APPLICATIONS OF CRYO-ELECTRON MICROSCOPY IN THE STUDIES OF VIRUS AND HOST INTERACTIONS

by

Yingyuan Sun

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biological Sciences

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Michael G. Rossmann, Chair
Department of Biological Sciences
Dr. Richard J. Kuhn
Department of Biological Sciences
Dr. Cynthia V. Stauffacher
Department of Biological Sciences
Dr. William A. Cramer
Department of Biological Sciences

Approved by:
Dr. Steven F. Konieczny
Head of the Graduate Program

Dedicated to my families

ACKNOWLEDGMENTS

I would like to express my gratefulness to Prof. Michael G. Rossmann, who has guided me through the last five years with full devotion. It has been a great honor to study under his wings. In him I see a true scientist, who focuses, who is energetic and passionate, who is humble and curious, and most importantly who loves what he is doing. Michael applauses every small step that I have made and works with me to overcome each of the setbacks. He supports me to explore my own ideas and provides directions whenever I need them. Without Michael, I would not be able to accomplish any of the works.

I am privileged to have Prof. William A. Cramer, Prof. Richard J. Kuhn and Prof. Cynthia V. Stauffacher on my advisory committee. I appreciate the helpful discussions and comments. They have taught me how to ask and answer questions in a scientific way. They are always patient and ready to help.

I am very fortunate to have some of the best collaborators. I work on the ΦX174 project with Prof. Bentley Fane at the University of Arizona who is an expert in the field. His humors and supportiveness always encourage me. Aaron Roznowski in his group is a fantastic colleague. Prof. Lois Pollack at Cornell University and her group were extremely helpful during my two visits to Ithaca. They taught me every detail of a SAXS experiment and made a lot of valuable comments on the biological aspect as well. I have also collaborated with Prof. Ann Palmenberg at the University of Wisconsin, Madison on the rhinovirus-C project who is a great virologist. It is always pleasant talking to her and she never hesitated to share with us her knowledge on viruses. Marchel and Kelly from her group worked closely with me and the project would not succeed without them. Dr. Stanislav D. Zakharov from Prof. Cramer's group assisted me in the preparations of liposomes that were used in the study of ΦX174.

I also have some of the greatest lab mates who have helped me in every aspect of my lab life. Thomas Klose is one of the most wonderful persons to work with. He is very patient and very knowledgeable. He spent hours and hours answering my questions, especially during my first two years in the lab. Yue Liu is another student of Michael who has provided tons of help and ideas. He exemplified a great structural biologist. Zhenguo Chen spent hours showing me how to operate the electron microscopes and patiently explained every detail to me. Lei Sun also worked on ΦX174. I followed her and learnt about the project during my rotation in the lab. She is very kind

and wise and taught me a lot. Xinzheng Zhang taught me to freeze my first EM grid. He is an expert on cryo-EM data processing and I benefited a lot in the conversations and discussions with him. Feng Long is experienced in molecular biology and biochemistry. She assisted me in the purification of fabs from antibodies. Sheryl Kelly is always ready to help whenever I need something, from submitting a manuscript to finding a pointer. I also thank Andrei Fokine, Geng Meng, Qianglin Fang, Jenny Wang, Saif Hassan, Geeta Buda, Chunliang Chen, Yangchao Dong and Ju Sheng. Many thanks to every former and current member of the Rossmann Lab, you have made my life here enjoyable and fruitful.

I want to thank Dr. William A. Cramer for discussions on the ΦX174 project, specifically the constitution of liposomes. I also attended two of his lectures on membrane protein complexes which are very helpful. I want to thank Dr. Cynthia Stauffacher for her inspirations since I joined the department. Her lectures on biophysical methods open a new world for me. I want to thank Dr. Michael Gribskov for his lectures on practical biocomputing. He examined our codes line by line and customized the lectures based on student's needs. I want to thank Dr. Daisuke Kihara for his lectures on bioinformatics which help me understand the principles of structure prediction and sequence alignments. I thank Dr. Wen Jiang for his brilliant lectures on electron microscopy. These lectures introduced me to the world of cryo-EM and taught me both mathematical theories and practical computations. I thank Dr. Jeff Bolin for the introductory course to X-ray crystallography.

I feel very lucky to work at Hockmeyer hall surrounded by so many talented and passionate colleagues. I want to thank Steve Wilson, Tim Schmidt, Valorie Bowman and Rebecca Harding who have provided help in many ways to make working in Hockmeyer easier and enjoyable.

I want to express my gratitude to my families. They have provided all the possible support for me to explore my interests and to pursue a career that excites me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Sun, Yingyuan PhD
Institution: Purdue University
Degree Received: December 2018
Title: Applications of Cryo-Electron Microscopy in the Studies of Virus and Host Interactions
Committee Chair: Michael G. Rossmann

Viruses are a group of contagious microbes that have compact structures, containing a nucleic acid core and a protein shell. The replication of viruses requires assistance from hosts which can be almost any cellular organism. Viral infections are often associated with diseases and have been a major threat to the human race. To cope with viral diseases, we need to understand viruses, including their structures, life cycle, pathogenesis and interactions with their hosts. The first structure of a human virus was determined by the Rossmann lab in 1985 using X-ray crystallography.

Thanks to the recent advances in both hardware and software, cryo-electron microscopy (cryo-EM) has emerged as a powerful tool to study virus structures. Cryo-EM allows structural determination for a wide range of specimens to high resolution comparable to what can be achieved by X-ray crystallography. Currently two techniques of cryo-EM are commonly used in structural virology: single particles analysis (SPA) and electron tomography (ET).

Single particle analysis has been used to determine the structures of viruses complexed with host factors in three studies that are to be discussed with more details in chapters 2-4.

The structure of B19 parvovirus complexed with Fabs of a neutralizing human antibody was determined to 3.2 Å resolution. This structure showed that amino acids from three neighboring VP2 proteins form a quaternary structure epitope. In addition, the structure of human rhinovirus-C (RV-C) complexed with its cellular receptor, CDHR3, was determined to 3.9 Å resolution. Despite the low occupancy of the receptors, a "powerful" localized 3D classification procedure helped to select viral particles that had more bound receptors. Furthermore, structures were determined to 10 Å resolution of bacteriophage ΦX174 bound to lipopolysaccharide (LPS) bilayers, before and after genome ejection. These structures showed a series of conformational changes that occurred when a phage penetrated the bacterial membranes. These studies are good examples of applying cryo-EM to investigate virus-host interactions.

However, single particle analysis requires samples to be isolated, homogenous and monodispersed. On the contrary, tomography allows *in situ* studies and is applicable to samples with more flexibility and more heterogeneity. In the case of ΦX174, the structural changes that are involved in the assembly of the H-tube during infection remains a huge mystery. To provide an environment that is more similar to the surface of a bacterial cell, LPS-containing liposomes were mixed with ΦX174 viruses. It was then observed that the ΦX174 particles bound to these liposomes in a very compact manner which was impossible interpret with single particle analysis. Using cryo-ET, 3D volumes of liposome-ΦX174 complexes were reconstructed and structural details were visualized by sub-tomogram classification and averaging.

The emergence of cryo-EM has not only made high-resolution structural studies possible but also broadened the scope of samples with which virologists could work. Moreover, studies on flexible and heterogeneous complexes between viruses and host factors are now possible using either single particle analysis or electron tomography. These techniques will help us to understand virus-host relationships and finally, to develop effective anti-viral therapies.

CHAPTER 1 INTRODUCTION

1.1 The co-evolution of electron microscopy and structural virology

1.1.1 Virus crystallography and its limitations

Viruses are the simplest organisms that can replicate, with the help of a host. A typical virus particle is composed of a nucleic acid core that carries the genetic information and a protein shell that protects the core. However, viruses have different compositions, different sizes and different shapes.

Lacking the necessary organelles and machineries, viruses rely on cellular organisms to accomplish replication. Viruses inhabit host cells and utilize the energy and machineries of the host to replicate genomes and synthesize proteins. As a consequence, viral infections often cause mild or severe diseases in their hosts. Virtually all living organisms, including humans, can serve as hosts for certain viruses. Humans have been fighting viruses for thousands of years, even without knowing the existence of viruses. So far over 200 virus species from 25 families have been identified as human pathogens (1).

The initial knowledge about virus structures came from X-ray analysis and electron microscopy. Watson and Crick had predicted that some of the spherical viruses have a cubic symmetry. Diffraction experiments on Turnip Yellow mosaic virus (2), Tomato Bushy Stunt Virus (3) and poliovirus (4) suggested that these viruses have icosahedral symmetry. These early studies led Aaron Klug and Don Caspar to formulate their theory on the organization of viral shells (5).

The first high resolution structures of intact spherical viruses were solved around 1980 using X-ray crystallography, including tomato bushy stunt virus (TBSV) by Steve Harrison (6) (figure 1.1A), southern bean virus (7) and human rhinovirus 14 (HRV14) by Michael Rossmann (8) (figure 1.1B). These breakthroughs were made possible by advances in the analysis of X-ray diffraction data. These early viral structures not only presented the first atomic view of viral capsid proteins and how these proteins are assembled into an icosahedral shell, but also showed astonishing similarities between plant viruses and animal viruses, particularly in the β-strands that formed the interior scaffold, which was later known as the “jelly-roll fold” (figure 1.1C).

For the past thirty years, X-ray crystallography has been one of the major approaches in the studies of viral structures. The structures of many icosahedral viruses have been determined to atomic resolutions by crystallography (9–14). However, virus crystallography also has its limitations.

Firstly, the presence of high order symmetry would smear structural features that do not obey the same symmetry in the same virus. For example, bacteriophage ΦX174 has a fourth structural protein, the DNA pilot protein H, which was missing in the crystal structure determined using icosahedral symmetry (15). This is because any non-icosahedral protein will have different orientations in crystal packing dominated by the icosahedral capsids.

Secondly, the requirement for well-diffracting crystals limits the scope of samples that are suitable for crystallography. Good crystals diffracting to high resolutions could only be produced for some icosahedral viruses, which are usually small and non-enveloped. However, many of the viruses that cause severe human diseases are pleomorphic and/or lipid enveloped including retroviruses (HIV), orthomyxoviruses (influenza virus), flaviviruses (Dengue virus), alphaviruses (Chikungunya virus), filoviruses (Ebola virus), coronaviruses (SARS-CoV), etc.

Thirdly, crystallography requires crystallization of viral particles or complexes. In practice, crystallography usually only works well with viral particles alone or complexes with small molecules, but not with complexes of viruses and receptors or antibodies. This is because these complexes could not be crystallized because of their large size and limited lattice contacts. These structural studies on more sophisticated viral complexes are important to understand the interaction between a virus and its host.

Lastly, a crystal structure represents only one conformation of the virus which could be different from the native structure due to purification, crystal packing or extensive averaging.

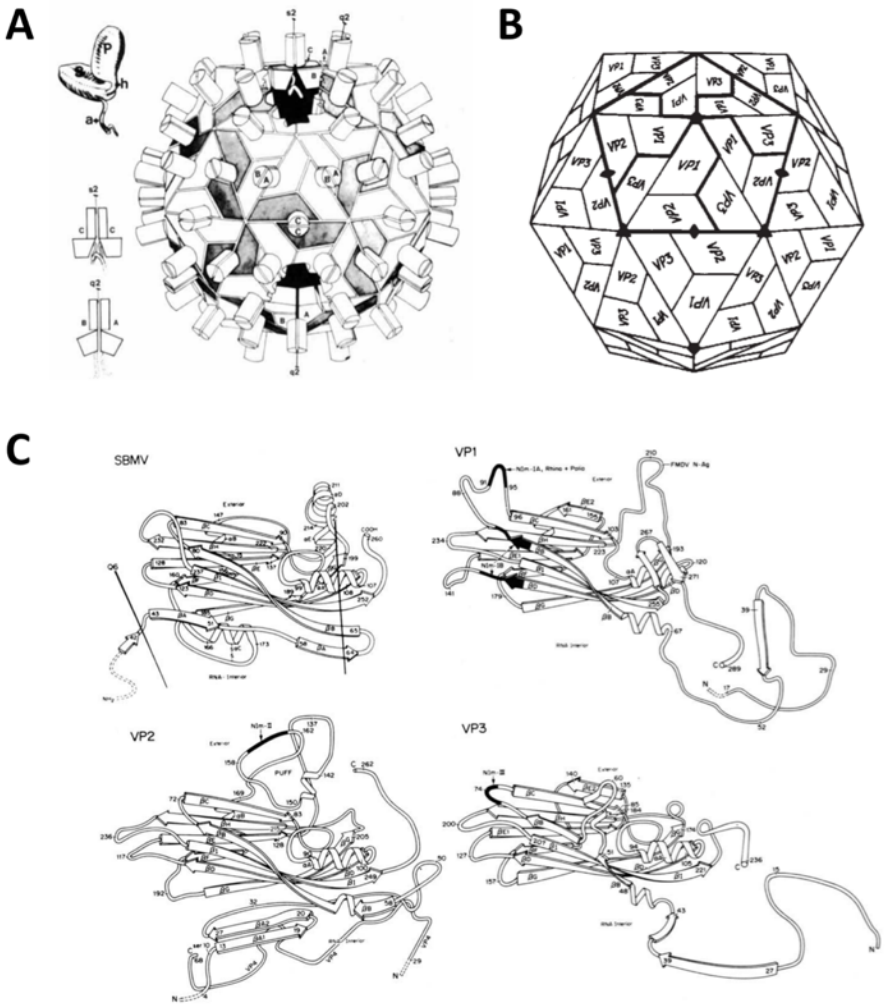

Figure 1.1 The structures of TBSV and HRV14 determined by X-ray crystallography. A. Packing of protein subunits in TBSV. A, B and C denote the three packing environment of the subunit. B. A schematic representation of the icosahedral capsid of HRV14 composing of three proteins. C. Cartoon diagrams showing the structural similarities between a plant virus, SBMV and the three larger capsid proteins of HRV14. (Panel A is reprinted from ref. 6 and Panel B and C are reprinted from ref. 8 with permissions from Springer Nature.)

1.1.2 The early development of electron microscopy and the era of "blobology"

The invention of the electron microscope in the 1920s allowed for the first time biologists to examine their specimens closely at near atomic resolution. Viruses became some of the first samples examined by electron microscopes because of their appropriate size and their relative easy availability. Based on micrographs of viruses, hypotheses were proposed on the assembly and

symmetry of different viruses (5, 16–18). At that time, in David DeRosier's words, virus electron microscopy was "ping-pong ball models of various macromolecular assemblies derived by eyeballing electron micrographs", and he "could not see how one could be sure that a particular model was correct" (19).

Aaron Klug and David DeRosier made the first attempt to reconstruct a 3D model of the tail of T4 bacteriophage in 1968 using negative stained electron micrographs (20) (figure 1.2A). The tail of T4 has a helical symmetry, and one 2D projection of the tail complex was in principle enough to generate a 3D model. In the same study, they also depicted a general process of 3D reconstruction of an object using 2D electron projections based on the central slice theorem (figure 1.2B). In 1970, Klug and colleagues calculated two reconstructions on spherical human wart virus and tomato bushy stunt virus (TBSV) (figure 1.2C) using multiple 2D projections from different orientations (21). A more complete theory was published in 1970 by Klug et al. (22). They also mentioned the issue of defocusing and discussed how to compensate it (23).

In the efforts to improve the quality and resolution of electron microscopic images, scientists realized that radiation damage caused by excessive exposure to electrons was a big problem for biological specimens (24). To minimize radiation damage and to keep the sample hydrated, Glaeser et al came up with the idea of freezing crystals of catalase and successfully recorded diffraction patterns at 4.5 Å (25).

In those early days of electron microscopy, viruses were the perfect sample for technique development for two reasons. First, viruses usually possess high order symmetry, either helical or icosahedral, which makes a small dataset interpretable. The highly symmetric particles also produce stronger signals and have more tolerance toward the noise from electron imaging. Second, viruses are easy to purify and have structural features that are easy to distinguish.

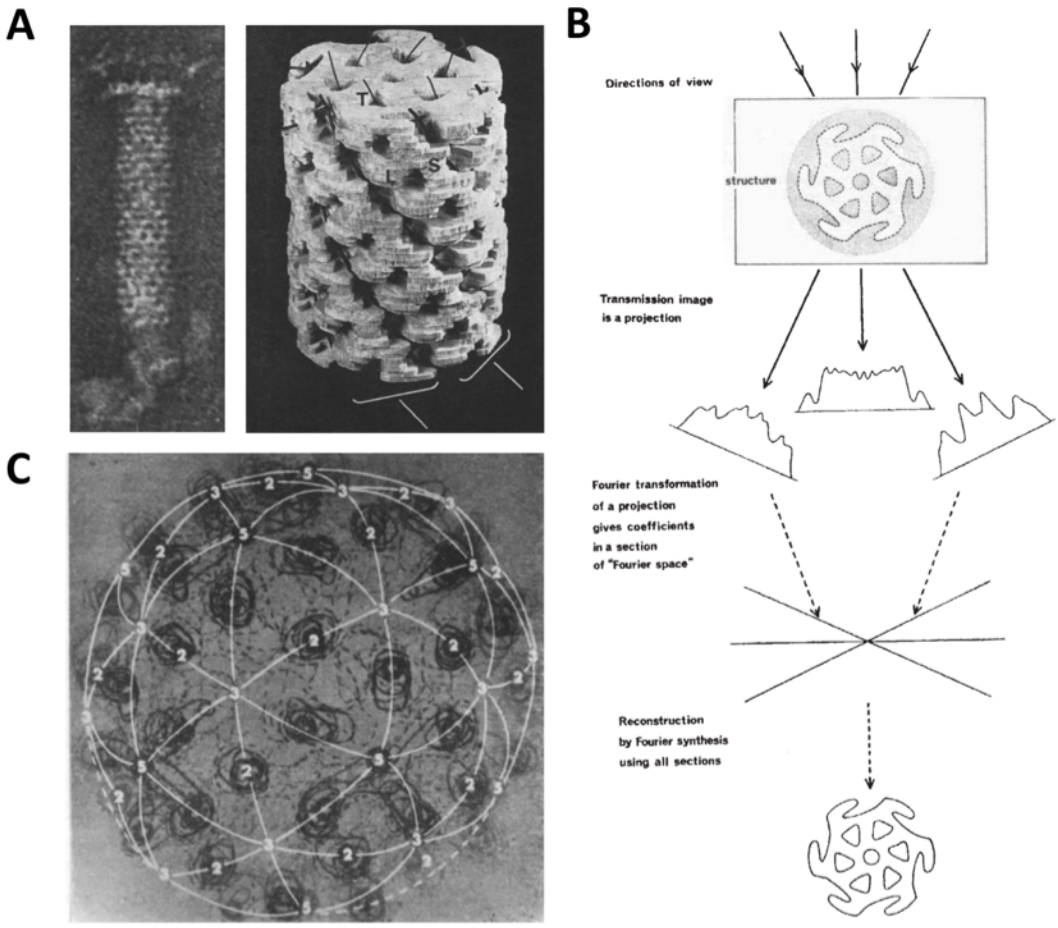

Figure 1.2 First reconstructions of T4 tail complex and TBSV from electron micrographs of negatively stained sample.
A. A negative-staining image of helical T4 tail (left) and a reconstructed model by Klug et al. using Fourier synthesis. B. A schematic diagram of reconstruction using projections from different orientation proposed by Klug et al.. C. A 3D wire model of TBSV reconstructed from electron micrographs. (Panel A and B are reprinted from ref. 20, and panel C is reprinted from ref. 21, both with permissions from Springer Nature.)

In 1975, Richard Henderson and Nigel Unwin, at the Medical Research Council Laboratory of Molecular Biology, used electron diffraction to determine the first structure of a membrane protein, bacteriorhodopsin (26). In their 7 Å resolution map (figure 1.3A), seven transmembrane helices were clearly seen, which was later acknowledged as a signature of G-protein coupled receptors.

At about the same time, Joachim Frank focused on the reconstruction of macromolecules with low or no symmetry in their non-crystalline form (27) and began to develop his SPIDER program (28). Using 2-dimensional autocorrelation algorithm, Frank for the first time produced 2D averaged images of glutamine synthetase (29) and the ribosomal 40S subunit (30). In 1987,

Joachim Frank and colleagues made the first reconstruction of an asymmetric macromolecule, *E. coli* 50S ribosome, using 489 particle projections selected from 5 tilt pair micrographs (figure 1.3B) (31).

Another fundamental advance regarding sample preparation for electron microscopic study was achieved by Jacques Dubochet. A complete procedure including sample freezing, grid transfer and column cooling was described in order to fulfill electron microscopic studies under a temperature of 130K (32, 33). Dubochet also showed that viruses embedded in vitreous ice, without any staining, still yielded sufficient contrast for structural studies (34). A reconstruction of Semliki Forest Virus with a resolution at about 35 Å was later produced from what Dubochet called "cryo-electron micrographs" (35). In the same paper, effects of different defocus values on the contrast (low resolution) and details of the viral particles (high resolution) were illustrated.

Richard Henderson, Joachim Frank and Jacques Dubochet shared the 2017 Nobel prize in chemistry "*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*".

Cryo-EM single particle reconstruction has been successfully used to determine the structures of many icosahedral viruses (36). Initially, the cryo-EM maps usually had a resolution of 10 – 30 Å and appeared as blobs of densities. Despite of the low resolutions, cryo-EM maps provided insights on the organizations of viral capsids and membranes (37, 38), and helped to determine the tertiary and quaternary structures of large protein complexes.

To build "pseudo-atomic" models for large molecular assemblies, it was a common strategy to fit atomic coordinates of the components that compose a macromolecular complex of interest, determined by X-ray crystallography, into a low-resolution cryo-EM map of the whole complex (39).

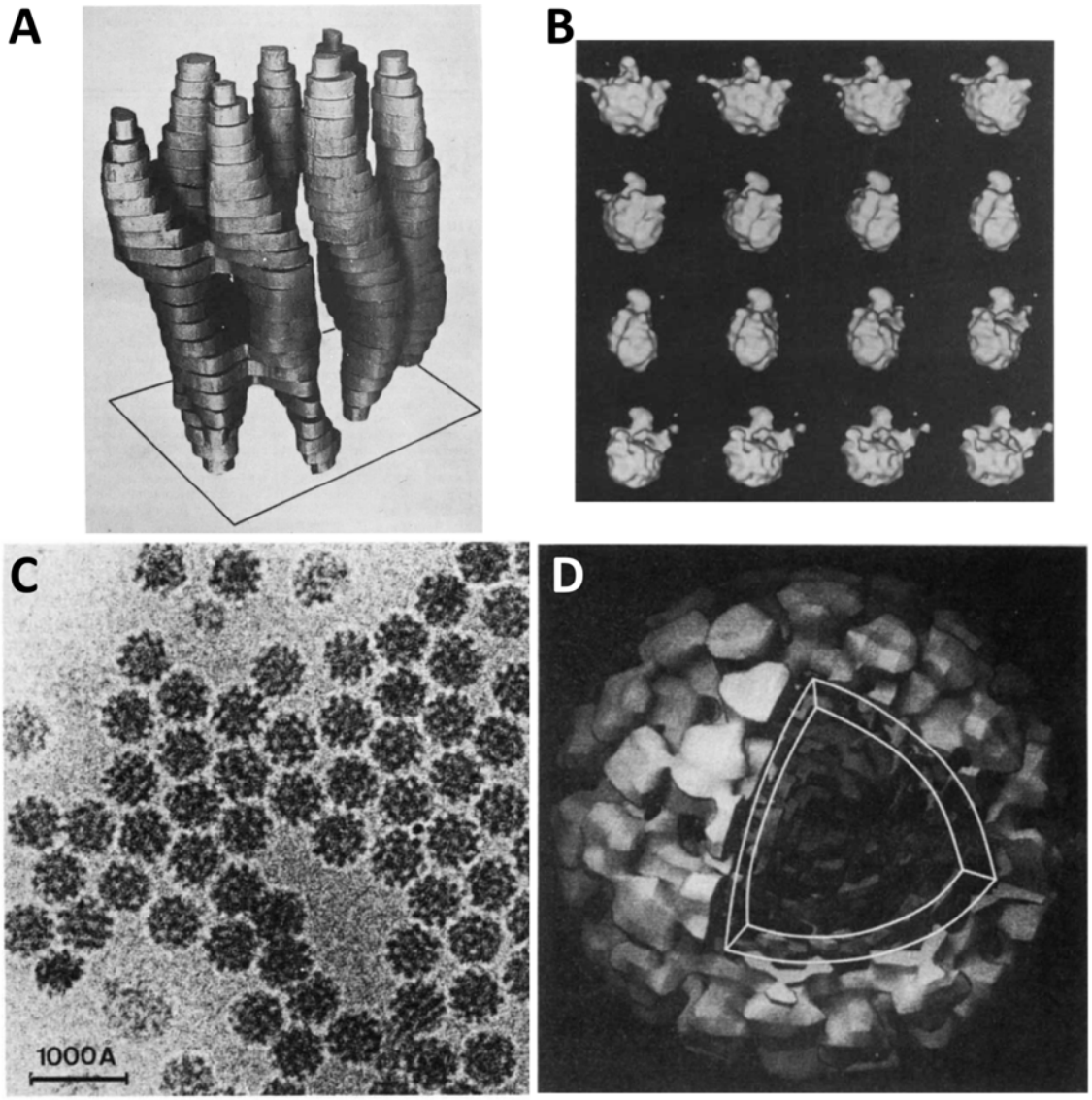

Figure 1.3 Milestones in the history of cryo-EM in 1970s and 1980s.
A. A 7Å model of bacteriorhodopsin acquired in 1974 using electron diffraction glucose-embedded 2D crystals of purple membrane by Richard Henderson and Nigel Unwin. B. A reconstruction of *E. coli* 50S ribosome using random conical tilt method by Frank Joachim et al.. C. A cryo-EM image of semliki forest virus particles embedded in vitreous ice. D. An icosahedral reconstruction of SFV by Dubochet et al.. (Panel A is reprinted from ref. 26, panel C is reprinted from ref. 34 and panel D is reprinted from ref. 35, all with permissions from Springer Nature; panel B is reprinted from ref. 31 with permission from John Wiley and Sons.)

