymol **579**: 51-86.

- Peabody, D. S. (2003). "A Viral Platform for Chemical Modification and Multivalent Display." J Nanobiotechnology **1**(1): 5.

- Pickett, G. G. and D. S. Peabody (1993). "Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein." Nucleic Acids Res **21**(19): 4621-4626.

- Pierrel, J. (2012). "An RNA Phage Lab: MS2 in Walter Fiers' laboratory of molecular biology in Ghent, from genetic code to gene and genome, 1963-1976." J Hist Biol **45**(1): 109-138.

- Poot, R. A., N. V. Tsareva, I. V. Boni and J. van Duin (1997). "RNA folding kinetics regulates translation of phage MS2 maturation gene." Proceedings of the National Academy of Sciences **94**(19): 10110-10115.

- Pumpens, P., R. Renhofa, A. Dishlers, T. Kozlovska, V. Ose, P. Pushko, K. Tars, E. Grens and M. F. Bachmann (2016). "The True Story and Advantages of RNA Phage Capsids as Nanotools." Intervirology **59**(2): 74-110.

- Reed, C. A., C. Langlais, V. Kuznetsov and R. Young (2012). "Inhibitory mechanism of the Qbeta lysis protein A2." Mol Microbiol **86**(4): 836-844.

- Reed, C. A., C. Langlais, I. N. Wang and R. Young (2013). "A(2) expression and assembly regulates lysis in Qbeta infections." Microbiology **159**(Pt 3): 507-514.

- Rietsch, A. and J. Beckwith (1998). "The genetics of disulfide bond metabolism." Annu Rev Genet **32**: 163-184.

- Roberts, J. W. and J. E. Steitz (1967). "The reconstitution of infective bacteriophage R17." Proceedings of the National Academy of Sciences **58**(4): 1416-1421.

- Robertson, H., R. E. Webster and N. D. Zinder (1968). "Bacteriophage coat protein as repressor." Nature **218**(5141): 533-536.

- Robertson, H. D. and H. F. Lodish (1970). "Messenger characteristics of nascent bacteriophage RNA." Proc Natl Acad Sci U S A **67**(2): 710-716.

- Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." J Struct Biol **192**(2): 216-221.

- Rolfsson, O., S. Middleton, I. W. Manfield, S. J. White, B. Fan, R. Vaughan, N. A. Ranson, E. Dykeman, R. Twarock, J. Ford, C. C. Kao and P. G. Stockley (2016). "Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2." J Mol Biol **428**(2 Pt B): 431-448.

- Rolfsson, O., K. Toropova, N. A. Ranson and P. G. Stockley (2010). "Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid." J Mol Biol **401**(2): 309-322.

- Rossmann, M. G. (2013). "Structure of viruses: a short history." Q Rev Biophys **46**(2): 133-180.

- Rumnieks, J. and K. Tars (2011). "Crystal structure of the read-through domain from bacteriophage Qbeta A1 protein." Protein Sci **20**(10): 1707-1712.

- Rumnieks, J. and K. Tars (2012). "Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M." BMC Microbiol **12**: 277.

- Rumnieks, J. and K. Tars (2017). "Crystal Structure of the Maturation Protein from Bacteriophage Qbeta." J Mol Biol **429**(5): 688-696.

- Rumnieks, J. and K. Tars (2018). "Protein-RNA Interactions in the Single-Stranded RNA Bacteriophages." Subcell Biochem **88**: 281-303.

- Sabo, D. L., E. Domingo, E. F. Bandle, R. A. Flavell and C. Weissmann (1977). "A guanosine to adenosine transition in the 3′ terminal extracistronic region of bacteriophage Qβ RNA leading to loss of infectivity." Journal of Molecular Biology **112**(2): 235-252.

- Sakurai, T., T. Miyake, T. Shiba and I. Watanabe (1968). "Isolation of a possible fourth group of RNA phage." Jpn J Microbiol **12**(4): 544-546.

- Scheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." J Struct Biol **180**(3): 519-530.

- Schmidt, B. F., B. Berkhout, G. P. Overbeek, A. van Strien and J. van Duin (1987). "Determination of the RNA secondary structure that regulates lysis gene

expression in bacteriophage MS2." Journal of Molecular Biology **195**(3): 505-516.

- Schmidt, J. M. (1966). "Observations on the adsorption of Caulobacter bacteriophages containing ribonucleic acid." J Gen Microbiol **45**(2): 347-353.

- Schmidt, J. M. and R. Y. Stanier (1965). "Isolation and Characterization of Bacteriophages Active against Stalked Bacteria." J Gen Microbiol **39**: 95-107.

- Schneemann, A. (2006). "The structural and functional role of RNA in icosahedral virus assembly." Annu Rev Microbiol **60**: 51-67.

- Schuppli, D., G. Miranda, S. Qiu and H. Weber (1998). "A branched stem-loop structure in the M-site of bacteriophage Qbeta RNA is important for template recognition by Qbeta replicase holoenzyme." J Mol Biol **283**(3): 585-593.

- Schwartz, F. M. and N. D. Zinder (1963). "Crystalline aggregates in bacterial cells infected with the RNA bacteriophage f2." Virology **21**(2): 276-278.

- Senear, A. W. and J. A. Steitz (1976). "Site-specific interaction of Qbeta host factor and ribosomal protein S1 with Qbeta and R17 bacteriophage RNAs." J Biol Chem **251**(7): 1902-1912.

- Shapiro, L. and I. Bendis (1975). RNA Phages of Bacteria Other Than E. coli. RNA Phages. N. D. Zinder. Cold Spring Harbor, Cold Spring Harbor Laboratory Press**:** 397-410.

- Shiba, T. and T. Miyake (1975). "New type of infectious complex of E. coli RNA phage." Nature **254**(5496): 157-158.

- Shiba, T. and Y. Suzuki (1981). "Localization of A protein in the RNA-A protein complex of RNA phage MS2." Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis **654**(2): 249-255.

- Si, Z., J. Zhang, S. Shivakoti, I. Atanasov, C. L. Tao, W. H. Hui, K. Zhou, X. Yu, W. Li, M. Luo, G. Q. Bi and Z. H. Zhou (2018). "Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate." PLoS Pathog **14**(12): e1007452.

- Silverman, P. M. and R. C. Valentine (1969). "The RNA injection step of bacteriophage f2 infection." J Gen Virol **4**(1): 111-124.

- Sirgel, F. A., J. N. Coetzee, R. W. Hedges and G. Lecatsas (1981). "Phage C-1: an IncC group; plasmid-specific phage." J Gen Microbiol **122**(1): 155-160.

- Sirohi, D., Z. Chen, L. Sun, T. Klose, T. C. Pierson, M. G. Rossmann and R. J. Kuhn (2016). "The 3.8 A resolution cryo-EM structure of Zika virus." Science **352**(6284): 467-470.

- Skripkin, E. A., M. R. Adhin, M. H. de Smit and J. van Duin (1990). "Secondary structure of the central region of bacteriophage MS2 RNA." Journal of Molecular Biology **211**(2): 447-463.

- Smirnova, G., N. Muzyka and O. Oktyabrsky (2012). "Transmembrane glutathione cycling in growing Escherichia coli cells." Microbiol Res **167**(3): 166-172.

- Smit, M. H. and J. van Duin (1993). "Translational initiation at the coat-protein gene of phage MS2: native upstream RNA relieves inhibition by local secondary structure." Molecular Microbiology **9**(5): 1079-1088.

- Stagg, S. M., G. C. Lander, J. Pulokas, D. Fellmann, A. Cheng, J. D. Quispe, S. P. Mallick, R. M. Avila, B. Carragher and C. S. Potter (2006). "Automated cryoEM data acquisition and analysis of 284742 particles of GroEL." J Struct Biol **155**(3): 470-481.