One of the examples was the structural determination of a flavivirus. Flaviviruses are a group of enveloped viruses that are highly pathogenic. The outer shell of a mature flavivirus consists of 180 copies of E (envelope) proteins and M (membrane) proteins, which obey icosahedral symmetry. However, flaviviruses could not be crystallized, potentially due to its flexibility caused

by the membrane envelope. The first structure of a flavivirus, dengue virus type 2 (DENV-2), was obtained by fitting 90 copies of the known structure of a DENV E protein dimer (40) into a low-resolution cryo-EM map of mature DENV-2 using the program EMfit (figure 1.4A&B) (39, 41). In this way, a complete model of the E protein shell was built. There were three monomers (1.5 dimers) in one icosahedral asymmetric unit. It was found that three parallel E dimers were assembled into a rhomboid shaped patch. Thirty copies of this patches associated to cover the surface of the entire DENV-2. These findings were striking at the time because the DENV E protein shell consists of 180 protomers but did not obey the classic quasi T=3 icosahedral symmetry.

A similar case was seen in a study on the spike proteins of human immunodeficiency virus (HIV). HIV is pleomorphic and the core of HIV is surrounded by a layer of lipid membrane. The trimeric protein spikes were found to be sparsely distributed on the membrane surface and are crucial in viral fusion with host membranes. A HIV trimer consists of three gp41-gp120 heterodimers. One of the earliest structures of intact HIV trimer was obtained by fitting a crystal structure of monomeric gp120 (42) into a map of the trimer determined by cryo-EM single particle reconstruction (figure 1.4C) (43).

The first single particle reconstruction that reached “near-atomic” resolution was achieved for rotavirus in 2008 (44) by Nikolaus Grigorieff and colleagues. The 3.8 Å map of icosahedral rotavirus inner capsid allowed *de novo* model building of most of the amino acids. This achievement was a result of non-icosahedral averaging of 13 VP6 trimers within one asymmetric unit after icosahedral averaging and a careful particle screening using a correlation coefficient cut-off. Between 2008 and 2013, six more single particle reconstructions at resolutions better than 4 Å were achieved (45–47). It is important to note that all of these high-resolution results were achieved on icosahedral viruses using electron microscopes operating at 300 kV, with around 250 thousand to 10 million equivalent subunits for structural averaging. These successes, as predicted in 1995 (48), confirmed the potential of cryo-EM to determine structures at resolutions that are comparable to X-ray crystallography.

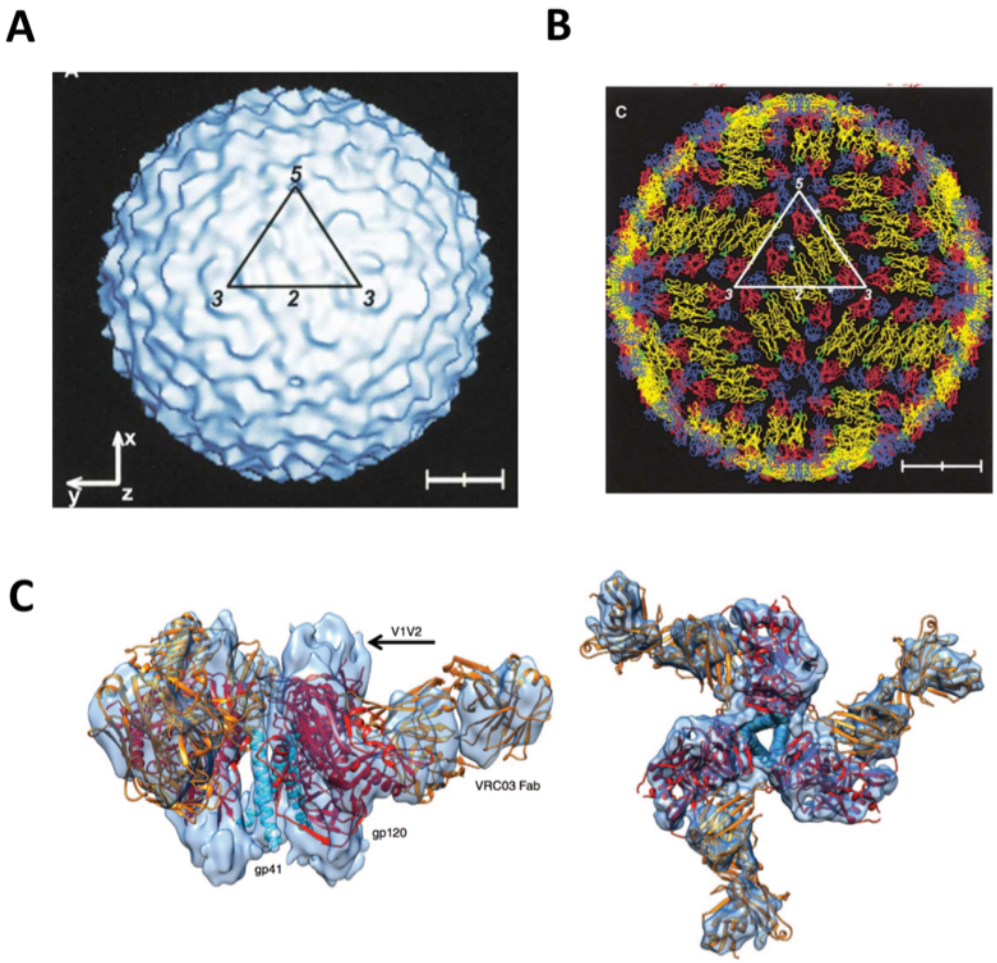

Figure 1.4 The combination of cryo-EM and X-ray crystallography help modeling structures of biological assemblies that cannot be crystallized.
A. A 24 Å map of mature Dengue virus subtype 2 (DENV-2) rendered as surface representation. B. A full assembly of the DENV-2 envelope proteins acquired by fitting the structure of E protein dimers into the map showed in panel A. C. Structures of HIV gp120 and a Fab fitted into a cryo-EM map of soluble HIV trimer at a resolution of 9 Å. (Panel A and B are reprinted from ref. 41 with permission from Elsevier; panel C is reprinted from ref. 43 with permission from Springer Nature.)

Another import progress was related to the validation of cryo-EM maps. It became widely accepted that a "gold standard" criterion proposed by Richard Henderson should be used to measure the resolution of a cryo-EM reconstruction (49). First, a dataset should be randomly divided into two halves with equivalent amounts of data. Date processing should then be performed for these two subsets independently, from building initial models to the iterative angular refinements. The accepted criterion for measuring the resolution is taken to be where the Fourier shell coefficient between the two maps was below 0.143, and corresponds to where the signal of

either half map becomes less than twice the noise. At the same time, the quality of the raw electron micrographs, 2D averages and the refined parameters for the projections need to be closely examined as well. Most importantly, the structure features present in a cryo-EM map should correspond with the nominal resolution. After all, resolution is just a number whereas a justifiable interpretation of a cryo-EM map is what matters.

1.1.3 The "resolution revolution"

From 2013, cryo-EM and structural biology have experienced a revolution marked by the structural determination of a transient receptor potential channel (TRPV1) to a resolution at 3.4 Å (figure 1.5A) (50, 51). The reconstruction of TRPV1 assumed a 4-fold rotational symmetry and used a total of 130,000 particles. This work for the first time showed that such high resolution could be achieved on a membrane protein with low-order symmetry by cryo-EM single particle reconstruction.

Two breakthroughs stood behind the results. The first one is a revolutionary direct electron detector, which is sensitive enough to enable high quality data collection in the form of movies. The second one is a motion-correction algorithm built based on data from the direct electron detector which calculates and compensates the relative beam-induced motion between frames of one "movie-like" image (52). The length of the exposure time for a single frame is usually between 0.05 and 0.20 seconds.

Electron micrographs of biological samples typically have low signal-to-noise ratios, because the total dose that can be applied is largely limited by the sample's susceptibility to radiation damage. Therefore, the detection of signals can substantially affect the quality of the micrographs. The already weak signal at high resolutions can be easily lost as a result of the error introduced in the detection and read-out processes. The new direct electron detectors are beneficial in two ways. First, a high Detective Quantum Efficiency (DQE) boosts the signal-to-noise ratio at all resolutions (53). The DQE of the K2 summit detector using "counting mode" is close to 0.8 at low resolutions and 0.2 at resolutions close to Nyquist (Figure 1.5, panel B). Second, the sensitivity to single electron events allows the recording of short frames at frequencies up to 1000Hz (54). These frames make it possible to conduct image processing like dose-weighting and correction of beam-induced motions (figure 1.5C).

The ideas of direct electron detector and motion correction were not new to the field. Early models of direct electron detectors had been proposed but had limitations (55, 56). A similar motion correction algorithm was published by Grigorieff et al. in 2012 (57) for a DE-12 detector made by Direct Electron Inc. It was shown that beam-induced motion existed not only in each frame as a whole, but also in each particle in the forms of rotation and translation. An ideal motion correction would be based on each particle.

Another factor that contributed to the rise of cryo-EM is automated data collection (58, 59). A single particle reconstruction requires a lot of data, especially when there is no symmetry in the sample. In the past, manual data collection lasting many days was exhausting and error-prone. Robust software that enables fully automated, continuous data collection without any compromise in image quality has greatly improved the efficiency of data collection. Currently there are three popular software packages to fulfill automated image acquisition: Legion developed by the Scripps Research Institute, SerialEM by the University of Colorado and EPU by FEI.

Structural virology has also experienced a big leap thanks to technique advances. In 2012, only 2 completely new structures that had resolutions better than 4 Å were deposited with the EMDB, whereas 38 were deposited in 2017. Moreover, these recently high-resolution results include not only icosahedral viruses alone, but also complexes with antibodies and receptors of their hosts. For example, in 2014 Plevka et al. determined three structures of EV71-Fab complexes at resolutions ranging from 9 to 16 Å (60). A similar study on another picornavirus, HRV16 complexed with Fabs, achieved 2.3 Å resolution with a K2 detector and powerful processing algorithms (61).

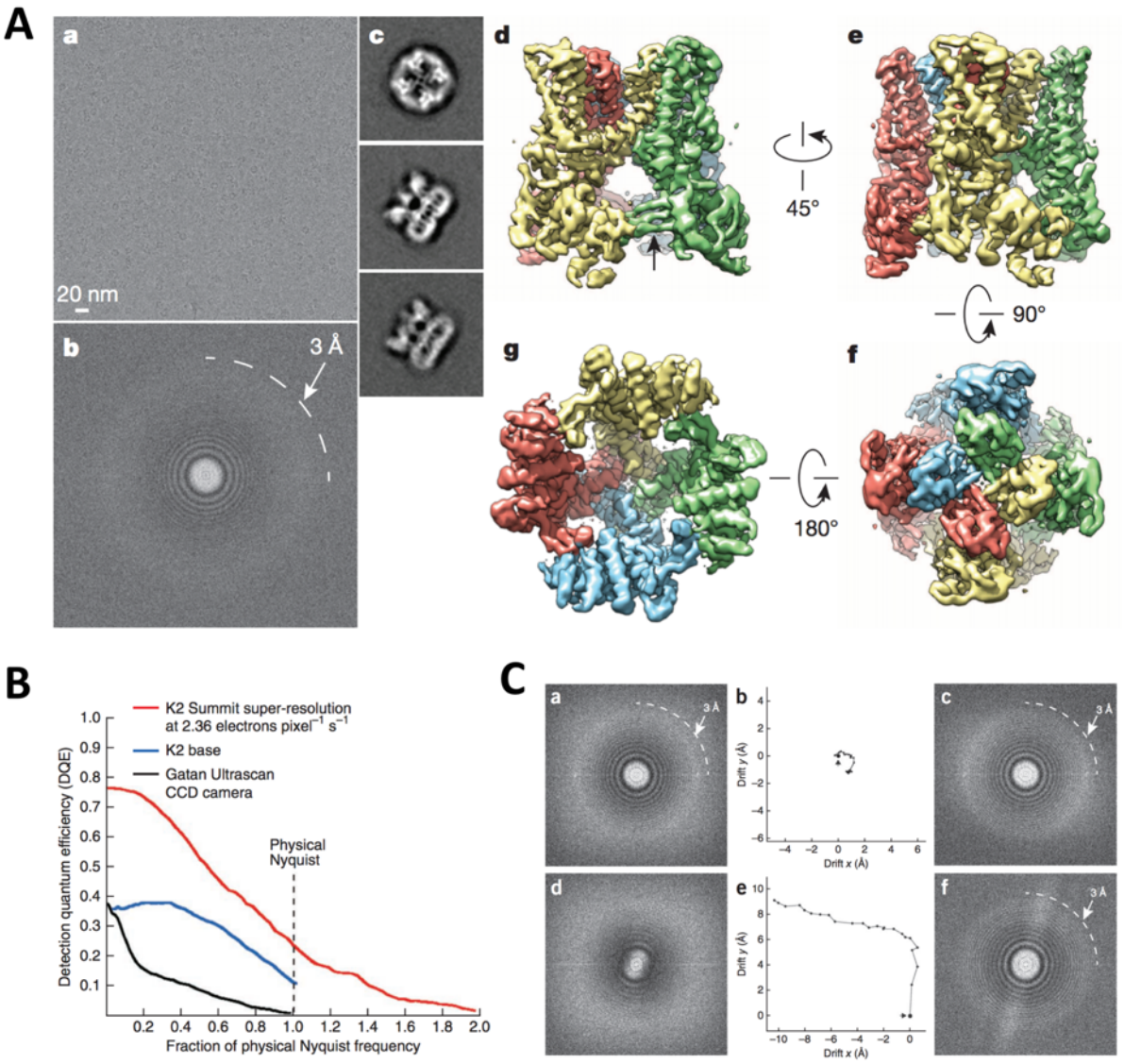

Figure 1.5 The structural determination of TRPV1 using a direct electron detector and motion-corrected images.

A. The reconstruction of TRPV1 to 3.4 Å resolution. (a. A micrograph of solubilized TRPV1 proteins embedded in vitreous ice. b, A Fourier transformed image showing Thon rings visible at high resolution. c. 2D class averages of TRPV1 where secondary elements can already be seen. d-f. Different views of the cryo-EM map of TRPV1.) B. A comparison of DQEs of K2 detector and traditional CCD. C. The correction of beam-induced motion helps restore high resolution information. (a&d. Thon rings before motion correction. b&e. Relative motion between frames plotted in two directions. c&f. Thon rings of the same images as in c&d after motion correction.) (Panel A is reprinted from ref. 50; panel B and C are reprinted from ref. 52. Both reprints are under permissions from Springer Nature.)

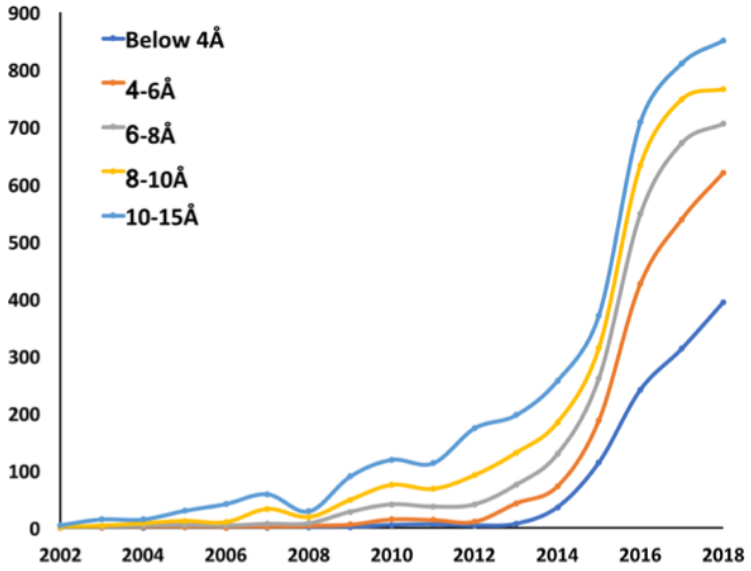

Figure 1.6 Number of maps deposited to EMDB by year and resolution.

Other than higher resolutions, the time needed for structural determination by cryo-EM has been much reduced, thanks to the user-friendly programs and customized pipelines (62). Faster data processing makes it possible to solve a structure "on the fly" and to adjust sample preparation accordingly. One of the programs, cryoSPARC, even supports semi-automated data processing and is able to generate high resolution results from raw micrographs without supervision (63). In the case of the first structure of Zika virus determined during an outbreak in 2016 (64), it took less than a month from harvest of the viruses to the achievement of the complete PDB model.

Since 2013, cryo-EM single particle reconstruction has emerged as a mainstream technique in structural biology to pursue high resolution structures (figure 1.6) (65, 66). Compared to X-ray crystallography, cryo-EM is sample-friendly, easy to manipulate and powerful in dealing with structural heterogeneities. At this point, the "revolution" is still going on. The highest resolution has been improved several times in the past three years (67, 68) and the lower barrier for sample size has been pushed to 64 kDa (69) while revisiting the structure of haemoglobin with a Volta phase plate.

1.2 Cryo-electron microcopy: a versatile tool box for structural virology

1.2.1 Handling structural heterogeneity by 2D and 3D classifications

Heterogeneity always exists in biological samples and obstructs high-resolution structure determinations. In crystallography, structurally different viruses or proteins have to be removed prior to crystallization. Whereas in electron microscopy, this problem can be addressed later in image processing.

A non-reference classification of all the input 2D projections (2D classification) has become a routine in the processing of cryo-EM data. Similar projections (the same macromolecules projected from similar orientations) are grouped together and an average is calculated for each class. This sorting process is usually accomplished in an iterative manner until the classes stabilize. In each of the iterations, an experimental projection is compared to the class averages from the last iteration. Both rotation and translation need to be considered. To reduce computing time, an autocorrelation algorithm is often used to enable translational-free comparison of two given images. A relative rotation can then be determined to find the best match. Finally, a projection can be classified into one of the groups that has the best correlation.

This process does not need any 3D model and is usually performed immediately after particle selection. The quality of the dataset, the orientation distribution of the particles and any existence of multiple structural states are usually revealed in 2D classifications. In structural virology, 2D classification is often used to separate full and emptied viral particles. Any Fab or receptor binding to the viral capsids should also be detectable in 2D classification (figure 1.7).

Any remaining structural heterogeneities can be further addressed by performing similar classification at a 3D level. Different software packages have different strategies for 3D classification but the general ideas are similar: projection matching (angular search) is carried out against multiple 3D references instead of one, so that the projections of structurally distinct states can be separated.

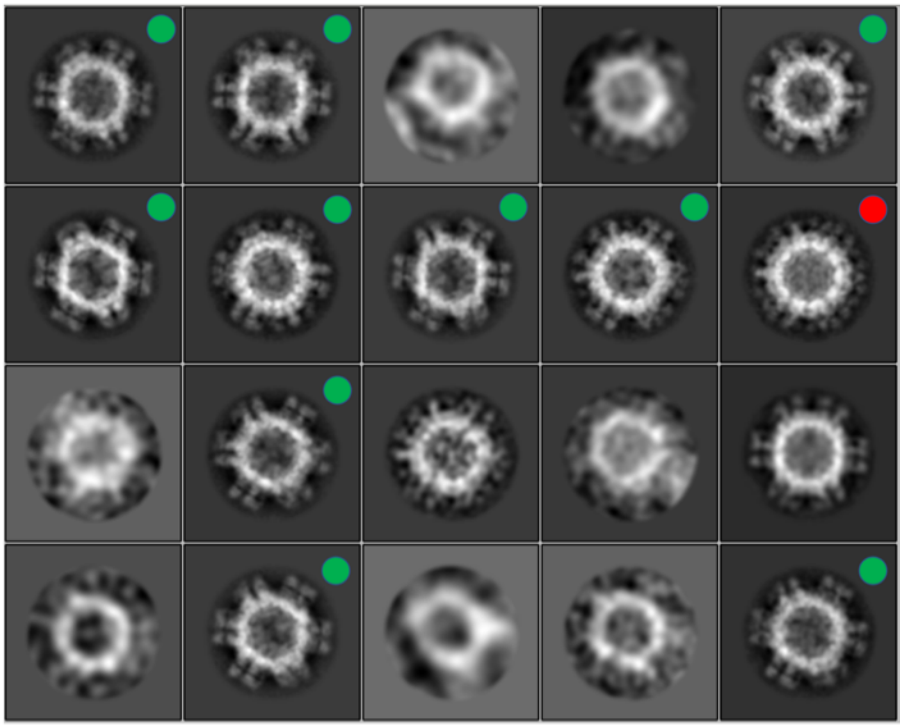

Figure 1.7 A 2D classification of B19-Fab complexes using RELION.
Green solid circles indicate good classes that were selected. The class with a red circle may have a lower occupancy of Fab. The rest of the classes were not selected because blurriness is observed indicating heterogeneity or damaged particles.

There are different ways to perform 3D classifications to serve different purposes. Here RELION (70) is used as an example and more details can be found in other references (71, 72).

A basic 3D classification is preferably carried out at the beginning of 3D auto refinement, without prior knowledge to the structural states existing in the data. In RELION, one 3D initial model is needed for the first round of classification and all particles will be randomly assigned into different classes. In the following iterations, global angular searches will be performed against all the models from the previous iteration until the classes stabilize. Refinements with finer angular and translational steps can then be performed to show the structural state represented by each of the classes.

In some cases, the structural difference is caused by flexibility and broken particles. Thus, the classes out of a 3D classification do not necessarily represent distinct structural states. Usually more homogeneous particles will be sorted into one class. In this case, 3D classification can be used as a procedure to select and remove bad particles. This has been shown to be helpful to improve resolutions.

Sometimes subtle differences in the conformation or composition of the particles may not be effectively detected in a crude 3D classification. In this case, 3D refinement against a consensus model can be performed first without any classification to obtain relatively accurate angles for all the projections. Then 3D classification without angular refinement can help to sort the particles that have already been aligned. Because the angles are fixed, the general architecture of the model will not diverge during the classification and minor structural differences might be detectable. At the end, the detail of the differences can be visualized by refinements within each of the classes against the corresponding models.

When structural heterogeneity is only present in a sub-area, masking is usually necessary and effective. A mask in cryo-EM is essentially a 3D volume with voxel values of either one or zero. In 3D classification, a mask can be applied to the new references to erase any density that is outside the region of interest. Therefore, in the projection matching process, only the masked region will be considered. This is usually helpful when there is domain flexibility or higher resolution is desired for a certain region (73, 74). An example of masked 3D classification on the complexes of ΦX174 and lipopolysaccharides receptors is given in figure 1.8. In this example, the interface between the phage and the receptors was initially not well resolved. Two consecutive masked 3D classifications were performed to classify the heterogeneity in this region.

However, a modestly homogenous sample is still one of the requisites to achieve good results. Obviously, no algorithm can recover information that is not present in the sample.

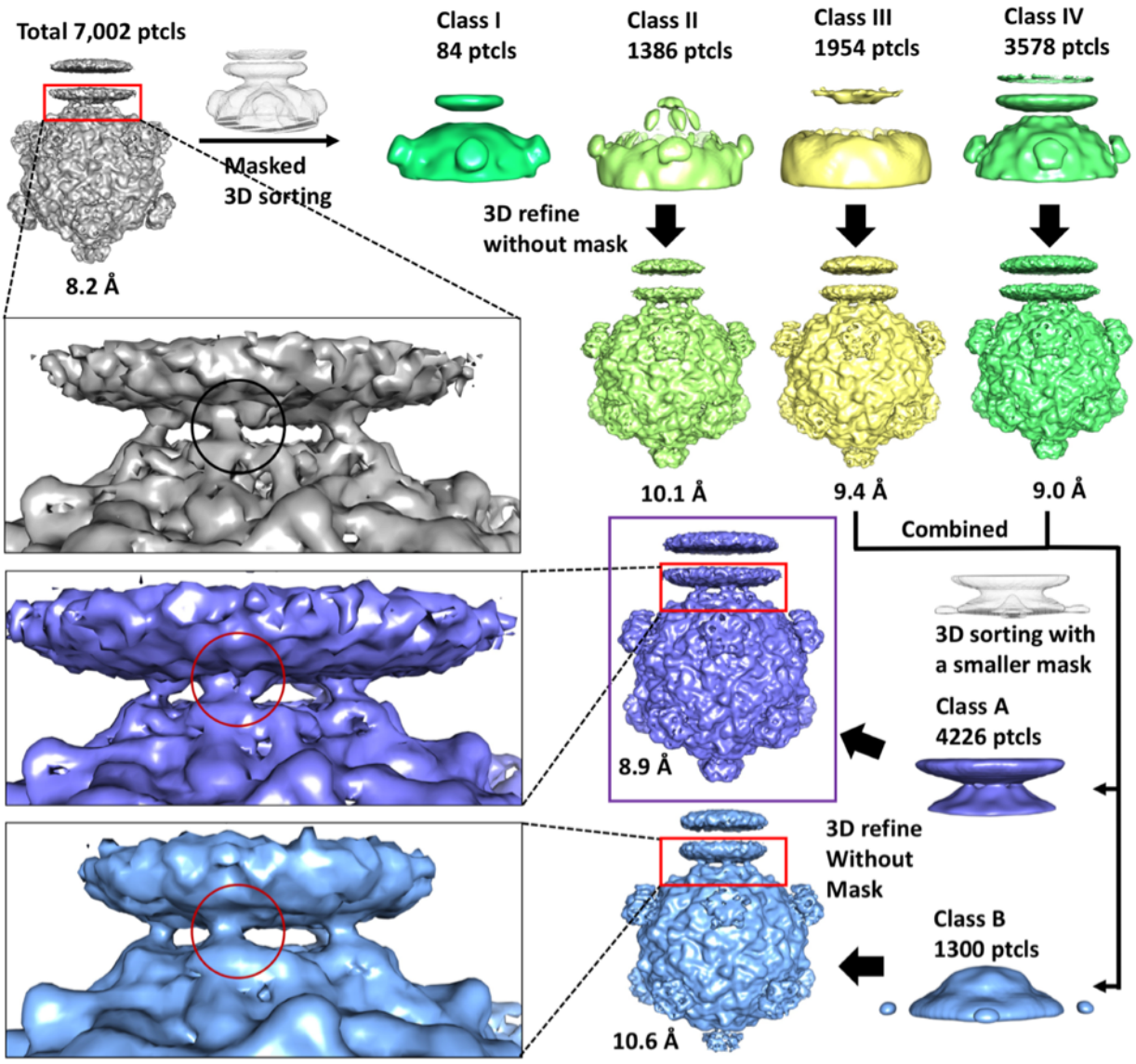

Figure 1.8 An example of masked 3D classification of ΦX174-LPS complexes. Two consecutive 3D classification using different masks helped sort the projections based on the features (black and red circles) in the interface region.

1.2.2 Pushing the resolution by correction of the Ewald sphere

In transmission electron microscopy, a micrograph is considered to be a projection along the z-axis and a single defocus value is assigned to this micrograph during contrast transfer function (CTF) estimation. An equivalent assumption is that the Ewald sphere of the electron scattering events has a very large radius (1/λ) because the wavelength of electrons is very small, and is therefore essentially a flat surface. This assumption leads to the Central Section Theorem which is the foundation for Fourier synthesis (reconstructing a 3D volume from a series of 2D projections)

used by all the single particles reconstruction programs (22). The assumption is almost correct when the sample is thinner than the depth of field (resolution dependent, usually smaller than 200 nm). However, when the sample being imaged is thicker than the depth of field, using the same strategy for data processing will limit the resolution (75).

In structural virology, it has been difficult to acquire high resolution structures for large viruses (76). This is because giant viruses are usually more flexible and the number of particles in one micrograph is very limited (Figure 1.9, panel A). Another important reason is that the sizes of these giant viruses, as large as thousands of Ångstroms, exceed the depth of fields of electron microscopes.

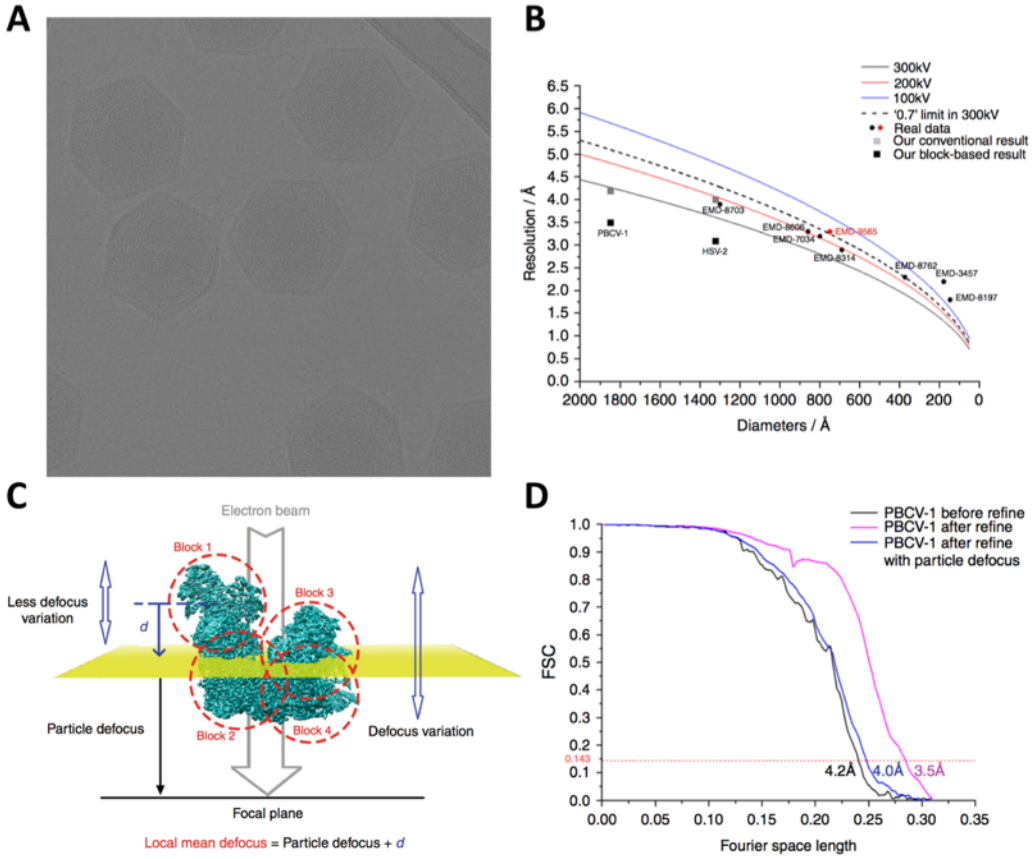

Figure 1.9 Limitations for large viruses and the Ewald sphere correction.
A. A micrograph of PBCV-1 imaged under 300 kV (642 x 664 nm). B. Resolution against sample size. C. A schematic representation of the "block-based reconstruction". Defocus gradient is compensated by divide a large sample into blocks. D. Block-based reconstruction effectively pushed the resolution of PBCV-1. (Figure in panel A is provided by Qianglin Fang. Panels B-D are reprinted from ref. 77. Link to license: https://creativecommons.org/licenses/by/4.0.)

To account for this problem, multiple algorithmic solutions have been proposed (77–80).

One of these methods is called “block-based reconstruction”. It divides one reconstruction into multiple blocks which will be assigned different defoci and then processed (figure 1.9C) (77). The structure of *Paramecium bursaria* chlorella virus (PBCV) capsid, with a diameter of 1900 Å, was previously determined to 8.5 Å assuming icosahedral symmetry (81). With the K2 detector and updated data processing, the resolution of PBCV-1 has been improved to 3.5 Å (Fang et al, unpublished data) even when using only 5-fold symmetry to account for the special vertex. Using the same method, another large virus, herpes simplex virus type 2 (HSV-2), has been reconstructed to a resolution of 3.1 Å (82). Another method to account for the defocus gradient, the “ single side-band image processing algorithm” (78) or “simple insertion method” (83), is implemented in Frealign (84) and cisTEM (85). In this method, the coefficients of individual projections are inserted into Fourier space as a curved sphere instead of a plane (86).

A similar issue also exists in cryo-electron tomography (cryo-ET), where the sample is usually thicker and images are deliberately tilted. A similar correction method customized for cryo-ET has been developed (87).

1.2.3 Beyond icosahedral symmetry

Macromolecules with higher symmetries have compressed structural information and provide more subunits to be averaged. Icosahedral symmetry that is possessed by many viruses has greatly facilitated the structural determinations and make viruses great models for technique development.

Icosahedral symmetry exists only in the viral capsid, whereas the genome or any enzymes packaged inside usually do not have the same symmetry. After being averaged by icosahedral symmetry in such cases, densities of asymmetric components are diminished in a reconstruction. Compared to crystallography, cryo-EM is more flexible in dealing with symmetry mismatches.

A simplest solution is to assume no symmetry at all during image processing (88–90). Bacteriophage MS2 has a single stranded RNA genome and a capsid composed of one maturation protein monomer and 89 coat protein dimers arranged with T=3 icosahedral symmetry. In a crystallographic study, neither the maturation protein nor the RNA genome was interpretable (91). In asymmetric cryo-EM reconstructions of MS2, both the ssRNA genome and the asymmetric maturation protein were resolved to resolutions better than had been attained crystallographically (figure 1.10) (92, 93).

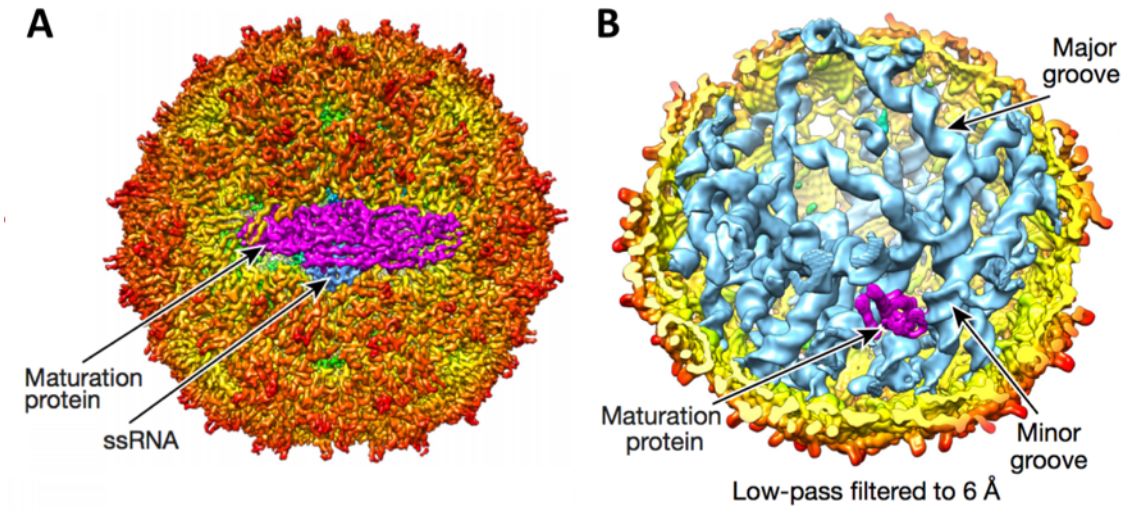

Figure 1.10 Cryo-EM asymmetric reconstruction of bacteriophage MS2.
A. The asymmetric map of MS2 capsid at 3.6 Å. The maturation protein monomer is colored in magenta. B. The interior of the MS2 reconstruction that has been low-pass filtered to 6 Å. The ssRNA genome is highlighted in light blue. The genome is connected to the maturation protein. (Reprinted from ref. 93 with permission from Nature Springer.)

There are two issues that come with asymmetric reconstructions of icosahedral viruses. First, 60 times more data would be required. Second, the orientation of the asymmetric component of a virus, e.g. its genome, has to be fixed relative to the icosahedral axes of a virus. In the case of MS2, the genome is rigidly connected to the maturation protein on the capsid and the extrusion of maturation protein helped to locate the special vertex in the angular search. In other words, if the genome does not have a fixed orientation relative to the icosahedral lattice, it would be almost impossible to correctly align the genome density.

The issue is usually referred to as symmetry mismatch. Symmetry mismatched components have to be processed independently using different symmetries.

One of the solutions is to perform masked or focused reconstruction. Bacteriophage T7 has an icosahedral head whereas there is also a portal complex at one of the vertices responsible for DNA packaging. The portal complex consists of multiple protein components which have different symmetries. Guo et al. used a FAR (focused asymmetric reconstruction) method to mask out certain regions and performed reconstruction of a single component at a time using the correct symmetry (94).

Another approach is to first carry out an icosahedral reconstruction and then subtract the icosahedral capsid from the micrographs. The remaining projections will then be used in an

asymmetric reconstruction (95). Cytoplasmic polyhedrosis virus (CPV) from the *Reoviridae* has a double stranded RNA genome and a transcriptional enzyme complex inside an icosahedral capsid. The asymmetric features were visualized after a reconstruction using capsid-subtracted projections. A work flow of the method is shown in figure 1.11. Projections of the icosahedral reconstruction of the capsid along the same orientations with the experimental images were produced. Then these projections will be subtracted from the "raw particles" using scaling factors determined by cross-correlation individually.

As mentioned above, the protein shells of many dangerous viruses do not have any symmetry and are resistant to single particle averaging. Another emerging cryo-EM technique, electron tomography, is suitable in the studies of pleomorphic viruses and virus-host interactions. This topic will be discussed in the next section.

1.2.4 *In situ* studies using cryo-electron tomography: a future direction

The central concept of cryo-EM reconstruction is the synthesis of 2D projections into 3D volumes, which requires the knowledge of the orientations of all the projections.

In single particle reconstructions, the orientation of each "particle" is acquired through iterative comparisons with all the possible projections of a 3D reference which is updated after every iteration. As a result, single particle analysis is limited to purified, structurally identical macromolecules. Whereas in cryo-electron tomography (cryo-ET), the stage of a microscope is tilted to generate different projections of the same ice-embedded region and therefore, the relative angles between the projections are known. Thus, 3D reconstruction of a single area of interest is possible and any structurally unique feature is preserved.

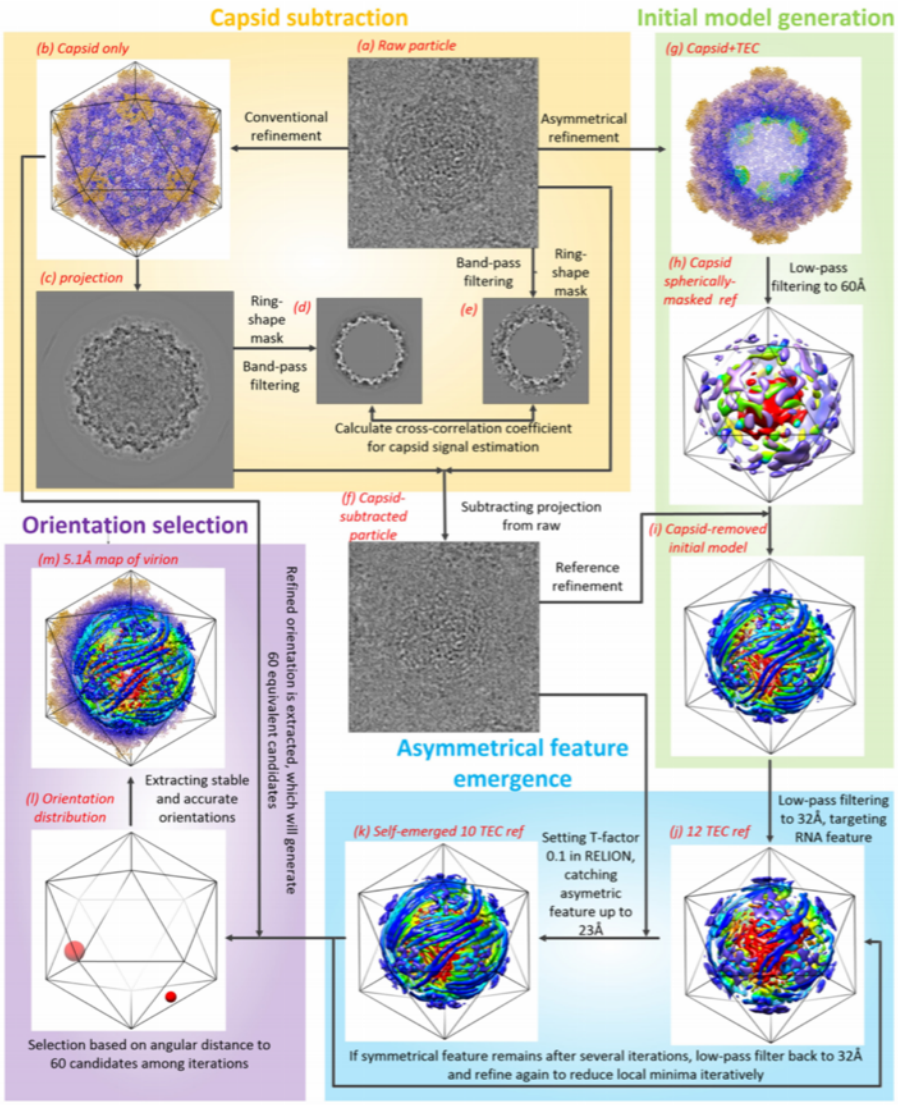

Figure 1.11 A procedure described by Zhang et al. to perform asymmetric reconstruction with the icosahedral capsid subtracted from the projections.
(Reprinted from ref. 95 with permission from Nature Springer)

Cryo-ET has been used to study the organization of many pleomorphic viruses, including the capsid (96, 97) and glycoprotein trimers (98, 99) of HIV, the nucleocapsid of filoviruses (100, 101) and the fusion influenza virus (102, 103). These viruses have irregular shapes and their glycoproteins are randomly distributed on the surface. One strategy to study the structures of these heterogeneous assemblies to sub-nanometer resolutions is sub-tomogram averaging (figure 1.12) (104). Essentially sub-volumes that contain smaller components like glycoprotein spikes and nucleocapsid proteins, are selected and cropped from the original tomograms. These smaller volumes, or sub-tomograms, are then aligned to a model before being averaged to achieve a new

model. This process is repeated and the quality of the model should improve over iterations. At the end, the orientations and centers for the sub-tomograms are no longer changing and the model also stabilizes.

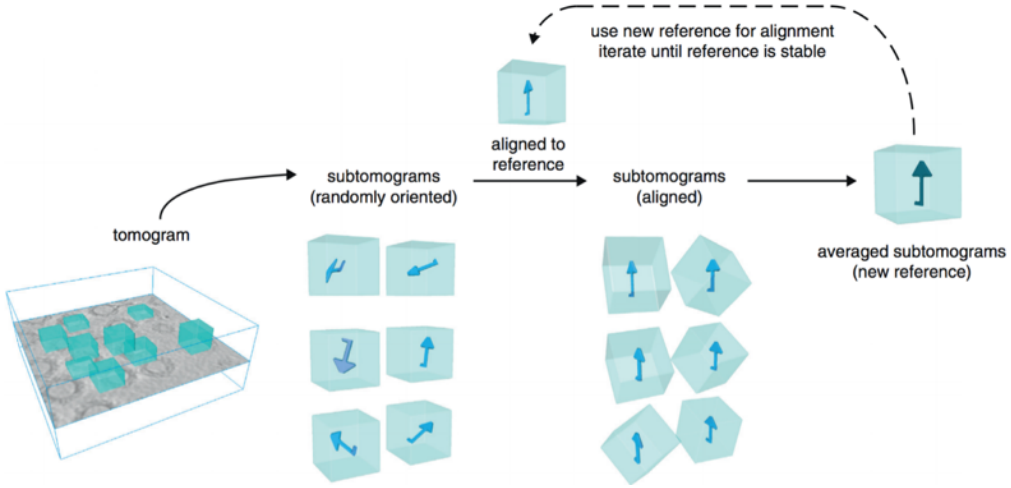

Figure 1.12 A schematic representation of subtomogram averaging.
Subtomograms are firstly cropped from tomograms and then aligned to a model by both translation and rotation. The average of newly aligned subtomograms is used as the model for the next round. This is repeated until the model and the alignments become stable. (Reprinted from ref. 104, with permission from Elsevier.)

The recent advances in detector technology, phase plate and automated data collection have also improved the quality of cryo-ET. In a study on the maturation of HIV capsid, by using an optimized strategy for data collection and averaging half a million subunits, an unprecedented resolution of 3.9 Å was achieved (figure 1.13A&B) (96). The resolution of the sub-tomogram averaging was further improved to 3.4 Å by utilizing an innovated method of CTF correction (87).

Cryo-ET is particularly suitable for studying dynamic virus-host interactions. Liposomes, mini-cells or extracted host membranes are often used to study the entry of different viruses, including influenza virus (105), arenaviruses (106), HIV (107), and bacteriophages (108–110). These reactions usually involve multi-components and each of the components may have several structural states. Cryo-ET allows 3D reconstruction of individual complexes and subsequently classifications based on different structural states. Although in many cases, cryo-ET is still limited in terms of resolutions, it can be compensated by high resolution structures determined by either single particle analysis or crystallography. Hallforsson et al. determined the crystal structure of a glycoprotein (Gn) of Rift Valley fever virus (RVFV) and a sub-tomogram average of pre-fusion

virions in contact with liposomes at low pH (figure 1.13C&D) (111). By combining the results from cryo-ET and crystallography, the authors proposed a model to show how the fusion loop of RVFV is hidden at neutral pH and how it is activated to trigger membrane fusion at acidic pH.

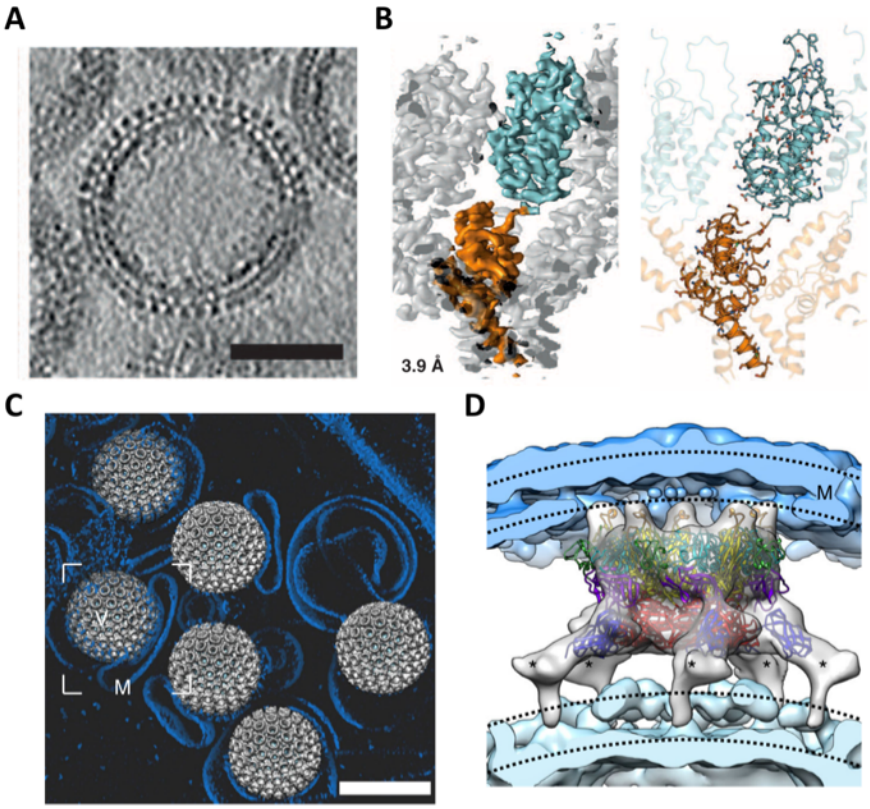

Figure 1.13 Applications of Cryo-ET in structural virology.
A. A slice through the tomogram of an HIV VLP particle treated with a maturation inhibitor. Scale bar, 50nm. B. Electron densities (left) and a model (right) of HIV capsid protein (CA) before spacer peptide (SP1) cleavage generated from subtomogram averaging at 3.9 Å. One CA-SP1 monomer is highlighted with the NTD in cyan and the CTD and SP1 in orange. C. A tomographic reconstruction of RVFV associated with liposomes at an acidic pH. The model was generated by placing the surface-rendered subtomogram averages of RVFV virions (grey) back into a tomogram. Scale bar, 100nm. D. A subtomogram average of the activated pentamer of Gn-Gc glycoprotein heterodimers in contact with liposomal membranes. The densities are fitted by crystal structures of Gn and Gc. The fusion loops from Gc are inserted into the outer leaflet of the liposomal membranes (blue). (Panels A and B are reprinted from ref. 96 with permissions from the American Association for the Advancement of Science. Panels C and D are reprinted from ref. 111 under license: https://creativecommons.org/licenses/by/4.0.)