- Staples, D. H., J. Hindley, M. A. Billeter and C. Weissmann (1971). "Localization of Q-beta maturation cistron ribosome binding site." Nat New Biol **234**(50): 202-204.

- Stass, R., S. L. Ilca and J. T. Huiskonen (2018). "Beyond structures of highly symmetric purified viral capsids by cryo-EM." Curr Opin Struct Biol **52**: 25-31.

- Steitz, J. A. (1968). "Identification of the A protein as a structural component of bacteriophage R17." Journal of Molecular Biology **33**(3): 923-936.

- Steitz, J. A. (1968). "Isolation of the A protein from bacteriophage R17." Journal of Molecular Biology **33**(3): 937-945.

- Stewart, E. J., F. Aslund and J. Beckwith (1998). "Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins." EMBO J **17**(19): 5543-5550.

- Stockley, P. G., N. A. Ranson and R. Twarock (2013). "A new paradigm for the roles of the genome in ssRNA viruses." Future Virology **8**(6): 531-543.

- Stockley, P. G., O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans and A. E. Ashcroft (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.

- Stockley, P. G., S. J. White, E. Dykeman, I. Manfield, O. Rolfsson, N. Patel, R. Bingham, A. Barker, E. Wroblewski, R. Chandler-Bostock, E. U. Weiss, N. A. Ranson, R. Tuma and R. Twarock (2016). "Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid." Bacteriophage **6**(1): e1157666.

- Strauss, E. G. and P. Kaesberg (1970). "Acrylamide gel electrophoresis of bacteriophage Qβ: Electrophoresis of the intact virions and of the viral proteins." Virology **42**(2): 437-452.

- Su, Q., D. Schuppli, T. Tsui Hc, M. E. Winkler and H. Weber (1997). "Strongly reduced phage Qbeta replication, but normal phage MS2 replication in an Escherichia coli K12 mutant with inactivated Qbeta host factor (hfq) gene." Virology **227**(1): 211-214.

- Takamatsu, H. and K. Iso (1982). "Chemical evidence for the capsomeric structure of phage q beta." Nature **298**(5877): 819-824.

- Takeshita, D. and K. Tomita (2010). "Assembly of Q{beta} viral RNA polymerase with host translational elongation factors EF-Tu and -Ts." Proc Natl Acad Sci U S A **107**(36): 15733-15738.

- Takeshita, D., S. Yamashita and K. Tomita (2012). "Mechanism for template-independent terminal adenylation activity of Qbeta replicase." Structure **20**(10): 1661-1669.

- Takeshita, D., S. Yamashita and K. Tomita (2014). "Molecular insights into replication initiation by Qbeta replicase using ribosomal protein S1." Nucleic Acids Res **42**(16): 10809-10822.

- Taketo, A. (1989). "RNA transfection of Escherichia coli by electroporation." Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1007**(1): 127-129.

- Tan, Y. Z., S. Aiyer, M. Mietzsch, J. A. Hull, R. McKenna, J. Grieger, R. J. Samulski, T. S. Baker, M. Agbandje-McKenna and D. Lyumkis (2018). "Sub-2 A Ewald curvature corrected structure of an AAV2 capsid variant." Nat Commun **9**(1): 3628.

- Taniguchi, T., M. Palmieri and C. Weissmann (1978). "QBeta DNA-containing hybrid plasmids giving rise to QBeta phage formation in the bacterial host." Nature **274**(5668): 223-228.

- Tars, K., K. Fridborg, M. Bundule and L. Liljas (2000). "The three-dimensional structure of bacteriophage PP7 from Pseudomonas aeruginosa at 3.7-A resolution." Virology **272**(2): 331-337.