However, the quality of tomographic reconstructions is still limited by a few factors:

The total dose of a tomogram has to be rationed to avoid radiation damage and the ice of a typical tomography sample is usually thick. These lead to very low signal-to-noise ratios for individual projections. Therefore, a tomogram reconstructed from these images is also very noisy;

Instrumentation is another big limiting factor. Due to the way that an EM grid is made, and the current design of a electron microscope, it is not able to achieve 180° rotation. Thus, a tilt-series can only cover part of the Fourier space (usually -60° to +60°) in back projection. This is known as the "missing wedge" problem. The "missing wedge" will introduce distortion in the density map and will also interfere with sub-tomogram alignments.

A third factor is that movements are usually introduced when the stage is tilted. Movements along the X- and Y- axes cause a change of the field of view. Movement along the Z axis causes a change of distance to the objective lens and defocus values as a consequence. The tilt also leads to a gradient of defocus across the image. All of these happen simultaneously, especially at high tilt angles.

If one or more of these issues are successfully addressed, which very likely will happen in near future, the quality of electron tomograms will be greatly improved and *in situ* structural studies at high resolution would become common. In the future, it will be possible to directly observe all kinds of biochemical reactions and biological processes, with detailed structural information, inside cells.

1.3 References

1. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. 2012. Human viruses: discovery and emergence. Philos Trans R Soc Lond B Biol Sci 367:2864–71.
2. Klug A, Finch JT. 1960. The symmetries of the protein and nucleic acid in turnip yellow mosaic virus: X-ray diffraction studies. J Mol Biol 2:201-IN10.
3. Caspar DLD. 1956. Structure of Bushy Stunt Virus. Nature 177:475–476.
4. Finch JT, Klug A. 1959. Structure of poliomyelitis virus. Nature 183:1709–1714.
5. Caspar DL, Klug A. 1962. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24.
6. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. 1978. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276:368–373.
7. Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AGW, Rayment I, Rossmann MG, Suck D, Tsukihara T. 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 286:33–39.
8. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, Johnson JE, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153.
9. Liu Y, Hill MG, Klose T, Chen Z, Watters K, Bochkov YA, Jiang W, Palmenberg AC, Rossmann MG. 2016. Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. Proc Natl Acad Sci U S A 113:8997–9002.
10. Liu Y, Sheng J, Fokine A, Meng G, Shin W-H, Long F, Kuhn RJ, Kihara D, Rossmann MG. 2015. Structure and inhibition of EV-D68, a virus that causes respiratory illness in children. Science 347:71–4.
11. Oliveira MA, Zhao R, Lee WM, Kremer MJ, Minor I, Rueckert RR, Diana GD, Pevear DC, Dutko FJ, McKinlay MA, Rossmann MG. 1993. The structure of human rhinovirus 16. Structure 1:51–68.
12. Kaufmann B, Simpson AA, Rossmann MG. 2004. The structure of human parvovirus B19. Proc Natl Acad Sci U S A 101:11628–33.

13. Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG, Compans RW. 1991. The three-dimensional structure of canine parvovirus and its functional implications. Science 251:1456–64.
14. Plevka P, Perera R, Cardosa J, Kuhn RJ, Rossmann MG. 2012. Crystal structure of human enterovirus 71. Science 336:1274.
15. McKenna R, Xia D, Willingmann P, IIag LL, Krishnaswamy S, Rossmann MG, Olson NH, Baker TS, Incardona NL. 1992. Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications. Nature 355:137–143.
16. Horne RW, Wildy P. 1961. Symmetry in virus architecture. Virology 15:348–373.
17. Klug A, Caspar DLD. 1961. The Structure of Small Viruses. Adv Virus Res 7:225–325.
18. Klug A, Finch JT. 1965. Structure of viruses of the papilloma-polyoma type. I. human wart type I human wart virus. J Mol Biol 11:403–423.
19. DeRosier D. 2010. 3D Reconstruction from Electron Micrographs: A Personal Account of its Development. Methods Enzymol 481:1–24.
20. De Rosier DJ, Klug A. 1968. Reconstruction of three dimensional structures from electron micrographs. Nature 217:130–134.
21. Crowther RA, Amos LA, Finch JT, De Rosier DJ, Klug A. 1970. Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. Nature 226:421–425.
22. Crowther RA, DeRosier DJ, Klug A. 1970. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. Proc R Soc A Math Phys Eng Sci 317:319–340.
23. Erickson HP, Klug A. 1971. Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs. Philos Trans R Soc B Biol Sci 261:105–118.
24. Glaeser RM. 1971. Limitations to significant information in biological electron microscopy as a result of radiation damage. J Ultrastruct Res 36:466–482.
25. Taylor KA, Glaeser RM. 1974. Electron diffraction of frozen, hydrated protein crystals. Science 186:1036–7.
26. Henderson R, Unwin PNT. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257:28–32.

27. Frank J. 1975. Averaging of low exposure electron micrographs of non-periodic objects. Ultramicroscopy 1:159–162.
28. Frank J, Shimkin B, Dowse H. 1981. Spider—A modular software system for electron image processing. Ultramicroscopy 6:343–357.
29. Frank J, Goldfarb W, Eisenberg D, Baker TS. 1978. Reconstruction of glutamine synthetase using computer averaging. Ultramicroscopy 3:283–290.
30. Frank J, Verschoor A, Boublik M. 1981. Computer averaging of electron micrographs of 40S ribosomal subunits. Science 214:1353–5.
31. Radermacher M, Wagenknecht T, Verschoor A, Frank J. 1987. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J Microsc 146:113–136.
32. Dubochet J, Lepault J, Freeman R, Berriman JA, Homo J-C. 1982. Electron microscopy of frozen water and aqueous solutions. J Microsc 128:219–237.
33. McDowall AW, Chang J-J, Freeman R, Lepault J, Walter CA, Dubochet J. 1983. Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. J Microsc 131:1–9.
34. Adrian M, Dubochet J, Lepault J, McDowall AW. 1984. Cryo-electron microscopy of viruses. Nature 308:32–36.
35. Vogel RH, Provencher SW, von Bonsdorff C-H, Adrian M, Dubochet J. 1986. Envelope structure of Semliki Forest virus reconstructed from cryo-electron micrographs. Nature 320:533–535.
36. Baker TS, Olson NH, Fuller SD. 1999. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol Rev 63:862–922, table of contents.
37. Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, Baker TS, Strauss JH, Rossmann MG, Kuhn RJ. 2003. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. Nat Struct Mol Biol 10:907–912.
38. Böttcher B, Wynne SA, Crowther RA. 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386:88–91.
39. Rossmann MG, Bernal R, Pletnev S V. 2001. Combining Electron Microscopic with X-Ray Crystallographic Structures. J Struct Biol 136:190–200.

40. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC. 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. Nature 375:291–298.
41. Kuhn RJ, Zhang W, Rossmann MG, Pletnev S V., Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG, Baker TS, Strauss JH. 2002. Structure of Dengue Virus: Implications for Flavivirus Organization, Maturation, and Fusion. Cell 108:717–725.
42. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K, O'Dell S, Perfetto S, Schmidt SD, Shi W, Wu L, Yang Y, Yang Z-Y, Yang Z, Zhang Z, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Simek M, Burton DR, Koff WC, Doria-Rose NA, Connors M, NISC Comparative Sequencing Program NCS, Mullikin JC, Nabel GJ, Roederer M, Shapiro L, Kwong PD, Mascola JR. 2011. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science 333:1593–602.
43. Bartesaghi A, Merk A, Borgnia MJ, Milne JLS, Subramaniam S. 2013. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. Nat Struct Mol Biol 20:1352–1357.
44. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N. 2008. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc Natl Acad Sci U S A 105:1867–72.
45. Grigorieff N, Harrison SC. 2011. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr Opin Struct Biol 21:265–273.
46. Zhang X, Sun S, Xiang Y, Wong J, Klose T, Raoult D, Rossmann MG. 2012. Structure of Sputnik, a virophage, at 3.5-Å resolution. Proc Natl Acad Sci U S A 109:18431–6.
47. Zhang X, Ge P, Yu X, Brannan JM, Bi G, Zhang Q, Schein S, Zhou ZH. 2013. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. Nat Struct Mol Biol 20:105–10.
48. Henderson R. 1995. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys 28:171.
49. Rosenthal PB, Henderson R. 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. J Mol Biol 333:721–745.

50. Liao M, Cao E, Julius D, Cheng Y. 2013. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504:107–112.
51. Cao E, Liao M, Cheng Y, Julius D. 2013. TRPV1 structures in distinct conformations reveal activation mechanisms. Nature 504:113–118.
52. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods 10:584–590.
53. McMullan G, Faruqi AR, Clare D, Henderson R. 2014. Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy 147:156–163.
54. McMullan G, Faruqi AR, Henderson R. 2016. Direct Electron Detectors, p. 1–17. *In* Methods in enzymology.
55. McMullan G, Faruqi AR, Henderson R, Guerrini N, Turchetta R, Jacobs A, van Hoften G. 2009. Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy 109:1144–1147.
56. Faruqi AR, Henderson R, Pryddetch M, Allport P, Evans A. 2005. Direct single electron detection with a CMOS detector for electron microscopy. Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip 546:170–175.
57. Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N. 2012. Beam-induced motion of vitrified specimen on holey carbon film. J Struct Biol 177:630–637.
58. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS, Carragher B. 2005. Automated molecular microscopy: The new Leginon system. J Struct Biol 151:41–60.
59. Mastronarde DN. 2003. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. Microsc Microanal 9:1182–1183.
60. Plevka P, Lim P-Y, Perera R, Cardosa J, Suksatu A, Kuhn RJ, Rossmann MG. 2014. Neutralizing antibodies can initiate genome release from human enterovirus 71. Proc Natl Acad Sci U S A 111:2134–9.

61. Dong Y, Liu Y, Jiang W, Smith TJ, Xu Z, Rossmann MG. 2017. Antibody-induced uncoating of human rhinovirus B14. Proc Natl Acad Sci U S A 114:8017–8022.
62. Lander GC, Stagg SM, Voss NR, Cheng A, Fellmann D, Pulokas J, Yoshioka C, Irving C, Mulder A, Lau P-W, Lyumkis D, Potter CS, Carragher B. 2009. Appion: An integrated, database-driven pipeline to facilitate EM image processing. J Struct Biol 166:95–102.
63. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296.
64. Sirohi D, Chen Z, Sun L, Klose T, Pierson TC, Rossmann MG, Kuhn RJ. 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science 352:467–70.
65. Nogales E. 2016. The development of cryo-EM into a mainstream structural biology technique. Nat Methods 13:24–7.
66. 2016. Method of the Year 2015. Nat Methods 13:1–1.
67. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JLS, Subramaniam S. 2015. 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science 348:1147–51.
68. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S. 2016. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. Cell 165:1698–1707.
69. Khoshouei M, Radjainia M, Baumeister W, Danev R. 2017. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. Nat Commun 8:16099.
70. Scheres SHW. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530.
71. Scheres SHW. 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods Enzymol 579:125–157.
72. Scheres SHW, Gao H, Valle M, Herman GT, Eggermont PPB, Frank J, Carazo J-M. 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat Methods 4:27–29.
73. Bai X, Rajendra E, Yang G, Shi Y, Scheres SH. 2015. Sampling the conformational space of the catalytic subunit of human γ-secretase. Elife 4.
74. Bai X, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SHW, Shi Y. 2015. An atomic structure of human γ-secretase. Nature 525:212–217.

75. DeRosier DJ. 2000. Correction of high-resolution data for curvature of the Ewald sphere. Ultramicroscopy 81:83–98.
76. Xiao C, Chipman PR, Battisti AJ, Bowman VD, Renesto P, Raoult D, Rossmann MG. 2005. Cryo-electron Microscopy of the Giant Mimivirus. J Mol Biol 353:493–496.
77. Zhu D, Wang X, Fang Q, Van Etten JL, Rossmann MG, Rao Z, Zhang X. 2018. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. Nat Commun 9:1552.
78. Russo CJ, Henderson R. 2018. Ewald sphere correction using a single side-band image processing algorithm. Ultramicroscopy 187:26–33.
79. Jensen GJ, Kornberg RD. 2000. Defocus-gradient corrected back-projection. Ultramicroscopy 84:57–64.
80. Downing KH, Glaeser RM. 2018. Estimating the effect of finite depth of field in single-particle cryo-EM. Ultramicroscopy 184:94–99.
81. Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Etten JL Van, Rossmann MG. 2011. Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. Proc Natl Acad Sci.
82. Yuan S, Wang J, Zhu D, Wang N, Gao Q, Chen W, Tang H, Wang J, Zhang X, Liu H, Rao Z, Wang X. 2018. Cryo-EM structure of a herpesvirus capsid at 3.1 Å. Science 360:eaao7283.
83. Wolf M, DeRosier DJ, Grigorieff N. 2006. Ewald sphere correction for single-particle electron microscopy. Ultramicroscopy 106:376–382.
84. Grigorieff N. 2007. FREALIGN: High-resolution refinement of single particle structures. J Struct Biol 157:117–125.
85. Grant T, Rohou A, Grigorieff N. 2018. cisTEM, user-friendly software for single-particle image processing. Elife 7.
86. Tan YZ, Aiyer S, Mietzsch M, Hull JA, McKenna R, Grieger J, Samulski RJ, Baker TS, Agbandje-McKenna M, Lyumkis D. 2018. Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. Nat Commun 9:3628.
87. Turoňová B, Schur FKM, Wan W, Briggs JAG. 2017. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. J Struct Biol 199:187–195.

88. Chang J, Weigele P, King J, Chiu W, Jiang W. 2006. Cryo-EM Asymmetric Reconstruction of Bacteriophage P22 Reveals Organization of its DNA Packaging and Infecting Machinery. Structure 14:1073–1082.

89. Jiang W, Chang J, Jakana J, Weigele P, King J, Chiu W. 2006. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature 439:612–616.

90. Liu X, Zhang Q, Murata K, Baker ML, Sullivan MB, Fu C, Dougherty MT, Schmid MF, Osburne MS, Chisholm SW, Chiu W. 2010. Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. Nat Struct Mol Biol 17:830–836.

91. Valegård K, Liljas L, Fridborg K, Unge T. 1990. The three-dimensional structure of the bacterial virus MS2. Nature 345:36–41.

92. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ. 2016. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nat Commun 7:12524.

93. Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. 2016. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature.

94. Guo F, Liu Z, Vago F, Ren Y, Wu W, Wright ET, Serwer P, Jiang W. 2013. Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. Proc Natl Acad Sci 110:6811–6816.

95. Zhang X, Ding K, Yu X, Chang W, Sun J, Hong Zhou Z. 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. Nature 527:531–534.

96. Schur FKM, Obr M, Hagen WJH, Wan W, Jakobi AJ, Kirkpatrick JM, Sachse C, Kräusslich H-G, Briggs JAG. 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. Science 353:506–8.

97. Schur FKM, Hagen WJH, Rumlová M, Ruml T, Müller B, Kräusslich H-G, Briggs JAG. 2014. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. Nature 517:505–508.

98. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S. 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature 455:109–113.

99. Tran EEH, Borgnia MJ, Kuybeda O, Schauder DM, Bartesaghi A, Frank GA, Sapiro G, Milne JLS, Subramaniam S. 2012. Structural Mechanism of Trimeric HIV-1 Envelope Glycoprotein Activation. PLoS Pathog 8:e1002797.

100. Bharat TAM, Riches JD, Kolesnikova L, Welsch S, Krähling V, Davey N, Parsy M-L, Becker S, Briggs JAG. 2011. Cryo-Electron Tomography of Marburg Virus Particles and Their Morphogenesis within Infected Cells. PLoS Biol 9:e1001196.

101. Wan W, Kolesnikova L, Clarke M, Koehler A, Noda T, Becker S, Briggs JAG. 2017. Structure and assembly of the Ebola virus nucleocapsid. Nature 551:394–397.

102. Noda T, Sugita Y, Aoyama K, Hirase A, Kawakami E, Miyazawa A, Sagara H, Kawaoka Y. 2012. Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. Nat Commun 3:639.

103. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC. 2006. Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A 103:19123–7.

104. Briggs JA. 2013. Structural biology in situ — the potential of subtomogram averaging. Curr Opin Struct Biol 23:261–267.

105. Fontana J, Cardone G, Heymann JB, Winkler DC, Steven AC. 2012. Structural changes in Influenza virus at low pH characterized by cryo-electron tomography. J Virol 86:2919–29.

106. Li S, Sun Z, Pryce R, Parsy M-L, Fehling SK, Schlie K, Siebert CA, Garten W, Bowden TA, Strecker T, Huiskonen JT. 2016. Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike. PLOS Pathog 12:e1005418.

107. Sougrat R, Bartesaghi A, Lifson JD, Bennett AE, Bess JW, Zabransky DJ, Subramaniam S. 2007. Electron Tomography of the Contact between T Cells and SIV/HIV-1: Implications for Viral Entry. PLoS Pathog 3:e63.

108. Hu B, Margolin W, Molineux IJ, Liu J. 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proc Natl Acad Sci 112:E4919–E4928.

109. Xu J, Gui M, Wang D, Xiang Y. 2016. The bacteriophage ϕ29 tail possesses a pore-forming loop for cell membrane penetration. Nature 534:544–547.

110. Hu B, Margolin W, Molineux IJ, Liu J. 2013. The bacteriophage t7 virion undergoes extensive structural remodeling during infection. Science 339:576–9.

111. Halldorsson S, Li S, Li M, Harlos K, Bowden TA, Huiskonen JT. 2018. Shielding and activation of a viral membrane fusion protein. Nat Commun 9:349.

CHAPTER 2 AN ICOSAHEDRAL RECONSTRUCTION OF B19 VIRUS-LIKE-PARTICLES COMPLEXED WITH FABS FROM A HUMAN ANTIBODY AT HIGH RESOLUTION

2.1 Chapter Abstract

B19 is a small DNA virus that belongs to the *Parvoviridae* family. B19 is one of the most common human pathogens. Antibodies to B19 are present in most human adults. B19 virus has a strong tropism to erythroid progenitor cells and is able to cause a series of medical conditions including fifth disease, arthritis, myocarditis, *hydrops fetalis* and aplastic crisis. No approved vaccine is available for B19 and there is a lack of structural characterization of any B19 epitopes. Here we present the cryo-EM structure of a B19 virus-like particle (VLP) complexed with Fab of a human neutralizing antibody, 860-55D. A model was built into the 3.2Å resolution map and the antigenic residues on the surface of B19 capsid were identified. Antibody 860-55D crosslinks the capsid of B19 by binding to a quaternary structure epitope formed by residues from three neighboring VP2 capsid proteins.

2.2 Introduction

B19 is a human parvovirus in genus *erythrovirus*. B19 virus primarily infects erythroid progenitor cells and is the etiological agent of childhood erythema infectiosum (fifth disease) (1, 2). B19 infection is also associated with rheumatoid arthritis and myocarditis (3 – 7). Infections in pregnant women of B19 may lead to *hydrops fetalis* or even fetal loss (1, 8, 9). B19 can cause severe medical problems if the patient has haematological or immunological deficiencies. For example, in patients with AIDS or sickle cell disease, B19 infection interferes with hematopoiesis and causes acute anemia (10 – 15). B19 virus has a linear single strand DNA genome packaged into a T=1 icosahedral capsid. The capsid has a diameter of about 260 Å and is composed of 60 capsid proteins, a mixture of VP1 and VP2. VP2 accounts for about 95% of the viral capsid proteins whereas VP1 accounts for about 5%. VP1 is essentially identical to VP2, with the exception of a unique region of 227 amino acids (VP1u) at the N-terminus of VP2. VP1u contains a phospholipase A2 (PLA_2) domain which is necessary for B19 infection (16). The structure of the mature virus is unknown whereas the structure of a virus-like particle (VLP) that only contains

VP2 has been determined (figure 2.1) (17). A few receptors have been suggested for B19. Erythrocyte P antigen has been shown to play a central role in B19’s tropism to bone marrow (18). The initial receptor binding induces the externalization of the VP1u which then binds to a co-receptor and initiates the viral uncoating process inside the host cell (19 – 23).

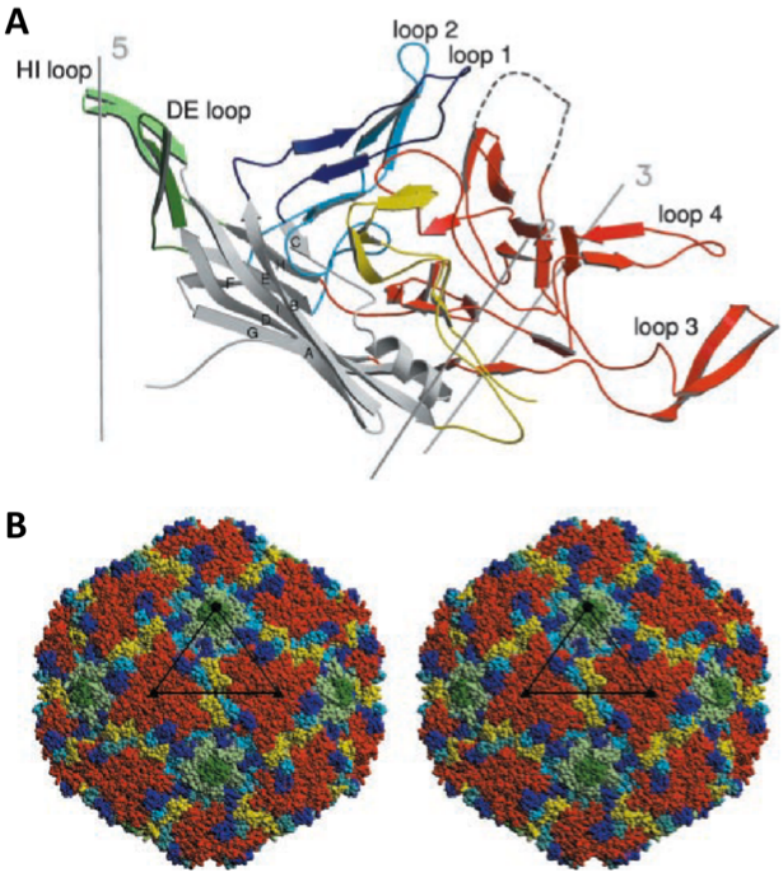

Figure 2.1 The structure of B19 VLP. A. The ribbon diagram of VP2.
The eight β-strands of the jelly roll fold are labeled from B to I. The loop insertions are highlighted in different colors. B. The stereo view of VP2 VLP. The color codes are the same as in panel A, with DE and HI loops at the 5-fold axes. An asymmetric unit is outlined by a black triangle. (Reprinted from ref. 17)

Human antibody 860-55D derived from an HIV positive patient and diminishes B19 infectivity by 50% when the antibody has a concentration of about 0.73 μg/ml or 4.9 nM (24). MAb 860-55D binds only to assembled viral particles and therefore has been used to detect mature B19 virions or VP2 VLPs (22, 25 – 28).

Currently, there is no approved vaccine for B19, and B19 infection is still a threat in certain circumstances. Structural investigations on virus-antibody interaction have been reported for many

parvovirus, but not for B19 (29 – 33). To study the antigenic properties of B19, we have determined to 3.2Å resolution the structure of a complex of B19 VLP consisting only of VP2 subunits and the Fab of MAb 860-55D using cryo-electron microscopy (cryo-EM) single particle analysis. The electron potential map shows that 860-55D recognizes a conformational epitope composed of three neighboring VP2 proteins close to the "canyon" around the 5-fold axes. By analogy with adeno-associated virus type 2 (AAV-2), the bound antibodies might be blocking the binding of the α5β1 integrin co-receptor and therefore preventing infection. Alternatively, the Fabs that crosslink the capsid proteins prevent uncoating.

2.3 Results and Discussions

2.3.1 The overall structure of B19-Fab complex

Purified VLPs were mixed with an excess of Fab molecules at 1:120 ratio (2 Fab molecules per VP2). The mixture was vitrified after 1 hour incubation at room temperature. The otherwise smooth VLPs became "fuzzy" in the electron micrographs because of the bound Fab molecules (figure 2.2). The cryo-EM data was processed using the jspr package (34) assuming icosahedral symmetry. A resolution of 3.2 Å was achieved with about 8,000 particles, as calculated by a "gold standard" FSC curve with a cutoff value of 0.143.

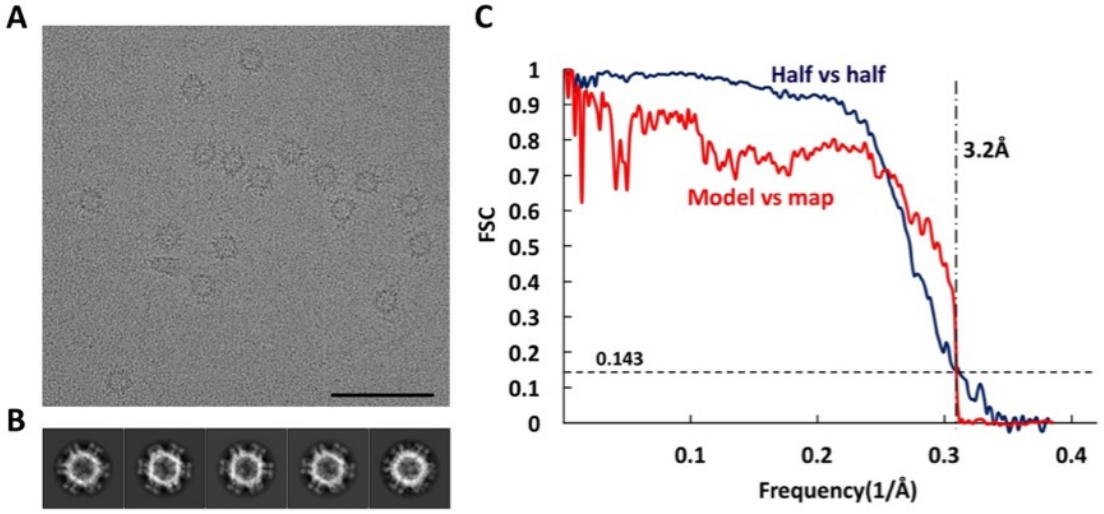

Figure 2.2 Cryo-EM processing of the B19-Fab complexes.
A, A raw micrograph of complexes embedded in vitreous ice. (bar = 100nm). B, Averages of representative 2D classes of the Fab-bound viral particles. C, Resolution evaluation using a gold

standard Fourier shell coefficient curve (dark blue) and a FSC curve between the model and the map (red).

The B19 VLPs were found to be decorated by 60 Fab molecules, one in each icosahedral asymmetric unit (Figure 2.3). The diameter of the complex is about 390 Å. The light- and heavy-chains of the Fab molecules bind close to the 5-fold vertices of the virus. The pseudo 2-fold axis of the variable domain was almost perpendicular to the surface of the capsid. Most of the side chains on VP2 could be built with confidence (figure 2.3B), except that loops 64-77, 300-312, 357-369, 397-400 and 525-535 are disordered. The quality of the Fab electron potential density was best closest to the viral capsid. There is not obvious difference in density height between the the capsid and the Fab molecules indicating a high occupancy of the Fab molecules. Given the 1:2 molecular ratio used in the sample preparation, the Fab molecules must be binding tightly to the B19 VLPs.

Table 2.1 Statistics of cryo-EM data collection, processing and model building.

Data collection and refinement parameters	
No. of movies used	1759
Defocus range (μm)	1.2 – 2.5
Electron does ($e^-/Å^2$)	38
No. of frames per movie	40
Exposure time per frame (ms)	200
Pixel size (Å/pixel)	1.30
No. of particles to start with	9120
No. of particles used for final map	7395
Symmetry imposed	icosahedral
Resolution of final map (Å)	3.22
Model refinement with NCS applied	
map-model correlation	0.736
all-atom clash	6.77
Ramachandran outliers	1.31%

The good quality of the potential density (figure 2.3C) allows the unambiguous identification of residues that form the epitope on the virus. However, no sequence information was available for the Fab. Therefore, the known structures of several homologous Fab molecules (IgG type 3, light chain type λ) were fitted into the density. PDB 5fhb (35) was used as the starting model. Five out of the six complementarity-determining region loops (CDR) fitted very well, leaving the CDR3 of the heavy chain (CDRH3) uninterpreted. Therefore, poly-alanine was used to model this region of 16 or 17 residues between the conserved cysteine and tryptophan that flank CDRH3.

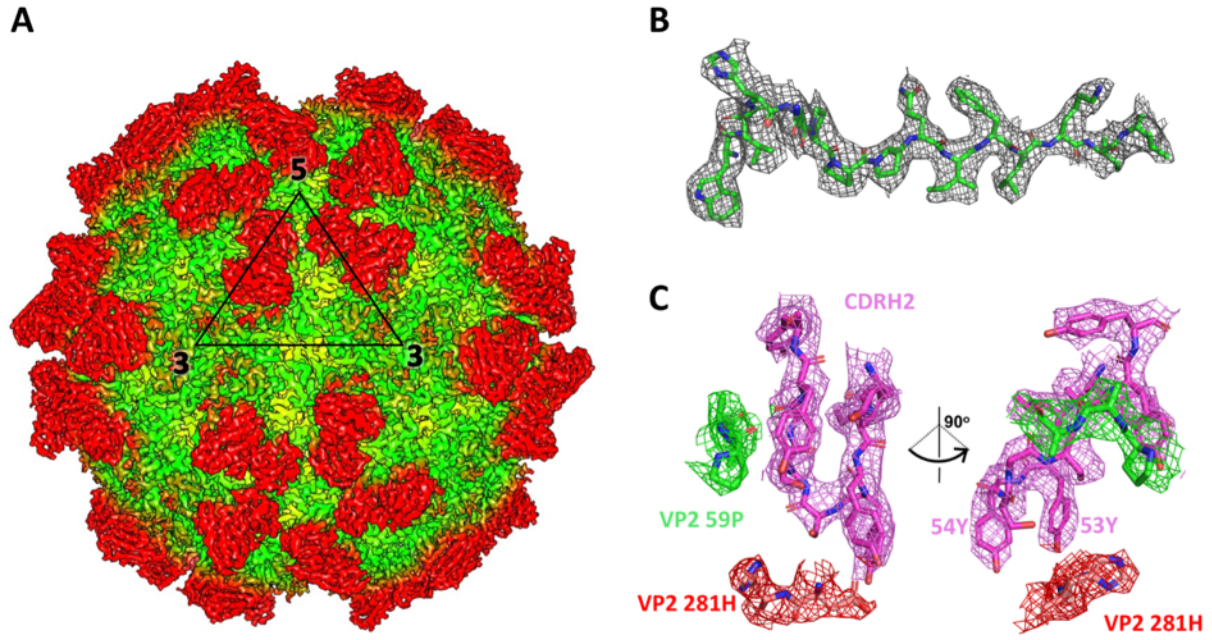

Figure 2.3 The structure of B19 VLP complexed with 860-55D Fab molecules.
A, Surface rendered cryo-EM map of the B19-Fab complex at 3.2 Å resolution. Yellow (100 Å), green (125 Å) and red (140 Å) coloring indicate increasing distances from the center of mass. B&C, Representative densities from the map of the B19-Fab complex with the fitted atomic model. B: Residues 447-461 of VP2; C: The interface between VP2 proteins and the Fab. Magenta: Fab CDRH2 residues 51-60; red: VP2 residues 280-281; green: VP2 residues 58-60.

2.3.2 The conformational epitope across three neighboring VP2 molecules

The epitope on B19 recognized by 860-55D is formed by residues distributed across three VP2 proteins, which will be referred to as P1, P2 and P3 (figure 2.4A). Among the six CDR loops, H1-3 and L1 make contact with the epitope. P1 and P2 are related by a 5-fold axis, and P2 and P3 are related by a 3-fold axis. P2 is located right beneath the bound Fab, whereas the HI loop of P1 extends above P2 and is inserted under the Fab molecule (the nomenclature of the VP2 loops

follows Kaufmann et al. (17)). P2 and P3 are interlocked and loop 3 of P3 extends between the Fab and P2, almost making contact with the HI loop of P1 (figure 2.4B). Loop 3 (P255-Y257 and L276-H281) of P3 is surrounded by the three CDR loops on the heavy chain (figure 2.5A). The HI loop in P1 sits between CDRH3 and CDRL1 and the tip of loop 1 (63S-78S) in P2 makes contact with CDRL1.

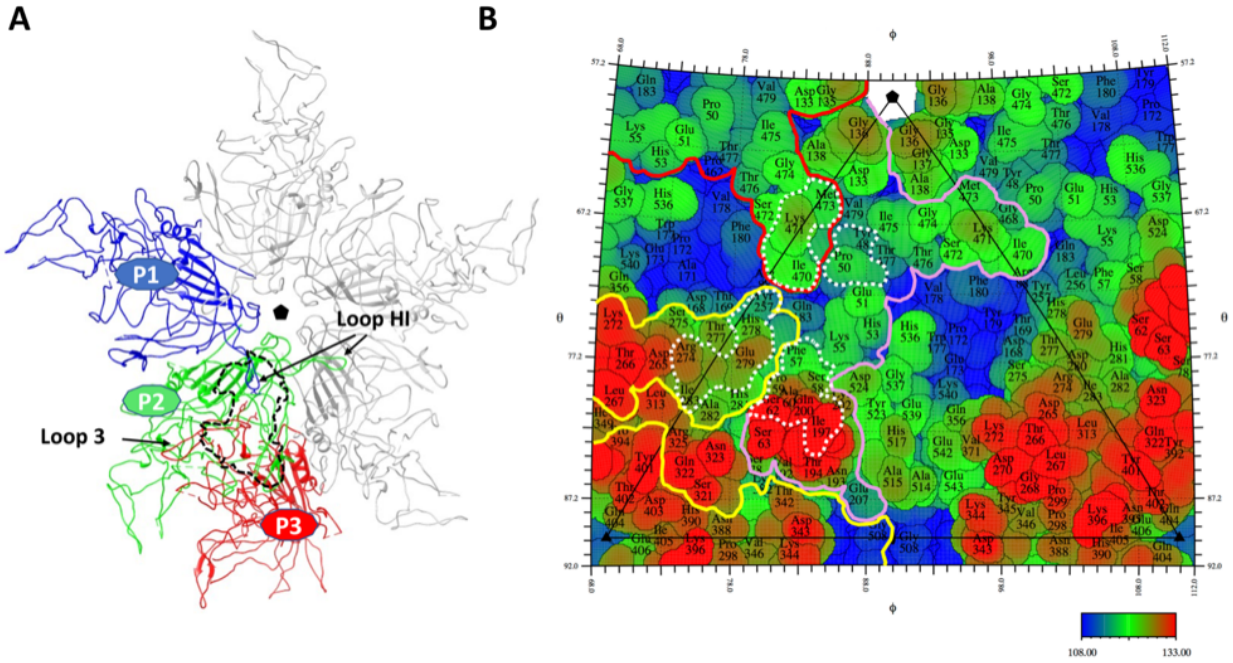

Figure 2.4 The quaternary structure epitope on B19 consists of residues from three VP2 proteins. A, VP2 proteins around an icosahedral 5-fold vertex. The black pentagon represents an icosahedral 5-fold axis perpendicular to the plane of the figure. The black dashed line outlines the footprint of one of the 60 bound Fab molecules. This Fab recognizes a quaternary epitope formed by three neighboring VP2 proteins, which are colored in blue, green and red. They are referred to as P1, P2 and P3, respectively. The VP2 proteins that do not interact with the same Fab molecule are colored in grey. B, A "roadmap" showing the surface residues of VP2 capsid projected onto a plane surface. Each of the residues is colored according to its distance from the center of the capsid. The black triangle outlines an asymmetric unit. The 2-fold axis is at the middle bottom, perpendicular to the image. The boundaries of different VP2 proteins are represented by thick solid lines (red: P1; pink: P2; yellow: P3). Residues that form the conformational epitope (within 4 Å from Fab) are surrounded by white dotted lines.

Residues that make contact with the Fab molecules were identified using the *ncont* function in the CCP4 suite with a cutoff distance of 4 Å (table 2.2). All the residues identified by the program were visually confirmed. Although an alanine was built into position 101 of the heavy chain, the electron density map indicates that residue 101 has a large side chain and is within 3.5

Å from F57 of P2 (figure 2.5B). Although F57 did not meet the distance criterion of 4 Å, it should be considered as a part of the epitope.

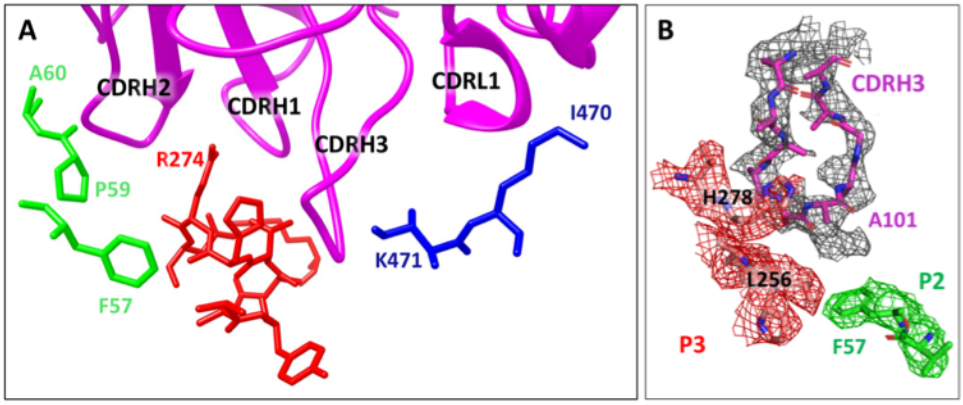

Figure 2.5 Interactions between VP2 proteins and MAb 860-55D.
A, Residues on loop 3 of P3 are interacting with all three CDRs on the heavy chain of the Fab (red). Residues (green) on loop 1 of P2 interact with CDRH2. The tip of the HI loop (blue) on P1 interacts with CDRH3 and CDRL1. B, CDRH3 of the Fab interacts with all three VP2 proteins. Residue 101 on the heavy chain has a big side chain which stacks with F57 on loop1 of P2.

In summary, the HI loop of P1, the loop 1 and loop 2 of P2, together with the loop 3 of P3 form a quaternary structure epitope on the surface of the B19 capsid. As a result, when a B19 virus is recognized and bound by antibody 860-55D, the capsid proteins are crosslinked. The number of copies of antibodies required to neutralize one virus and whether the crosslinking does prevent the uncoating process remain to be studied.

Table 2.2 Residues of VP2 that form the epitope.

Variable regions on 860-55D	VP2 copy no. (fig. 2.4)	VP2 residues within 4 Å
CDRL1	1	I470, K471, M473
	2	P50, Y48
CDRH1	3	R274, E279, H278
CDRH2	2	S58, P59, A60, S62, D200, I197
	3	R274, E279, D280, I283
CDRH3	1	I470
	2	F57*
	3	L256, Y257, H278

2.3.3 Structural alteration of VP2 by Fab binding

The root mean square deviation between the atoms in the bound and unbound VLPs was only 1.20 Å. Therefore, binding of the Fab molecules did not significantly change the structure of the VLPs (17) although there are a few loop regions that change their structures as much as 10 Å for some atoms (K471) (figure 2.6). A major difference occurs in the HI loop (amino acids 467-474) which is a part of the epitope. The HI loop is located close to a 5-fold vertex, and extends through the boundary between the capsid proteins. With the Fab molecules bound, the tip of the HI loop is bend by about 60^{o} toward the Fab and interacts with both the CDRH2 and CDRH3 loops. A similar but less significant conformational change occurs in the loop 2 which is slightly bend due to the interaction with CDRH2. Among the five disordered loops in the cryo-EM map, only 300-312 is also disordered in the crystallographic structure of the unbound VLP. This is possibly due to stabilization by crystalline packing.

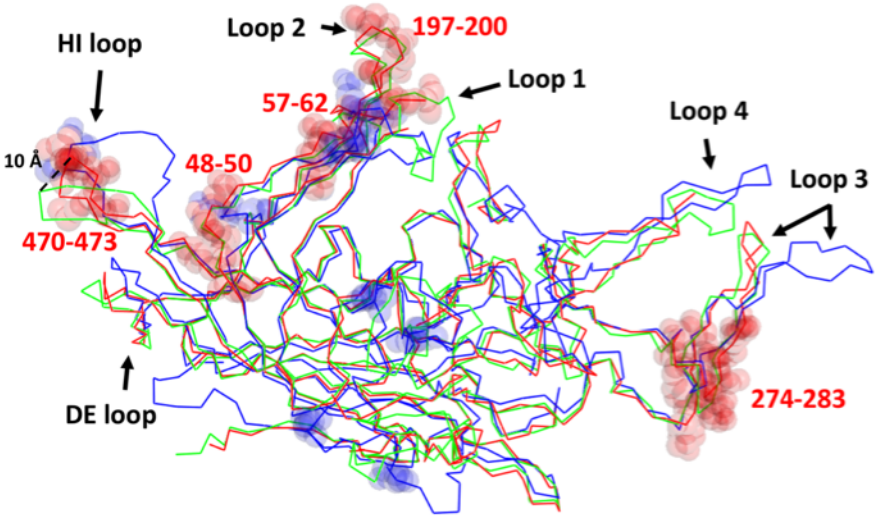

Figure 2.6 Comparison of B19 VP2 with fab bound (red), B19 VP2 without fab bound (PDB-1s58, green) and AAV-2 VP3 (PDB-1lp3, blue).
The most significant conformational change caused by Fab binding occurs at the HI loop, with a 10 Å difference in C_{α} of K471. The residues that form the epitopes of 860-55D (B19, red) and A20 (AAV-2, blue) are represented in transparent spheres.

2.3.4 The mechanism of neutralization

Antibodies that recognize viral capsid proteins are able to neutralize infections by inhibiting genome release (uncoating for non-enveloped viruses or membrane fusion for enveloped viruses),

blocking cell attachment, or inducing irreversible conformational changes (or even disassembly) of the virion.

B19 is structurally similar to AAV-2 (17, 36). AAV-2 uses a primary receptor, heparin sulfate proteoglycan, for initial cellular attachment and then binds to co-receptors (αVβ5 integrin and AAVR) for internalization (37–39). Similarly, B19 utilizes blood P antigen for attachment and α5β1 integrin as a cellular co-receptor. A neutralizing antibody of AAV-2, MAb A20, allows the viral attachment whereas it inhibits the internalization of the viruses (40).

Superimposing the structures of B19 VP2 and AAV-2 VP3 showed that A20 has a very similar conformational epitope with 860-55D (figure 5). Four of the five regions on AAV-2 that form the epitope of A20, overlap with the epitope of 860-55D in the alignment (32). On the other hand, the epitope of 860-55D does not overlap with the footprint of heparin sulfate on AAV-2 (41). Therefore, the neutralization of B19 by MAb 860-55D is likely achieved by blocking receptor binding to α5β1 integrin. However, it is also possible that 860-55D abrogates B19's ability to uncoat inside a host cell by crosslinking the capsid proteins.

2.4 Methods and Materials

2.4.1 Generation of B19 VP2-only virus-like particles and Fab of 860-55D

The procedure described here was based on the protocol described in previous studies (17, 42). The VP2 capsids were produced by a baculovirus-based expression system. VP2 protein was expressed in *Spodoptera frugiperdes* (Sf-9) insect cells and assembled into icosahedral (T=1) capsids. The AcMNPV-VP2 construct described previously was used to infect Sf9 cells at around 0.5 x 10^6 cells/ml as a suspension in Sf-900™ II media (Thermo Fisher Scientific) without any antibiotics. The cell culture was shaken continuously at 27^{o}C and harvested 70 hours after infection by centrifugation at 4000 g. The pallet was resuspended in lysis buffer (10 mM Tris/HCl, pH 7.4; 10 mM NaCl, 15 mM MgCl2, 0.5% Triton X-100, supplemented with protease inhibitors). The cells were lysed by either repeated freezing and thawing or mild sonication. The lysate was centrifuged at 2000 g to eliminate the remaining cells and large cellular debris. The recombinant capsids in the lysate were sedimented into a 30% CsCl cushion by centrifugation at 150,000 g. No band could be seen after the centrifugation and a mouse antibody (ab64295, abcam) was used to determine the location of the capsids. The lower part of the cushion was extracted and added to a

solution containing 36% CsCl, 50 mM Tris/HCl (pH 8.7) and 25 mM EDTA. The mixture was centrifuged again at 150,000 g for 20-24 hours until a density equilibrium was reached. An opaque band was extracted and dialyzed against PBS-Mg (8 mM Na_2HPO_4, 1.47 mM KH_2PO_4, 137 mM NaCl, 2.68 mM KCl, 0.49 mM $MgCl_2$). The solution was then concentrated to the desired concentration.

Fab fragments of human antibody 860-55D were produced following the instructions of the Pierce™ Fab preparation kit (Thermo Fisher Scientific). First, 0.5 mg IgG was desalted by passing through a Zeba spin desalting column. Then the flow-through was mixed with equilibrated immobilized papain at 37 ^{o}C for 4 hours for digestion. The Fab fragments were purified from the Fc fragments by passing through an equilibrated protein A column. Additional washing was done to achieve optimal recovery. The flow-through fractions were combined and concentrated to obtain the purified Fab fragments.

2.4.2 Sample vitrification and electron microscopy imaging

Purified VLPs and Fab was mixed at a molar ratio of 1:120 (2 Fab molecules for each VP2 protein). The mixture was incubated at room temperature (22^{o}C) for one hour. Aliquots of 3 μl of the virus-Fab mixture were applied to glow-discharged lacey carbon grids (400 mesh copper, lot #200617, Ted Pella INC.) before the grids were blotted for 4-5 seconds and plunged into liquid ethane using a cryoplunge 3 system (Gatan). A total of 1759 movies of the vitrified complexes embedded in vitreous ice were collected in two sessions using a Titan Krios microscope operated at 300kV with a Gatan K2 direct electron detector in counting mode (3838 x 3710). Automated data collection was enabled by Leginon (43). The nominal magnification was 22,500 which produced a pixel size of 1.30 Å. The defocus range was set to 1.2 - 2.5 μm. Every movie consisted of 40 frames and each frame had an exposure time of 200ms. The dose rate received by the detector was 8 e^-/pixel/s generating a total dose of 38 $e^-/Å^2$ for each movie.

2.4.3 Image processing

Relative motion between the frames within a movie was corrected using MotionCorr (44) that had been modified by Wen Jiang at Purdue University. Contrast transfer function (CTF) estimation was calculated with CTFFIND3 (45). Semi-auto particle boxing was carried out using the program *e2boxer.py* in the EMAN2 package (46). A total of 9,120 particles were found and confirmed by

visual inspection. A subsequent 2D classification was calculated by RELION (47) with a 400 Å diameter mask. A total of 7,395 particles were selected. Hereafter the dataset was randomly divided into two half subsets. The two subsets were then independently refined against randomly generated initial models using the jspr package (34) assuming icosahedral symmetry. Parameters including orientation, center, defocus, astigmatism, scale, beam tilt and magnification anisotropy were refined for each of the particles through multiple iterations until convergence was reached. The Fourier shell coefficient (FSC) was calculated between the resulting maps of the two subsets. The resolution was 3.22 Å according to the 0.143 criterion (48). Then the subsets were combined to generate a final map. The map was then low-pass filtered to 3.2 Å.

2.4.4 Model building and refinement

The previously determined structure of VP2 was fitted into the cryo-EM electron potential density (17). Further modeling was done using Coot (49). For the Fab molecule, the variable region from another IgG molecule (PDB: 5fhb) was first fitted into the density. If a specific residue did not agree with the density, it was replaced by alanine. Most residues of the sequence could fit easily into the density except for the region around CDRH3. The CDRH3 loop of 860-55D consists of 17 amino acids (between C92 and W110) whereas the template structure (PDB 5fhb) has 19 amino acids. Residues 73-88 and 96-106 were changed to alanines in Coot. The models for the capsid and the Fab molecule were combined and refined together in real space using Phenix (50). The icosahedral symmetry was then applied to the model of one VP2-Fab complex to generate the entire capsid. Finally, the entire capsid-Fab complex was refined to reduce clashing and to maximize correlation with the experimental electron potential density using Phenix with NCS restraints. To evaluate the overall agreement between the model and the electron potential density, a map was calculated based on the model. The FSC curve was then calculated between the model map and the original cryo-EM map.

2.4.5 Data deposition

The final map, two half maps and the FSC xml file have been deposited with the Electron Microscopy Data Bank (EMDB) under session code EMD-9110. The coordinates of one VP2-Fab complex has been deposited as PDB-6MF7.

2.5 Chapter Acknowledgements

This chapter is reprinted from a submitted manuscript “Structure of parvovirus B19 decorated by Fab molecules from a human antibody” by “Yingyuan Sun, Thomas Klose, Yue Liu, Michael G. Rossmann” to *Journal of Virology* which is currently under review.