- Tomita, K., N. Ichihashi and T. Yomo (2015). "Replication of partial double-stranded RNAs by Qbeta replicase." Biochem Biophys Res Commun **467**(2): 293-296.

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

- Toropova, K., P. G. Stockley and N. A. Ranson (2011). "Visualising a viral RNA genome poised for release from its receptor complex." J Mol Biol **408**(3): 408-419.

- Trown, P. W. and P. L. Meyer (1973). "Recognition of template RNA by Qβ polymerase: Sequence at the 3′-terminus of Qβ 6S RNA." Archives of Biochemistry and Biophysics **154**(1): 250-262.

- Tsukada, K., M. Okazaki, H. Kita, Y. Inokuchi, I. Urabe and T. Yomo (2009). "Quantitative analysis of the bacteriophage Qbeta infection cycle." Biochim Biophys Acta **1790**(1): 65-70.

- Twarock, R., R. J. Bingham, E. C. Dykeman and P. G. Stockley (2018). "A modelling paradigm for RNA virus assembly." Curr Opin Virol **31**: 74-81.

- Valegard, K., L. Liljas, K. Fridborg and T. Unge (1990). "The three-dimensional structure of the bacterial virus MS2." Nature **345**(6270): 36-41.

- Valegard, K., J. B. Murray, P. G. Stockley, N. J. Stonehouse and L. Liljas (1994). "Crystal structure of an RNA bacteriophage coat protein-operator complex." Nature **371**(6498): 623-626.

- Valentine, R. C. (1958). "Contrast in the electron microscope image." Nature **181**(4612): 832-833.

- Valentine, R. C. and M. Strand (1965). "Complexes of F-Pili and RNA Bacteriophage." Science **148**(3669): 511-513.

- Valentine, R. C. and H. Wedel (1965). "The extracellular stages of RNA bacteriophage infection." Biochemical and Biophysical Research Communications **21**(2): 106-112.

- van Duin, J. (1988). Single-Stranded RNA Bacteriophages. The Bacteriophages. R. Calendar. New York, Plenum Press**:** 117-167.

- van Duin, J. and N. V. Tsareva (2006). Single-stranded RNA phages. The Bacteriophages. R. Calendar. Oxford, Oxford University Press**:** 175-196.

- Vasilyev, N. N., Z. S. Kutlubaeva, V. I. Ugarov, H. V. Chetverina and A. B. Chetverin (2013). "Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qbeta phage replicase." Nat Commun **4**: 1781.

- Vinuela, E., I. D. Algranati and S. Ochoa (1967). "Synthesis of virus-specific proteins in Escherichia coli infected with the RNA bacteriophage MS2." Eur J Biochem **1**(1): 3-11.

- Wahba, A. J., M. J. Miller, A. Niveleau, T. A. Landers, G. G. Carmichael, K. Weber, D. A. Hawley and L. I. Slobin (1974). "Subunit I of Qβ Replicase and 30 S Ribosomal Protein Sl of Escherichia coli EVIDENCE FOR THE IDENTITY OF THE TWO PROTEINS." The Journal of Biological Chemistry **249**: 3314-3316.

- Wang, I.-N., D. E. Dykhuizen and L. B. Slobodkin (1996). "The evolution of phage lysis timing." Evolutionary Ecology **10**(5): 545-558.

- Wang, J. C., S. Mukhopadhyay and A. Zlotnick (2018). "Geometric Defects and Icosahedral Viruses." Viruses **10**(1).

- Watanabe, I., T. Sakurai, K. Furuse and A. Ando (1979). ""Pseudolysogenization" by RNA phage Q beta." Microbiol Immunol **23**(11): 1077-1083.

- Watanabe, T. and M. Okada (1964). "New Type of Sex Factor-Specific Bacteriophage of Escherichia Coli." J Bacteriol **87**: 727-736.

- Weber, H., M. A. Billeter, S. Kahane, C. Weissmann, J. Hindley and A. Porter (1972). "Molecular basis for repressor activity of Q replicase." Nat New Biol **237**(75): 166-170.

- Weber, K. and W. Konigsberg (1975). Proteins of the RNA phage. RNA Phages. N. D. Zinder. Cold Spring Harbor, Cold Spring Harbor Laboratory**:** pp. 51 - 84.

- Weiner, A. M. and K. Weber (1971). "Natural Read-through at the UGA Termination Signal of Qβ Coat Protein Cistron." Nature New Biology **234**(50): 206-209.

- Weissmann, C. (1974). "The making of a phage." FEBS Letters **40**(S1): S3-S9.

- Weissmann, C., M. A. Billeter, H. M. Goodman, J. Hindley and H. Weber (1973). "Structure and function of phage RNA." Annu Rev Biochem **42**: 303-328.

- Weissmann, C., G. Feix and H. Slor (1968). "In vitro Synthesis of Phage RNA: The Nature of the Intermediates." Cold Spring Harbor Symposia on Quantitative Biology **33**(0): 83-100.

- Winter, R. B. and L. Gold (1983). "Overproduction of bacteriophage Qβ maturation (A2) protein leads to cell lysis." Cell **33**(3): 877-885.

- Witherell, G. W., J. M. Gott and O. C. Uhlenbeck (1991). Specific Interaction between RNA Phage Coat Proteins and RNA**:** 185-220.

- Yin, J. and J. Redovich (2018). "Kinetic Modeling of Virus Growth in Cells." Microbiol Mol Biol Rev **82**(2).

- Yoffe, A. M., P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart and A. Ben-Shaul (2008). "Predicting the sizes of large RNA molecules." Proc Natl Acad Sci U S A **105**(42): 16153-16158.

- Yu, G., K. Li, P. Huang, X. Jiang and W. Jiang (2016). "Antibody-Based Affinity Cryoelectron Microscopy at 2.6-A Resolution." Structure **24**(11): 1984-1990.

- Yumura, M., N. Yamamoto, K. Yokoyama, H. Mori, T. Yomo and N. Ichihashi (2017). "Combinatorial selection for replicable RNA by Qbeta replicase while maintaining encoded gene function." PLoS One **12**(3): e0174130.

- Zhang, J., K. V. Gorzelnik, M. F. Schmid and W. Chiu (2016). Single-Particle CryoEM of Macromolecular Complexes. Encyclopedia of Cell Biology**:** 5-13.

- Zhang, X., L. Jin, Q. Fang, W. H. Hui and Z. H. Zhou (2010). "3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry." Cell **141**(3): 472-482.

- Zhang, Y., V. A. Kostyuchenko and M. G. Rossmann (2007). "Structural analysis of viral nucleocapsids by subtraction of partial projections." J Struct Biol **157**(2): 356-364.

- Zhao, L., M. Kopylov, C. S. Potter, B. Carragher and M. G. Finn (2019). "Engineering the PP7 Virus Capsid as a Peptide Display Platform." ACS Nano **13**(4): 4443-4454.

- Zheng, S. Q., E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." Nat Methods **14**(4): 331-332.

- Zhong, Q., A. Carratala, S. Nazarov, R. C. Guerrero-Ferreira, L. Piccinini, V. Bachmann, P. G. Leiman and T. Kohn (2016). "Genetic, Structural, and Phenotypic Properties of MS2 Coliphage with Resistance to ClO2 Disinfection." Environ Sci Technol **50**(24): 13520-13528.

- Zhu, D., X. Wang, Q. Fang, J. L. Van Etten, M. G. Rossmann, Z. Rao and X. Zhang (2018). "Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions." Nat Commun **9**(1): 1552.

- Zinder, N. D. (1965). "RNA phages." Annu Rev Microbiol **19**: 455-472.

- Zinder, N. D. (1980). "Portraits of viruses: RNA phage." Intervirology **13**(5): 257-270.