We thank Susanne Modrow from University of Regensburg for kindly providing the 860-55D antibodies. We thank Geng Meng for helpful discussions. This work is supported by NSF grant MCB-1515260 awarded to M.G.R.

2.6 References

1. Chorba T, Coccia P, Holman RC, Tattersall P, Anderson LJ, Sudman J, Young NS, Kurczynski E, Saarinen UM, Moir R, Lawrence DN, Jason JM, Evatt B. 1986. The Role of Parvovirus B19 in Aplastic Crisis and Erythema Infectiosum (Fifth Disease). J Infect Dis 154:383–393.
2. Anderson MJ, Lewis E, Kidd IM, Hall SM, Cohen BJ. 1984. An outbreak of erythema infectiosum associated with human parvovirus infection. J Hyg (Lond) 93:85–93.
3. Takahashi Y, Murai C, Shibata S, Munakata Y, Ishii T, Ishii K, Saitoh T, Sawai T, Sugamura K, Sasaki T. 1998. Human parvovirus B19 as a causative agent for rheumatoid arthritis. Proc Natl Acad Sci U S A 95:8227–32.
4. Saal JG, Steidle M, Einsele H, Muller CA, Fritz P, Zacher J. 1992. Persistence of B19 parvovirus in synovial membranes of patients with rheumatoid arthritis. Rheumatol Int 12:147–151.
5. Bock C-T, Klingel K, Kandolf R. 2010. Human Parvovirus B19–Associated Myocarditis. N Engl J Med 362:1248–1249.
6. Molina KM, Garcia X, Denfield SW, Fan Y, Morrow WR, Towbin JA, Frazier EA, Nelson DP. 2013. Parvovirus B19 Myocarditis Causes Significant Morbidity and Mortality in Children. Pediatr Cardiol 34:390–397.
7. Kuhl U, Pauschinger M, Seeberg B, Lassner D, Noutsias M, Poller W, Schultheiss H-P. 2005. Viral Persistence in the Myocardium Is Associated With Progressive Cardiac Dysfunction. Circulation 112:1965–1970.
8. Woernle CH, Anderson LJ, Tattersall P, Davison JM. 1987. Human Parvovirus B19 Infection During Pregnancy. J Infect Dis 156:17–20.
9. Anand A, Gray ES, Brown T, Clewley JP, Cohen BJ. 1987. Human Parvovirus Infection in Pregnancy and Hydrops Fetalis. N Engl J Med 316:183–186.
10. Saarinen U, Chorba T, Tattersall P, Young N, Anderson L, Palmer E, Coccia P. 1986. Human parvovirus B19-induced epidemic acute red cell aplasia in patients with hereditary hemolytic anemia. Blood 67.

11. Frickhofen N, Abkowitz JL, Safford M, Berry JM, Antunez-de-Mayolo J, Astrow A, Cohen R, Halperin I, King L, Mintzer D, Cohen B, Young NS. 1990. Persistent B19 Parvovirus Infection in Patients Infected with Human Immunodeficiency Virus Type 1 (HIV-1): A Treatable Cause of Anemia in AIDS. Ann Intern Med 113:926.
12. Pattison JR, Jones SE, Hodgson J, Davis LR, White JM, Stroud CE, Murtaza L. 1981. Parvovirus infections and hypoplastic crisis in sickle-cell anaemia. Lancet (London, England) 1:664–5.
13. Rao SP, Miller ST, Cohen BJ. 1992. Transient Aplastic Crisis in Patients With Sickle Cell Disease. Am J Dis Child 146:1328.
14. Kurtzman GJ, Ozawa K, Cohen B, Hanson G, Oseas R, Young NS. 1987. Chronic Bone Marrow Failure Due to Persistent B19 Parvovirus Infection. N Engl J Med 317:287–294.
15. Kurtzman G, Meyers P, Cohen B, Amunullah A, Young N. 1988. Persistence B19 Parvovirus infection as a cause of severe chronic aneamia in children with acute lymphocytic leukaemia. Lancet 332:1159–1162.
16. Filippone C, Zhi N, Wong S, Lu J, Kajigaya S, Gallinella G, Kakkola L, Söderlund-Venermo M, Young NS, Brown KE. 2008. VP1u phospholipase activity is critical for infectivity of full-length parvovirus B19 genomic clones. Virology 374:444–452.
17. Kaufmann B, Simpson AA, Rossmann MG. 2004. The structure of human parvovirus B19. Proc Natl Acad Sci U S A 101:11628–33.
18. Brown KE, Anderson SM, Young NS. 1993. Erythrocyte P antigen: cellular receptor for B19 parvovirus. Science 262:114–7.
19. Munakata Y, Saito-Ito T, Kumura-Ishii K, Huang J, Kodera T, Ishii T, Hirabayashi Y, Koyanagi Y, Sasaki T. 2005. Ku80 autoantigen as a cellular coreceptor for human parvovirus B19 infection. Blood 106:3449–56.
20. Weigel-Kelley KA, Yoder MC, Srivastava A, Guichard J, Leroy C, Tchernia G, Coulombel L. 2003. Alpha5beta1 integrin as a cellular coreceptor for human parvovirus B19: requirement of functional activation of beta1 integrin for viral entry. Blood 102:3927–33.
21. Bönsch C, Zuercher C, Lieby P, Kempf C, Ros C. 2010. The globoside receptor triggers structural changes in the B19 virus capsid that facilitate virus internalization. J Virol 84:11737–46.

22. Bonsch C, Kempf C, Ros C. 2008. Interaction of Parvovirus B19 with Human Erythrocytes Alters Virus Structure and Cell Membrane Integrity. J Virol 82:11784–11791.
23. Kaufmann B, Chipman PR, Kostyuchenko VA, Modrow S, Rossmann MG. 2008. Visualization of the externalized VP2 N termini of infectious human parvovirus B19. J Virol 82:7306–12.
24. Gigler A, Dorsch S, Hemauer A, Williams C, Kim S, Young NS, Zolla-Pazner S, Wolf H, Gorny MK, Modrow S. 1999. Generation of neutralizing human monoclonal antibodies against parvovirus B19 proteins. J Virol 73:1974–9.
25. Ros C, Baltzer C, Mani B, Kempf C. 2006. Parvovirus uncoating in vitro reveals a mechanism of DNA release without capsid disassembly and striking differences in encapsidated DNA stability. Virology 345:137–147.
26. Bonsch C, Zuercher C, Lieby P, Kempf C, Ros C. 2010. The Globoside Receptor Triggers Structural Changes in the B19 Virus Capsid That Facilitate Virus Internalization. J Virol 84:11737–11746.
27. Leisi R, Ruprecht N, Kempf C, Ros C. 2013. Parvovirus B19 uptake is a highly selective process controlled by VP1u, a novel determinant of viral tropism. J Virol 87:13161–7.
28. Lowin T, Raab U, Schroeder J, Franssila R, Modrow S. 2005. Parvovirus B19 VP2-Proteins Produced in Saccharomyces cerevisiae: Comparison with VP2-Particles Produced by Baculovirus-Derived Vectors. J Vet Med Ser B 52:348–352.
29. Kaufmann B, López-Bueno A, Mateu MG, Chipman PR, Nelson CDS, Parrish CR, Almendral JM, Rossmann MG. 2007. Minute virus of mice, a parvovirus, in complex with the Fab fragment of a neutralizing monoclonal antibody. J Virol 81:9851–8.
30. Organtini LJ, Lee H, Iketani S, Huang K, Ashley RE, Makhov AM, Conway JF, Parrish CR, Hafenstein S. 2016. Near-Atomic Resolution Structure of a Highly Neutralizing Fab Bound to Canine Parvovirus. J Virol 90:9733–9742.
31. Wikoff WR, Wang G, Parrish CR, Cheng RH, Strassheim ML, Baker TS, Rossmann MG. 1994. The structure of a neutralized virus: canine parvovirus complexed with neutralizing antibody fragment. Structure 2:595–607.
32. McCraw DM, O'Donnell JK, Taylor KA, Stagg SM, Chapman MS. 2012. Structure of adeno-associated virus-2 in complex with neutralizing monoclonal antibody A20. Virology 431:40–49.

33. Gurda BL, Raupp C, Popa-Wagner R, Naumer M, Olson NH, Ng R, McKenna R, Baker TS, Kleinschmidt JA, Agbandje-McKenna M. 2012. Mapping a neutralizing epitope onto the capsid of adeno-associated virus serotype 8. J Virol 86:7739–51.
34. Guo F, Jiang W. 2014. Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses, p. 401–443. *In* Methods in molecular biology (Clifton, N.J.).
35. Misasi J, Gilman MSA, Kanekiyo M, Gui M, Cagigi A, Mulangu S, Corti D, Ledgerwood JE, Lanzavecchia A, Cunningham J, Muyembe-Tamfun JJ, Baxa U, Graham BS, Xiang Y, Sullivan NJ, McLellan JS. 2016. Structural and molecular basis for Ebola virus neutralization by protective human antibodies. Science 351:1343–6.
36. Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, Chapman MS. 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A 99:10405–10.
37. Summerford C, Samulski RJ. 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J Virol 72:1438–45.
38. Samulski RJ, Summerford C, Bartlett JS. 1999. αVβ5 integrin: a co-receptor for adeno-associated virus type2 infection. Nat Med 5:78–82.
39. Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, Jae LT, Wosen JE, Nagamine CM, Chapman MS, Carette JE. 2016. An essential receptor for adeno-associated virus infection. Nature 530:108–112.
40. Wobus CE, Hügle-Dörr B, Girod A, Petersen G, Hallek M, Kleinschmidt JA. 2000. Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. J Virol 74:9281–93.
41. O'Donnell J, Taylor KA, Chapman MS. 2009. Adeno-associated virus-2 and its primary cellular receptor—Cryo-EM structure of a heparin complex. Virology 385:434–443.
42. Kaufmann B, Baxa U, Chipman PR, Rossmann MG, Modrow S, Seckler R. 2005. Parvovirus B19 does not bind to membrane-associated globoside in vitro. Virology 332:189–198.
43. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS, Carragher B. 2005. Automated molecular microscopy: The new Leginon system. J Struct Biol 151:41–60.

44. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods 10:584–590.
45. Mindell JA, Grigorieff N. 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol 142:334–347.
46. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. 2007. EMAN2: An extensible image processing suite for electron microscopy. J Struct Biol 157:38–46.
47. Scheres SHW. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530.
48. Rosenthal PB, Henderson R. 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. J Mol Biol 333:721–745.
49. Emsley P, Cowtan K. 2004. *Coot*: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 60:2126–2132.
50. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr 66:213–221.

CHAPTER 3 A STRUCTURAL STUDY ON THE INTERACTION BETWEEN RHINOVIRUS-C AND ITS RECEPTOR

3.1 Chapter Abstract

Rhinovirus C (RV-C) is a picornavirus, which was discovered in 2006 and has been associated with childhood asthmas. In a genome wide association study (GWAS), a single nucleotide mutation in cadherin related family member 3 (CDHR3) was identified to have significant impact on the chance that a child is hospitalized because of asthma exacerbation. Further studies suggested that CDHR3 serves as a cellular receptor for RV-C. This single nucleotide polymorphism influences the localization of CDHR3 and therefore alters an individual's susceptibility to RV-C infection.

To understand the interaction between RV-C and CDHR3, we have determined the structure of RV-C complexed with the extracellular cadherin-like domain 1-3 (EC1-3) of CDHR3 using cryo-EM single particle reconstruction. The initial map of the complex was at a resolution of 3.9 Å. Using 3D classification, RV-C particles with more bound receptors were selected. After refinements using these selected particles, a final map with prominent density for CDHR3 at a resolution of 5.1 Å was achieved. The structure showed that the EC1 of CDHR3 binds to VP3 and VP2 close to the 3-fold axes on the surface of RV-C.

3.2 Introduction

Rhinoviruses are a group of single stranded RNA viruses that belong to *Picornaviridae*. Similar to other picornaviruses, rhinoviruses have small, non-enveloped, icosahedral capsids with diameters of approximately 300 Å.

Rhinoviruses cause common colds and are among the most widespread human pathogens. The first crystallographic structure of a picornavirus, which is also the first of an animal virus, was determined for human rhinovirus 14 (HRV14) in 1985 (1). Structures of many picornaviruses have since been determined by either X-ray crystallography (2–4) or cryo-EM.

Rhinoviruses (RVs) can be further divided into three groups: RV-A, RV-B and the newly identified RV-C (5, 6). RV-C has been shown to be related to some respiratory diseases, especially the severe exacerbations of asthma in children (7–10).

The structure of RV-C (strain 15) has recently been determined by cryo-EM (Figure 3.1) (11). RV-C adopts an overall similar architecture to RV-A and RV-B. Three capsid proteins, VP1, VP2 and VP3 form a quasi T=3 icosahedral shell, with a fourth minor protein VP4 located on the inner surface. There is a plateau around each 5-fold vertex which is surrounded by depressions known as the "canyon". The structure of RV-C differs from other rhinoviruses primarily in two aspects. The surface of RV-C is spiky with finger-like extrusions between VP1 and VP2. This "finger" region is highly antigenic. The hydrophobic pocket, which exists in the VP1 proteins under the floor of the "canyon" in most of other rhinoviruses, is collapsed in RV-C and cannot accommodate any "pocket factor" (a short lipid).

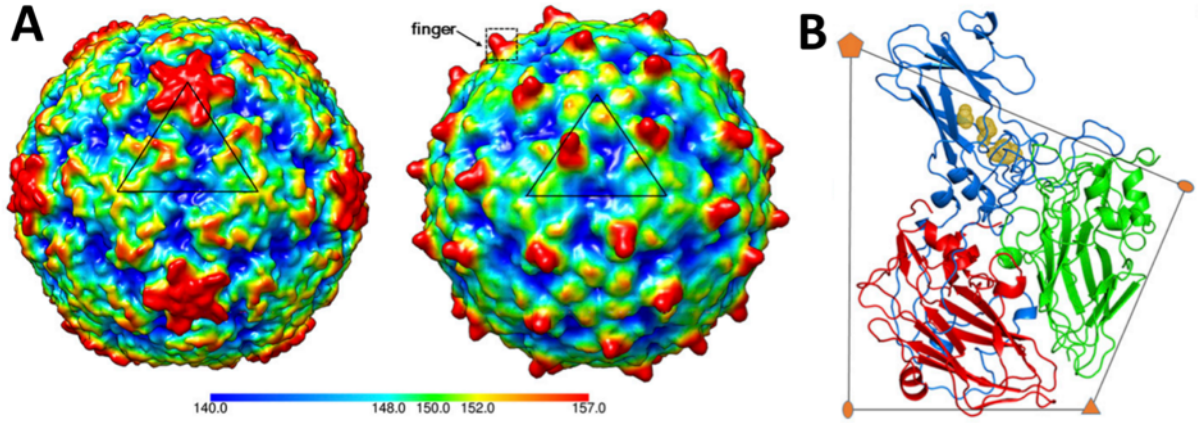

Figure 3.1 The cryo-EM structure of RV-C.
A. A comparison between the structures of HRV-B14 (PDB-4RHV, left) and RV-C (PDB-5K0U, right). The surface rendered representations are shown for the low-pass filtered maps. The surface is colored by distance to the center. B. VP1 (blue), VP2 (green), and VP3 (red) within an asymmetric unit are shown. The hydrophobic pocket, which is commonly present in VP1 under the canyon floor, is collapsed in RV-C (the collapsed volume is colored in gold). (Reprinted from ref. 11)

The major class of rhinoviruses use the intercellular adhesion molecule 1 (ICAM-1) as their receptor (12, 13), and the minor group uses the low-density lipoprotein receptor (LDLR) as the receptor (14). However, neither ICAM-1 nor LDLR are used by RV-C. In a genome wide association study (GWAS) on childhood asthma exacerbations, cadherin-related family member 3 (CDHR3) was identified as a susceptibility gene (15). CDHR3 was later shown to act as a receptor for RV-C and therefore mediates asthma exacerbations caused by RV-C infections (16).

Cadherins are a class of glycoproteins located on the surface of cells that contain multiple repeats of extracellular cadherin-like (EC) domains (17–20). An EC domain has about 110 amino acids and has a β-strand-based fold that resembles immunoglobulin domains. Cadherin is a large superfamily of cell surface receptors and plays an important role in cell-to-cell recognition and adhesion. CDHR3 is a cadherin-like membrane protein expressed mainly in lungs and airway epithelium. CDHR3 consists of a small cytoplasmic domain, a single transmembrane helix and six tandemly connected extracellular cadherin-like domains. The single amino acid mutation identified in the GWAS (C529 to Y529) occurs in the fifth extracellular domain and shifts the localization of CDHR3 towards the cellular surface (Figure 3.2).

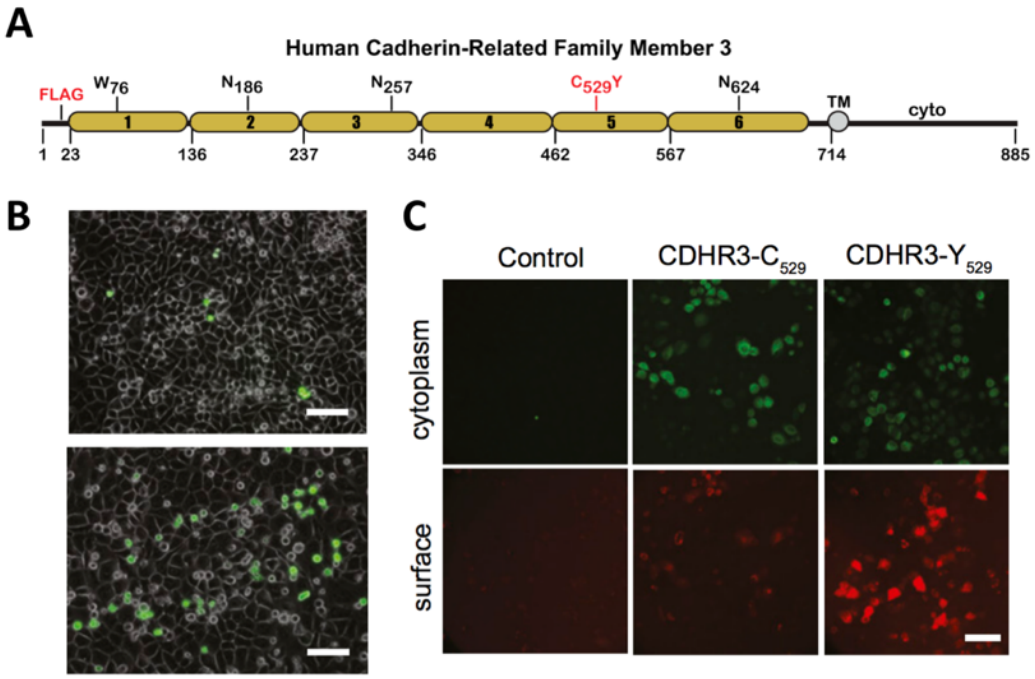

Figure 3.2 CDHR3 as a receptor of RV-C.
A. The organization of human CDHR3 construct used in the study. B. Infectivity of RV-C in HeLa cells expressing CDHR3^{C529} (upper) or CDHR3^{Y529} (lower) indicated by GFP expression. C. The localizations of CDHR3 in HeLa cells. (Reprinted from ref. 16)

Initially, it was shown that EC1 and EC2 were necessary for the binding between CDHR3 and RV-C (figure 3.3B) in immunoprecipitation assays. However, when Watters et al. used recombinant CDHR3 proteins expressed in *E. coli*, EC1+3 (Δ2) also showed binding affinity with RV-C viruses (figure 3.3C). It appeared that EC2 helps the correct folding of EC1 in HeLa cells and only EC1 is necessary for the binding.

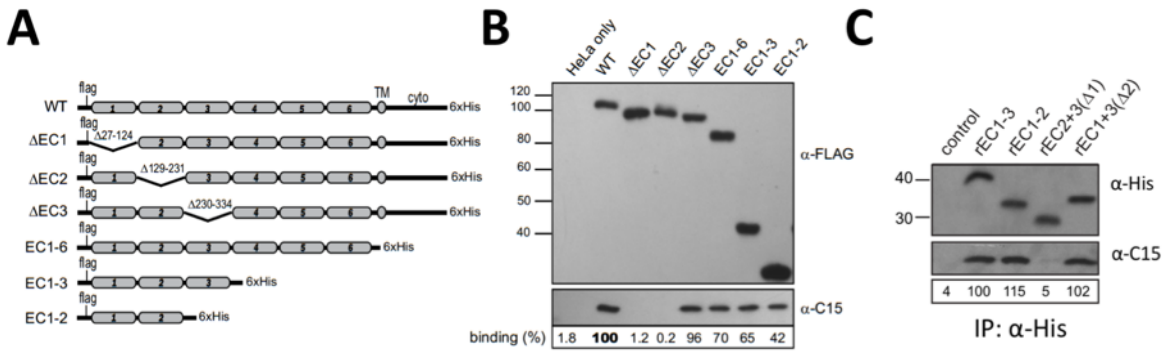

Figure 3.3 The binding between RV-C and different constructs of CDHR3.
A. A cartoon showing the domain organization of different forms of CDHR3 constructs. B. The binding between different domains of CDHR3 to RV-C assayed by immunoprecipitation in HeLa cell lysates. C. The binding between RV-C and different CDHR3 proteins refolded from *E. coli* inclusion bodies. (Unpublished data, figures provided by Ann C. Palmenberg and Kelly Watters.)

To investigate the binding between RV-C and CDHR3, we have determined the structure of RV-C complexed with the extracellular cadherin domains (EC1-3) of CDHR3. Although the densities of the receptors were poorer compared to that of the capsid, EC1 was clearly visible and seen to be interacting with RV-C in the region between the "fingers" and the 3-fold vertices. The following fitting results suggested that a potential receptor binding site is formed by a cleft in VP3 and the BC and the HI loops of VP2.

3.3 Results and Discussions

3.3.1 The cryo-EM structure of the complex of RV-C and CDHR3.

To characterize the receptor binding of RV-C, three N-terminal extracellular cadherin-like (EC1-3) domains of human CDHR3 were expressed and purified. Purified RV-C (strain 15a) virions from CDHR3-expressing HeLa cells were mixed with EC1-3 proteins at a molecular ratio of approximately 1:20 (20 receptors per VP1-4 protomer) to achieve a better occupancy. The structure of the complex was then studied by cryo-EM single particle analysis.

A total of 3570 images were collected and 23,826 particles were boxed. After two rounds of 2D classifications using the program RELION (21), 19,876 particles were selected for 3D refinement. The structure of the native full RV-C virion (EMD-8189) was low-pass filtered and was used as the initial model. Using the *jspr* (22) software package, the icosahedral reconstruction of the complexes was refined to 3.9 Å resolution when all the particles were used.

In this initial cryo-EM map of the receptor bound RV-C, additional densities were found close to the 3-fold axes compared to the map of the native RV-C (EMD-8189) (figure 3.4A&B). However, these densities, potentially representing part of the bound EC1-3, were not as easily interpretable as the RV-C capsid proteins. The average density height of theses CDHR3 densities was only about one fifth of the capsid density and the connectivity was poor. This was potentially caused by both a low occupancy and additional flexibility due to the other two domains (EC2-3) that could move around freely.

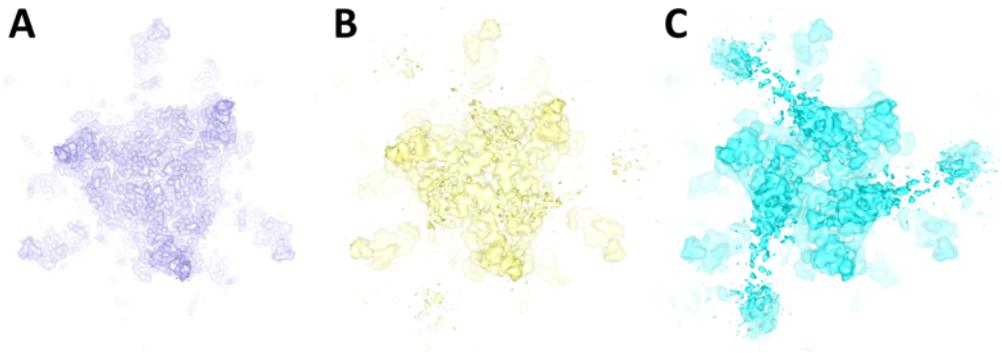

Figure 3.4 A comparison of the densities adjacent to a 3-fold vertex on RV-C.
Native RV-C (EMd-8189) (A), an initial reconstruction of the RV-C complex with CDHR3 EC1-3 (B) and a reconstruction using only one class (C) are shown as surface rendered maps, with a 3-fold axis perpendicular to the plane of image.

A 3D classification process that was focused in the receptor region was then performed to select particles that had more bound receptors (figure 3.5). A minor class with about 5,000 particles had stronger density for CDHR3 than did the other classes (figure 3.5D) and was kept for further 3D refinement (Figure 3.5E). The resulting map using these particles has an improved density of CDHR3 (figure 3.4C) and has 5.1 Å resolution. This map, with 5,000 selected particles, was used for the structural analysis of the interaction between RV-C and CDHR3 EC1.

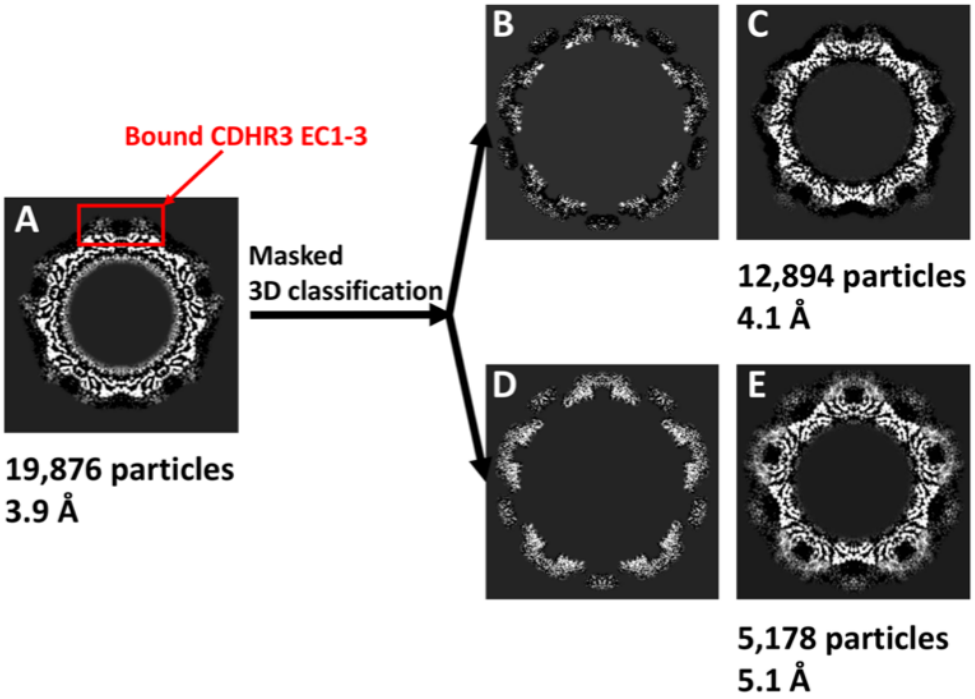

Figure 3.5 Masked 3D classification to improve quality of the density for CDHR3. Sections of electron density maps along a 5-fold axis are shown. A. The reconstruction with all particles selected after 2D classification. B&D. Masked maps after 3D classification without updating orientations. One of the classes had significantly stronger density for CDHR3. C&E. Reconstructions of each class after refinement.

3.3.2 Interpretation of the complex structure

The structure of RV-C when complexed with CDHR3 is very similar to the native virion (11), with 60 copies of VP1-VP4 (no VP4 in emptied particles) forming an icosahedral shell. Extra densities can be found between the "fingers" and close to the 3-fold axes (Figure 3.6).

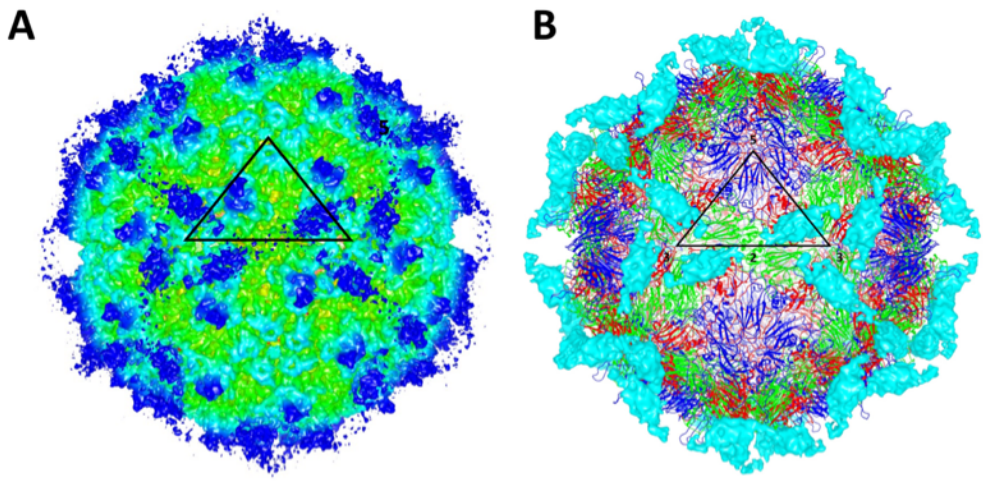

Figure 3.6 The structure of RV-C complexed with CDHR3.
A. A surface rendered cryo-EM map of the complex with the 2-fold axis perpendicular to the plane of the figure. The surface is colored by radius. (Blue: >160 Å; cyan: 140-150 Å; green: 120-140 Å; yellow: 110-120 Å) The black triangle outlines one asymmetric unit. B. The difference map between the complex and RV-C alone (EMD-8189). The difference density (cyan) is superimposed with the model of full RV-C capsid (PDB-5K0U; blue: VP1; green: VP2; red: VP3).

A difference map was calculated by subtracting the capsid densities (EMD-8189) from the map of the RV-C and CDHR3 complex. Both maps had been low-pass filtered to 5 Å prior to the subtraction. The remaining densities are sixty symmetrical equivalent blobs on the surface of RV-C capsid, about 150 Å from the center of the virus. Each of these density blob has the size and shape corresponding to one cadherin-like domain. A predictive model of EC1 of human CDHR3 (16) was fitted into the density using EMfit (23) (Figure 3.7). Although the orientation of bound EC1 could not be accurately determined, the results of EMfit showed that the C-terminus of EC1 was pointed at the 2-fold axes. The residual density at the N-terminus agreed well with the fact that the EC1-3 construct used for structural determination had a flag peptide (DYKDDDDK) tagged to the N-terminus of the protein.

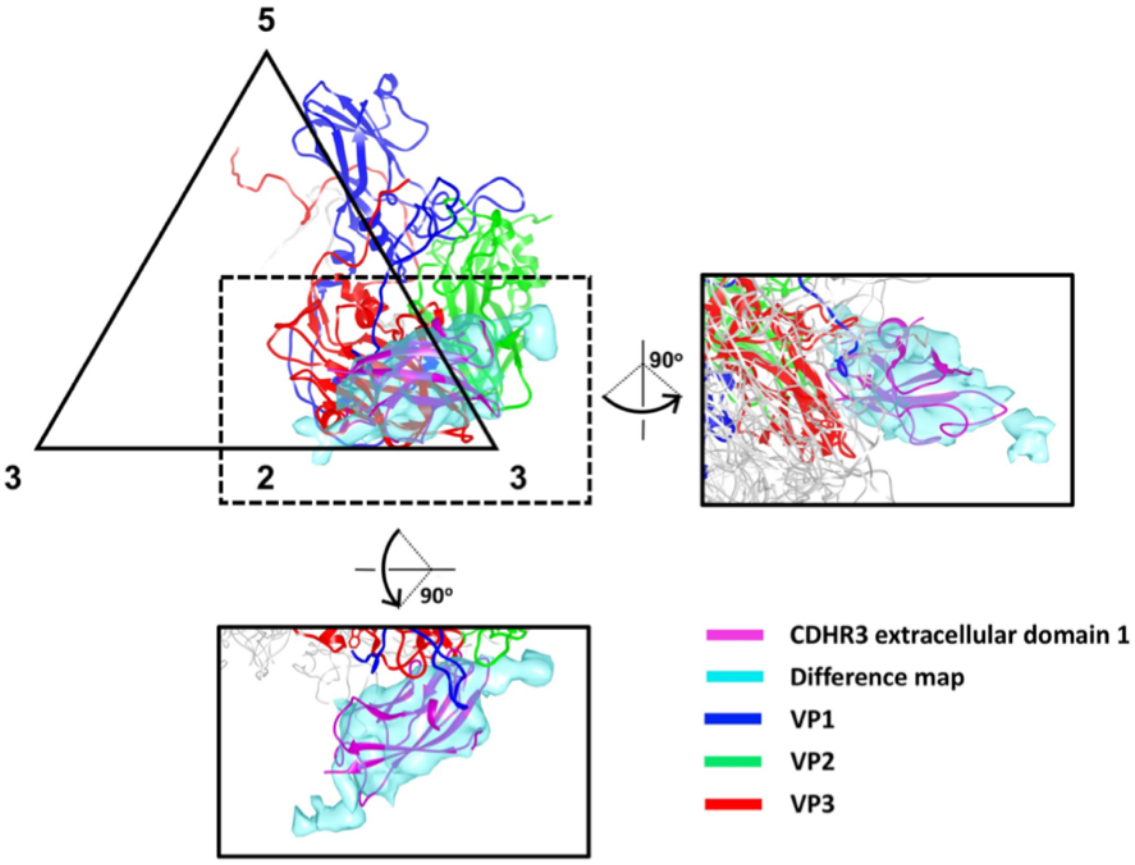

Figure 3.7 The fitting of a predicted model of CRHR3 EC1 into the difference map with the N-terminus oriented toward the 3-fold axis.
The triangle shows the relative position of the protomer inside an asymmetric unit. A predicted model of human CDHR3 EC1 (magenta) is placed into the difference density (cyan).

3.3.3. The receptor binding site on RV-C

A "roadmap" was calculated to show the footprint of EC1 on the surface of RV-C (Figure 3.8). Different from other picornaviruses which bind to ICAM-1 (e.g. HRV14, coxsackievirus A), CD155 (e.g. poliovirus-1) or CAR (e.g. coxsackievirus B) through the "canyon" around the 5-fold axes (24–28), RV-C binds to CDHR3 through a hydrophobic cleft formed by VP3 and VP2 close to the 3-fold axes (Figure 3.9). The β-strands B (70-ITVT-73) and I (203-GTG-205) of VP3 compose the floor and the loop insertions AB (E61), BC (74-KKT-76, E79) and HI (G202) of VP3 form the edges of the cleft. The loop insertions BC (72-QVG-74) and HI (233-LRP-235) of VP2 bind to the N-terminus of EC1. The carboxyl end of VP1 (274-PLI-276) is within the vicinity of the C-terminus of EC1 and may also contribute to the interaction. The amino acids that may participate in the receptor binding are summarized in table 3.1.

Among these residues, K74 and K75 of VP3, 75S, 233L and 234R of VP2 are conserved in different RV-C strains, but not in rhinoviruses A or B. It is anticipated that these residues play more important roles in the recognition of CDHR3.

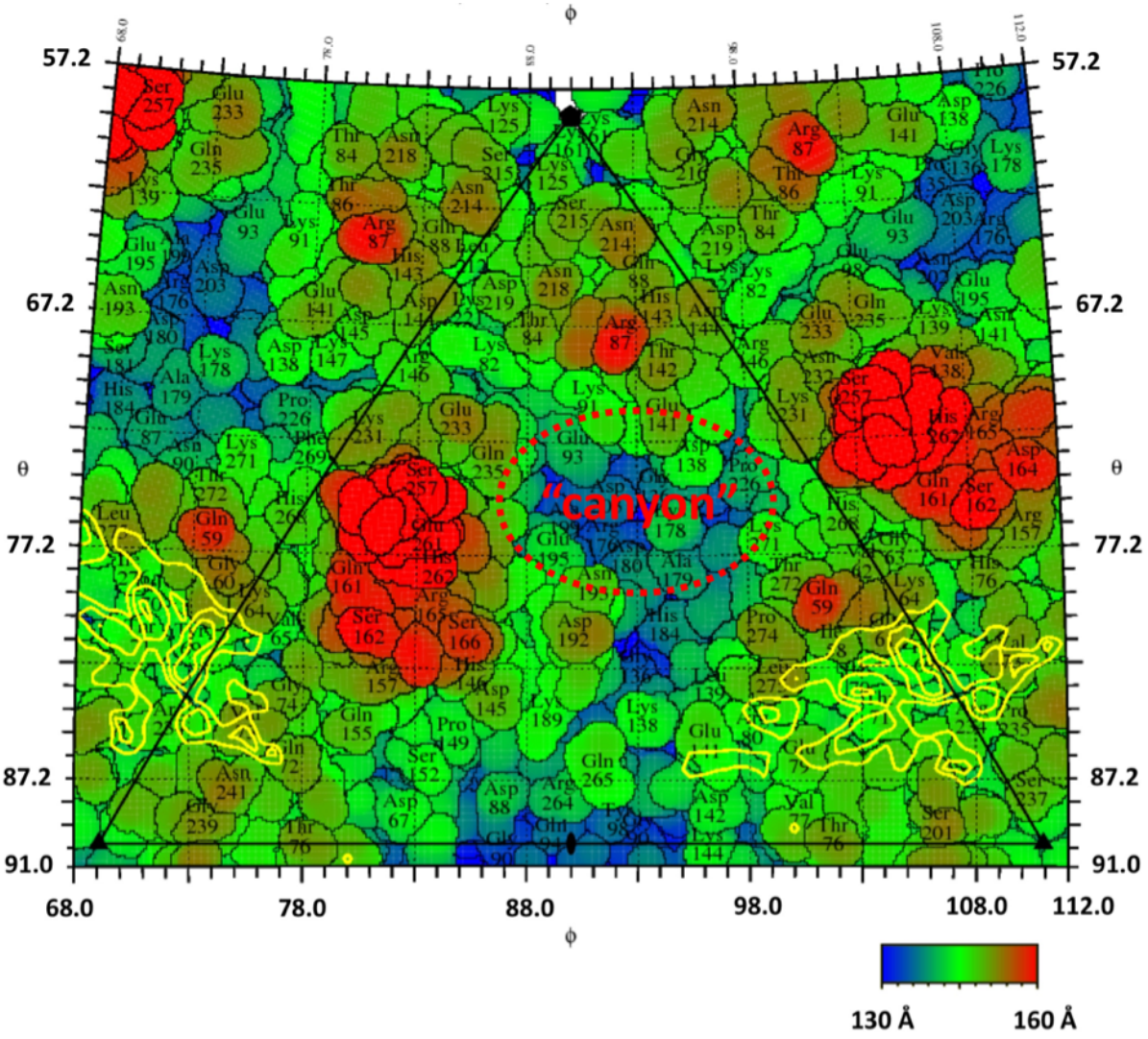

Figure 3.8 The foot print of CDHR3 on the surface of RV-C.
A "roadmap" shows the surface of RV-C within one asymmetric unit (black triangle), colored by distance (Å) to the center of the virus. The numbers on the side represent the angles of the corresponded spot on the surface relative to the 2-fold axis (middle bottom) which is normal to the plane of this page. The yellow contour lines represent the density of the difference map at a height of 150 Å.

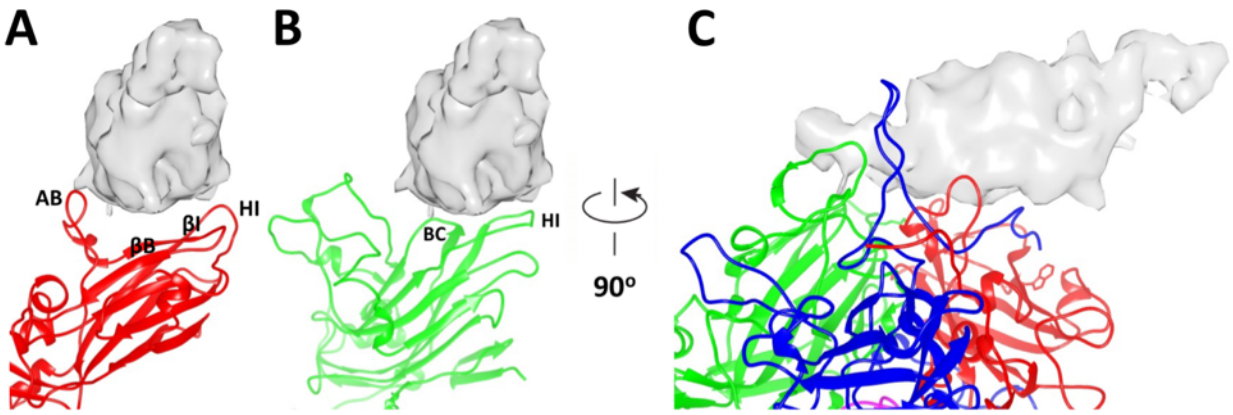

Figure 3.9 The receptor binding site of CDHR3 on RV-C.
A&B. A cleft formed by the β-strands and insertions of VP3 (A, red) and VP2 (B, green) accommodates EC1 of CDHR3. Panel A and B have the same viewing angle with only VP3 (A) or VP2 (B) shown. One of the blobs in the difference map, corresponding to one EC1 molecule, is colored in grey. C. A side view (A&B rotated by 90°) of EC1 bound to one RV-C protomer (VP1: blue, VP2: green, VP3: red, VP4: magenta).

Table 3.1 RV-C residues that are potentially involved in the binding CDHR3.

VP1	L275, A278
VP2	V73, G74, S75; L233, R234, P235
VP3	Q59, G60, E61; K74, K75, T76, E79, R80; G202, G203, T204

These results suggest that RV-C adopts a receptor binding strategy that is different from other rhinoviruses. A cleft formed mainly by VP3 and two loop insertions in VP2 (BC and HI) together form the receptor binding site on the surface of RV-C and binds to the N-terminal domain, EC1, of CDHR3. Similar to all cadherin-like proteins, the correct folding of CDHR3 depends on calcium ions that bind to the interface between the cadherin-like extracellular domains. The position 529 is located between EC5 and EC6 and is involved in Ca^{2+} binding. The replacement of this tyrosine by cysteine causes the misfolding and recycling of CDHR3. The transition of the tyrosine at position 529 to cysteine is probably a strategy that humans have developed during evolution to defend against RV-C.

EC1 has a shape similar to an American football and binds to the viral surface with one side of the football. If the extracellular domains of CDHR3 extend perpendicularly from the cell surface as expected for other cadherins, the binding will cause a big change of the orientation of the CDHR3 N-terminal domains. This may destabilize the conformation of the entire protein and

trigger the endocytosis of the entire complex. Thus, RV-C is able to enter the cell upon attachment and prepare for uncoating in endosomes.

3.4 Methods and Materials

3.4.1 Virus and protein preparation

The preparation of infectious RV-C particles was described in ref. 11. The cadherin-like domain 1-3 (EC1-3) of human CDHR3 was prepared following a previously described protocol (16).

3.4.2 Cryo-electron microscopy

Native RV-C virions (0.1 mg/ml) were mixed with excessive EC1-3 (5-8mg/ml) at a volume ratio of 1:1 and then were incubated at 4^{o}C overnight. Aliquots of 3 μl of the mixture were applied onto glow-discharged lacey carbon grids (400 mesh copper, lot #200617, Ted Pella INC.). Then the grids were blotted for 3 seconds and plunged into liquid ethane cooling by liquid nitrogen using a CP3 plunger (Gatan). The girds were transferred to a Titan Krios (FEI) operated at 300 kV for data collection. A total of 3570 movies were collected with 0.2s frames using a K2 summit detector (Gatan, 3710 x 3838), at a dose rate of 8 e^{-}/s/pixel. The nominal magnification was 81000 and the corresponding pixel size was 1.73 Å. The total dose for each movie was 32 e^{-}/ $Å^{2}$.

3.4.3 Image processing

Beam-induced motion within frames were corrected by Motioncor2 (29). Defocuses of the motion corrected images were estimated using CTFFIND4 (30). A template matching particle selection was done using FindEM (31) that is incorporated in the *appion* pipeline (32). A total of 23,826 particles were selected using two projections of a EM map of native RV-C as the template. Two rounds of non-reference 2D classification were carried out with RELION/1.4 (21). Classes that did not produce an average with RV-C features were discarded and 19,876 particles were left. The dataset was hereafter divided into two half subsets to comply with the Gold standard resolution evaluation. All 3D processing and refinements were performed using *jspr* (22) assuming icosahedral symmetry. An initial refinement all the particles achieved a map at a resolution of 4.5 Å which was later improved to 3.9 Å with proper masking and weighting in 3D reconstructions.

A focused 3D classification was performed subsequently to improve the density of CDHR3. In each iteration of the classification, three different references were first multiplied by a 3D mask and only the region of bound receptors was left. Then the particles were sorted using projection matching without updating orientations, centers, scales, etc. Two major classes were achieved with distinct occupancies of receptor binding. The class with a much stronger CDHR3 density has 5,178 particles. It is important to note that the classification was performed within each half dataset independently. The resulting classes in each half were similar but had slightly different number of particles. The further 3D refinement with these selected 5,178 particles led to a lower resolution at 5.1 Å but a better density of bound receptors.

3.4.4 Generation of the difference map and fitting

The map for native RV-C (11, EMD-8189) was low-pass filtered to 5 Å. This filtered map was then subtracted from the map of the receptor bound RV-C to obtain a difference map. The subtraction was accomplished in *chimera* (33) using a scaling factor of 6.5 on the complex map. The difference was then low-pass filtered to 6 Å to facilitate structure interpretation.

The coordinates of a predicted model of EC1 of human CDHR3 were fitted into the 6 Å difference map using *EMfit* (23). A step size of 10^{o} was used in the global angular search and local searches were performed based on each of the top 5 fitting results in the global search ("climbing" in *EMfit*). The best fitting after the local searches was not significantly better than the runner-ups due to the lack of high resolution features of the difference map. The top three results returned all had *sumF* values around 0.4.

3.5 Future plan

The study on RV-C's interaction with CDHR3 is an ongoing project and more need to be accomplished to address different aspects of the problem.

Cadherins are a group of proteins on cell surface mediating cell-to-cell interactions (19). Both *cis*- and *trans*- interactions exist between the first and second domains of cadherin extracellular segment. This agrees with our observation using EM that EC1-3 of CDHR3 formed aggregations, which certainly reduced the amount of protein that was capable of binding to RV-C particles.

To overcome the low occupancy caused by cadherin polymerization, more truncated forms of CDHR3 were tested. EC1 and EC1-2 were expressed successfully and exhibited a better binding

with RV-C virions. These two forms of proteins will be used for structural determinations aimed at a better understanding of the receptor binding of RV-C.

3.6 Chapter Acknowledgements

This chapter includes unpublished data. The maps and models have not been deposited to any databases. As a result, a delay of publication has been required for this thesis.
This work is a collaboration with Ann C. Palmenberg at University of Wisconsin, Madison.

3.7 References

1. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H-J, Johnson JE, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153.
2. Liu Y, Sheng J, Fokine A, Meng G, Shin W-H, Long F, Kuhn RJ, Kihara D, Rossmann MG. 2015. Structure and inhibition of EV-D68, a virus that causes respiratory illness in children. Science 347:71–4.
3. Hogle JM, Chow M, Filman DJ. 1985. Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–65.
4. Plevka P, Perera R, Cardosa J, Kuhn RJ, Rossmann MG. 2012. Crystal structure of human enterovirus 71. Science 336:1274.
5. Lamson D, Renwick N, Kapoor V, Liu Z, Palacios G, Ju J, Dean A, St. George K, Briese T, Ian Lipkin W. 2006. MassTag Polymerase‐Chain‐Reaction Detection of Respiratory Pathogens, Including a New Rhinovirus Genotype, That Caused Influenza‐Like Illness in New York State during 2004–2005. J Infect Dis 194:1398–1402.
6. Arden KE, McErlean P, Nissen MD, Sloots TP, Mackay IM. 2006. Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections. J Med Virol 78:1232–1240.
7. Khetsuriani N, Lu X, Teague WG, Kazerouni N, Anderson LJ, Erdman DD. 2008. Novel human rhinoviruses and exacerbation of asthma in children. Emerg Infect Dis 14:1793–6.
8. Bizzintino J, Lee W-M, Laing IA, Vang F, Pappas T, Zhang G, Martin AC, Khoo S-K, Cox DW, Geelhoed GC, McMinn PC, Goldblatt J, Gern JE, Le Souëf PN. 2011. Association between human rhinovirus C and severity of acute asthma in children. Eur Respir J 37:1037–42.
9. Calvo C, Garcia ML, Pozo F, Reyes N, Pérez-Breña P, Casas I. 2009. Role of rhinovirus C in apparently life-threatening events in infants, Spain. Emerg Infect Dis 15:1506–8.
10. Miller EK, Khuri-Bulos N, Williams J V., Shehabi AA, Faouri S, Al Jundi I, Chen Q, Heil L, Mohamed Y, Morin L-L, Ali A, Halasa NB. 2009. Human rhinovirus C associated with wheezing in hospitalised children in the Middle East. J Clin Virol 46:85–89.

11. Liu Y, Hill MG, Klose T, Chen Z, Watters K, Bochkov YA, Jiang W, Palmenberg AC, Rossmann MG. 2016. Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. Proc Natl Acad Sci U S A 113:8997–9002.
12. Staunton DE, Merluzzi VJ, Rothlein R, Barton R, Marlin SD, Springer TA. 1989. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56:849–853.
13. Greve JM, Davis G, Meyer AM, Forte CP, Yost SC, Marlor CW, Kamarck ME, McClelland A. 1989. The major human rhinovirus receptor is ICAM-1. Cell 56:839–847.
14. Hofer F, Gruenberger M, Kowalski H, Machat H, Huettinger M, Kuechler E, Blaas D. 1994. Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. Proc Natl Acad Sci U S A 91:1839–42.
15. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, den Dekker HT, Husby A, Sevelsted A, Faura-Tellez G, Mortensen LJ, Paternoster L, Flaaten R, Mølgaard A, Smart DE, Thomsen PF, Rasmussen MA, Bonàs-Guarch S, Holst C, Nohr EA, Yadav R, March ME, Blicher T, Lackie PM, Jaddoe VW V, Simpson A, Holloway JW, Duijts L, Custovic A, Davies DE, Torrents D, Gupta R, Hollegaard M V, Hougaard DM, Hakonarson H, Bisgaard H. 2014. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet 46:51–55.
16. Bochkov YA, Watters K, Ashraf S, Griggs TF, Devries MK, Jackson DJ, Palmenberg AC, Gern JE. 2015. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. Proc Natl Acad Sci U S A 112:5485–90.
17. Patel SD, Ciatto C, Chen CP, Bahna F, Rajebhosale M, Arkus N, Schieren I, Jessell TM, Honig B, Price SR, Shapiro L. 2006. Type II cadherin ectodomain structures: implications for classical cadherin specificity. Cell 124:1255–68.
18. Patel SD, Chen CP, Bahna F, Honig B, Shapiro L. 2003. Cadherin-mediated cell-cell adhesion: sticking together as a family. Curr Opin Struct Biol 13:690–8.
19. Brasch J, Harrison OJ, Honig B, Shapiro L. 2012. Thinking outside the cell: how cadherins drive adhesion. Trends Cell Biol 22:299–310.

20. Ciatto C, Bahna F, Zampieri N, VanSteenhouse HC, Katsamba PS, Ahlsen G, Harrison OJ, Brasch J, Jin X, Posy S, Vendome J, Ranscht B, Jessell TM, Honig B, Shapiro L. 2010. T-cadherin structures reveal a novel adhesive binding mechanism. Nat Struct Mol Biol 17:339–47.
21. Scheres SHW. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530.
22. Guo F, Jiang W. 2014. Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses, p. 401–443. *In* Methods in molecular biology (Clifton, N.J.).
23. Rossmann MG, Bernal R, Pletnev S V. 2001. Combining Electron Microscopic with X-Ray Crystallographic Structures. J Struct Biol 136:190–200.
24. Baggen J, Hurdiss DL, Zocher G, Mistry N, Roberts RW, Slager JJ, Guo H, van Vliet ALW, Wahedi M, Benschop K, Duizer E, de Haan CAM, de Vries E, Casasnovas JM, de Groot RJ, Arnberg N, Stehle T, Ranson NA, Thibaut HJ, van Kuppeveld FJM. 2018. Role of enhanced receptor engagement in the evolution of a pandemic acute hemorrhagic conjunctivitis virus. Proc Natl Acad Sci U S A 115:397–402.
25. Xiao C, Bator-Kelly CM, Rieder E, Chipman PR, Craig A, Kuhn RJ, Wimmer E, Rossmann MG. 2005. The Crystal Structure of Coxsackievirus A21 and Its Interaction with ICAM-1. Structure 13:1019–1033.
26. Rossmann MG, He Y, Kuhn RJ. 2002. Picornavirus-receptor interactions. Trends Microbiol 10:324–31.
27. Kolatkar PR, Bella J, Olson NH, Bator CM, Baker TS, Rossmann MG. 1999. Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. EMBO J 18:6249–59.
28. He Y, Chipman PR, Howitt J, Bator CM, Whitt MA, Baker TS, Kuhn RJ, Anderson CW, Freimuth P, Rossmann MG. 2001. Interaction of coxsackievirus B3 with the full length coxsackievirus-adenovirus receptor. Nat Struct Biol 8:874–878.
29. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14:331–332.
30. Mindell JA, Grigorieff N. 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol 142:334–347.

31. Roseman A. 2004. FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. J Struct Biol 145:91–99.
32. Lander GC, Stagg SM, Voss NR, Cheng A, Fellmann D, Pulokas J, Yoshioka C, Irving C, Mulder A, Lau P-W, Lyumkis D, Potter CS, Carragher B. 2009. Appion: An integrated, database-driven pipeline to facilitate EM image processing. J Struct Biol 166:95–102.
33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

CHAPTER 4 THE PENETRATION OF BACTERIOPHAGE ΦX174 THROUGH BACTERIAL CELL WALLS

4.1 Chapter Abstract

Unlike tailed bacteriophages which use a preformed tail for transporting their genomes into a host bacterium, the ssDNA bacteriophage ΦX174 is tail-less. Using cryo-electron microscopy and time-resolved small angle x-ray scattering, we show that lipopolysaccharides (LPS) form bilayers that interact with ΦX174 at an icosahedral 5-fold vertex and induce the ssDNA genome ejection. The structures of ΦX174 complexed with LPS have been determined for the pre- and post ssDNA ejection states. The ejection is initiated by the loss of the G protein spike that encounters the LPS, followed by conformational changes of two polypeptide loops on the major capsid F proteins. One of these loops mediates viral attachment and the other participates in making the 5-fold channel at the vertex contacting the LPS.

4.2 Introduction

Tailed bacteriophages use their tail complexes for host cell recognition, adsorption and penetration of bacterial cell walls (1–4). However, it remains unclear how tail-less phages like ΦX174 perform tail-associated functions. ΦX174 is a small icosahedral, tail-less bacteriophage with a maximum diameter of about 320 Å. The capsid contains a 5.3 kb circular, single-stranded DNA genome, which encodes four structural proteins: the capsid protein F, the spike protein G, the genome associated protein J and the DNA pilot protein H. The ΦX174 structure has T=1 icosahedral symmetry (figure 4.1A) (5), with 60 F proteins forming the capsid decorated by twelve spikes on the vertices, each containing five G proteins (6).

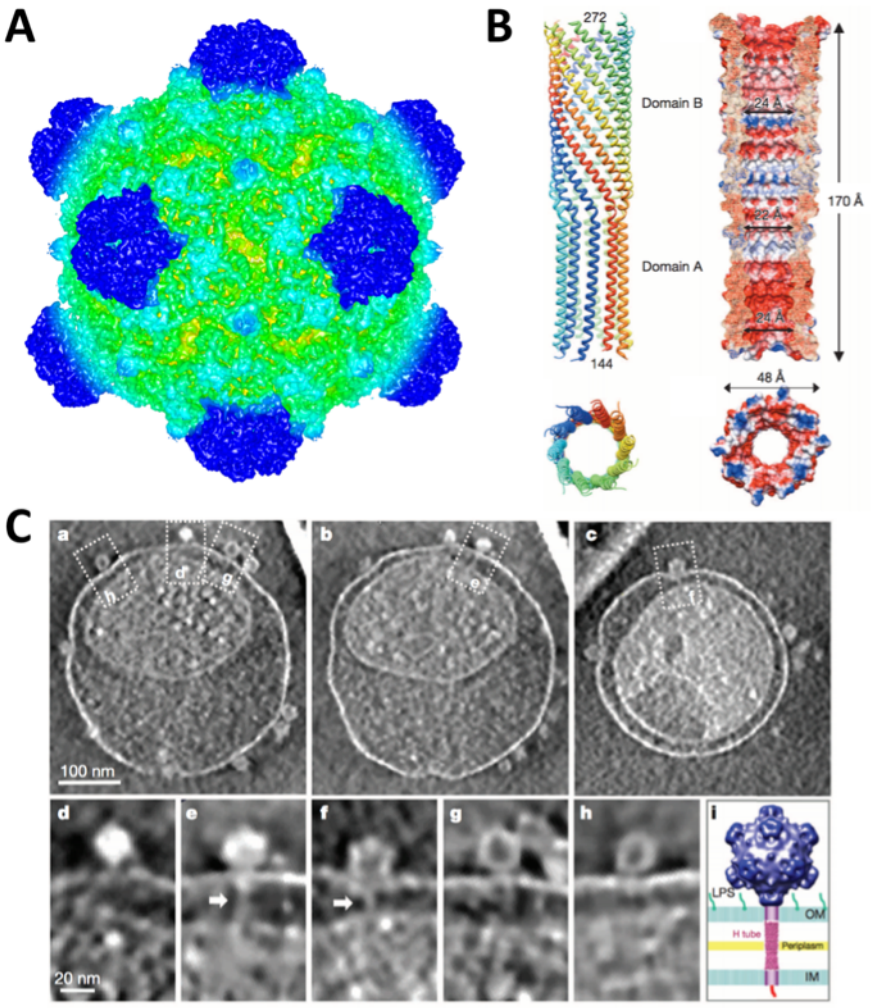

Figure 4.1 The structures of ΦX174 and its "tail".
A. Surface rendered cryo-EM map of ΦX174 colored by distance to the center (green: 120-135Å, cyan: 135-150Å, blue: >150Å). Twelve G spikes are mainly colored in blue. B. The structure of the H tube. Ten coil-coil helices are assembled into a 170 Å-long α-helical barrel. C. Tomographic studies showing ΦX174 in different states while infecting bacterial cells. A tubular structure similar to the "H-tube" is seen penetrating through the host membranes. a-c. Slices of the tomograms, showing phage particles that are infecting the mini cells. d-h. ΦX174 particles in different structural states during infection. i. A model of the "H-tube" penetrating across the cellular wall of a host cell. (Panel B and C are reprinted from ref. 9 with permissions from Springer Nature.)

The major capsid protein F consists of a "jelly roll" fold that has eight antiparallel 🚲 strands (B to I) and seven loop insertions. F proteins bind to G proteins through two of the longest insertions, loop EF (His73-Pro234) and HI (Pro292-Gln401). The N-terminal portion of loop EF forms a bulge on the surface of the capsid between the 5-fold and 2-fold axes. On one side of the "bulge", Tyr158 interacts with the G protein whereas the other side of the "bulge" is mostly

exposed. The majority of known mutations affecting assembly are found in these two loops (7). Overall the F proteins of homologous tail-less phages are highly conserved (8), with the exception of these long loop insertions. Prior to or at the onset of infection, the H proteins assemble into a tube that translocates the genome across the host cell wall (figure 4.1B&C) (9). However, the trigger that initiates this process is unknown.

Due to their potential to trigger the DNA ejection in vitro, lipopolysaccharides (LPS) have been used to study the mechanisms of infection of many tailed bacteriophages (2, 10). ΦX174-like viruses from Microviridae also use LPS as cellular receptors (11–15). LPS molecules are located on the distal leaflet of the outer membrane of Gram-negative bacteria and consist of three major components: hydrophobic lipid A tails, a hydrophilic polysaccharide core (R core), and repeating O-antigenic polysaccharide chains (O-antigen). The oligosaccharide composition varies between different bacterial strains. Both proteins G and H are known to interact with the LPS of suitable hosts in a species-specific manner (12, 14). Mutations affecting host cell recognition and/or the kinetics of DNA ejection have been isolated and found to be in genes F, G and H (16–18).

Here we show that the G and F proteins at one of the 5-fold vertices recognize and interact with an LPS receptor to initiate DNA ejection. Infectious particles were incubated with LPS to identify the structural changes that led to DNA ejection. In vitro, purified LPS molecules formed membrane-like structures with two parallel layers to which ΦX174 particles were attached. The conformational changes of the ΦX174 particles during this reaction were initially characterized by small-angle X-ray scattering (SAXS). The structures of full and emptied ΦX174, complexed with these LPS bilayers, were then determined to about 10 Å resolution by single particle cryo-electron microscopy (cryo-EM). In both cases, one of the G spikes was missing and the virus was interacting with LPS through the exposed EF loops on the F proteins. After the genome had been ejected, the channel formed by the F proteins at the unique 5-fold axis remained open.

4.3 Results

4.3.1. ΦX174 ejects its genome upon LPS treatment

The loss of infectivity was measured to assess the effectiveness of LPS to trigger a response from the virion. After incubation at 33 °C for 20 minutes, the infectivity of the group containing LPS derived from Salmonella typhimurium decreased by over 90% while that of the control group

containing only buffer decreased by 30% (figure 4.2A). To evaluate the specificity of this reaction, LPS from ΦX174 insensitive E. coli strain was tested in a similar manner, but failed to reduce the infectivity. In time course experiments, most of the infectivity was lost during the first 5 minutes (figure 4.2B), which is consistent with in vivo studies (17, 19, 20). Both native and LPS-treated ΦX174 particles were examined by negative stain EM. This showed that over 90% of the particles had lost their genomes after a 20-minute incubation with LPS (figure 4.2 C&D). These results suggest that Salmonella typhimurium derived LPS triggers ΦX174 genome ejection and could be used in the subsequent structural studies.

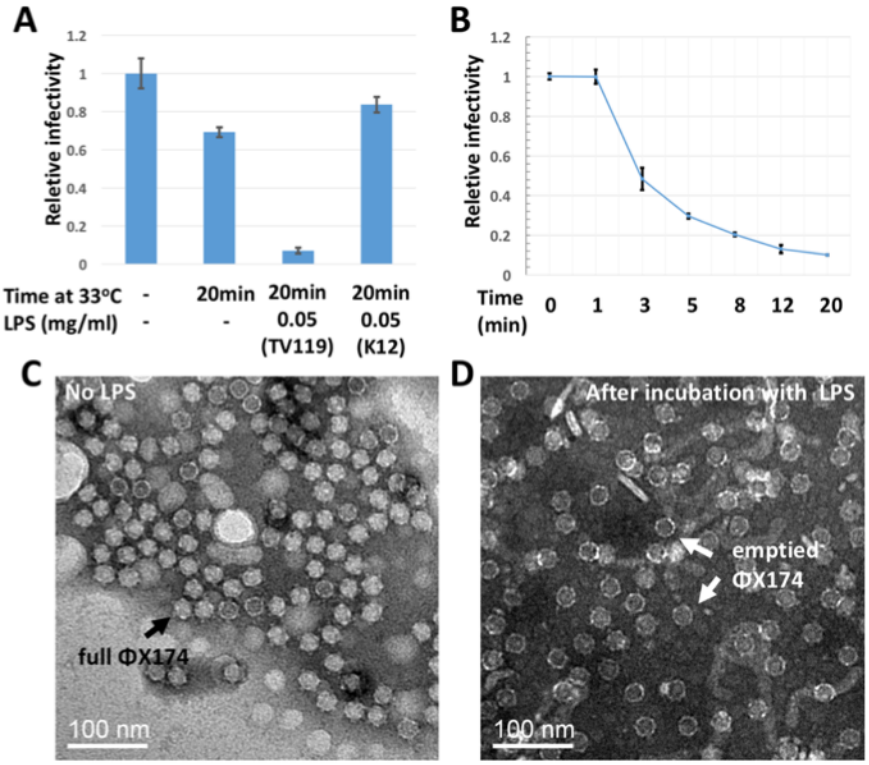

Figure 4.2 Ejection of the ΦX174 genome during incubation with LPS.
A. Infectivity of ΦX174 with a concentration of 109 PFU/ml after incubation with no or specified LPS species. TV119 represents *Salmonella typhimurium* strain TV119 (Ra mutant), and K12 represent E. coli strain K12. B. Normalized infectivity sampled at different time points of the reaction. C&D. Negative stained full infectious ΦX174 before and after incubation with LPS, respectively. Over 80 percent of the ΦX174 had become emptied (white arrows).

4.3.2. ΦX174 particles lose their symmetry and sphericity during the reaction with LPS

SAXS was used to study the LPS-induced conformational response of ΦX174 in solution. Model structures aided in the interpretation of features reported by SAXS. For regularly shaped particles with well-defined dimensions, such as full or empty phages, the scattering profiles display a series of maxima and minima, whose positions are determined by particle structure (21). Theoretical SAXS profiles for both types of particle are shown in figure 4.3A. Their differences are characterized by two features: (1) reduced forward scattering intensity (I(q=0)) of the empty particles relative to full particles, which reflects their lower mass, and (2) a shift towards lower angles in the positions of the maxima and minima, which reflects an increase in radius of gyration (Rg). Rg is determined by the mean distance between all pairs of the electrons in the sample and, hence, is larger for the empty particles. Any loss of rotational symmetry affects the scattering profile by decreasing the depth of the first minimum (without a change in position) as illustrated in figure 4.3B.

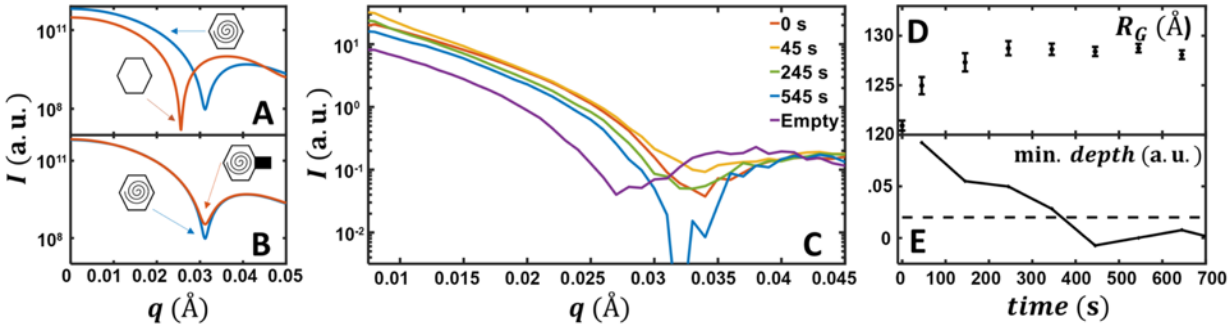

Figure 4.3 Time resolved SAXS data of ΦX174 genome ejection.
A&B. Simulated SAXS profiles of empty capsid, capsid with DNA, and full capsid with asymmetric addition. Panel A illustrates the impact, on the scattering profile, of increasing the particle's radius of gyration. Panel B illustrates the impact of losing rotational symmetry, by addition of a model projection at only one vertex. For this model scenario, the mass of the addition is insufficient to significantly alter the overall radius of gyration, hence the peak position appears unchanged. C. Experimental Time Resolved SAXS data showing the loss of symmetry and fullness over time of the virion. D&E. Radius of gyration and depth of the first minimum as a function of time. The dotted line represents the noise level in the SAXS data.

Scattering profiles for the full phage (acquired before the addition of LPS, t=0 data point) and empty phages (acquired using degraded procapsids lacking DNA and the external scaffolding protein) are shown in figure 4.3C. For time resolved data, LPS was manually mixed with ΦX174

at 4˚C, which allows attachment but not ejection. The reactions were loaded into a temperature controlled sample cell held at 33˚C. SAXS profiles collected at various times during the reaction show changes in the average size, mass, and symmetry of the phage particles. The first time resolved profile, acquired 45 s after LPS addition, displays a loss of depth in the first minimum and an increase in the intensity at a scattering angle of zero degrees, I(0). These changes indicate a loss of symmetry and increase in particle mass, respectively, which is consistent with LPS binding. Subsequent scattering profiles (acquired after 245 and 545 s) show a leftward shift and increased first-minimum depth, indicating an increasing Rg and recovery of symmetry. I(0) decreases below that of the initial t=0 state, reflecting a net loss of particle mass over the reaction's course. Time courses for the Rg and depth of the first minimum are shown in figure 4.3 D and E, respectively. In summary, the SAXS data show an initial increase in particle mass and loss of symmetry after LPS addition. Viral particles binding LPS at a single vertex would produce this result. At later time points, particle symmetry increases while Rg and particle mass decrease, which is consistent with genome ejection and phage release from the LPS bilayer.

4.3.3. Cryo-EM single particle reconstructions of ΦX174-LPS complexes

Cryo-EM single particle analysis was performed to obtain greater structural detail regarding the changes occurring upon LPS interaction. Purified ΦX174 particles were mixed with LPS at 33°C for 1min before being frozen in vitreous ice. Most of the particles were found attached to LPS bilayers by a single vertex (figure 4.4A, black arrows) but some seem to be attached by more than one vertex (figure 4.4A, white arrow). The attached particles were "boxed" and subjected to non-reference two-dimensional (2D) classification using the software package RELION (22).

Among the 10,000 selected particles, approximately 7,000 were empty while 2,400 still contained DNA. Compared to the results of the infectivity assay at an equivalent time point, where all particles were sampled, the percentage of emptied particles was much higher in LPS-complexed ΦX174. This suggested that binding to LPS bilayers is crucial for DNA ejection. Full and emptied particles were separately subjected to 3D analysis using RELION (22). A 5-fold rotational symmetry was imposed during the processing. A masked 3D classification process in RELION was used to select groups of moderately homogeneous particles (Materials and Methods). Approximately 2,400 full particles went into the final reconstruction of full ΦX174, attaining a

final resolution of 10.2 Å (figure 4.4B, left), whereas 2,500 particles were used to generate a 9.8 Å resolution map of emptied ΦX174 (figure 4.4B, right).

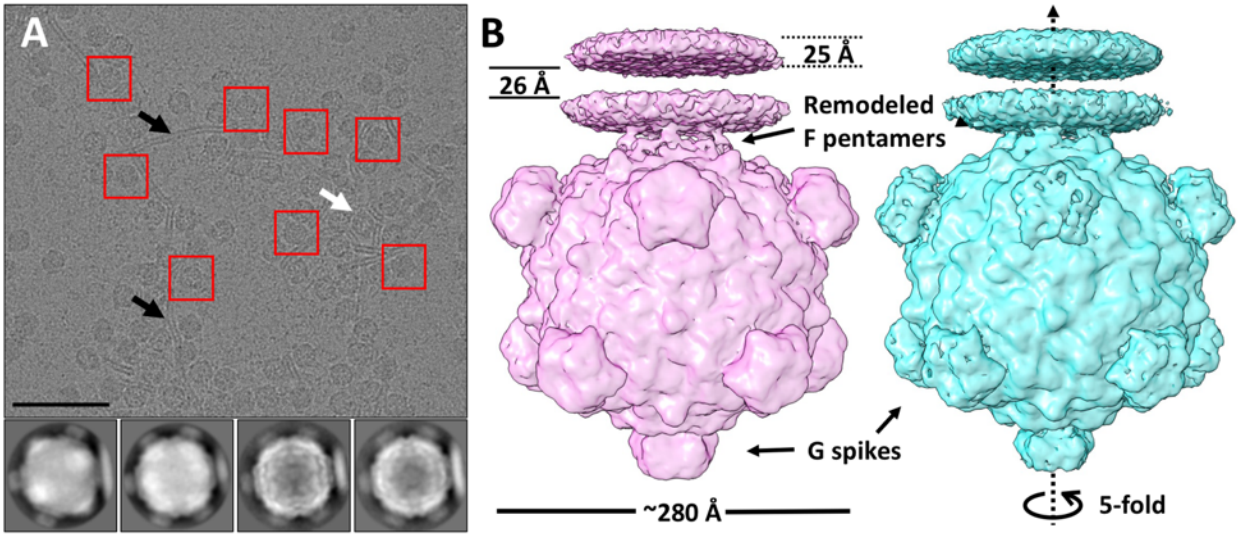

Figure 4.4 Cryo-EM single particle processing of ΦX174 complexed with LPS.
A. (Top) A raw image of ΦX174 incubated with LPS (Scale bar, 100 nm) with examples of selected particles (red boxes) bound to LPS bilayers (black arrows). Some particles are attached to bilayers at more than one vertex (white arrows). (Bottom) Examples of 2D classes generated by RELION. Full and emptied particles were well separated. B. Reconstructions of ΦX174 complexed with a LPS disk before (pink, full) and after (cyan, emptied) the ssDNA genome is ejected. These two states of ΦX174 have overall similar architectures, with differences in F proteins at the ΦX174 –LPS interface and the G protein spikes.

In both the 2D class averages and the 3D reconstructions, LPS is seen as a bilayered disk. The thickness of each layer is about 25 Å, separated by about 26 Å. The higher electron density in each of the bilayers is the consequence of the greater electron scattering power of the heavier phosphate atoms (23). In both maps, there was no density at the interacting 5-fold vertex that would account for the G spike, suggesting that the spike has dissociated from the rest of the capsid upon binding to LPS. The absence of a G spike at the interface leaves the F protein pentamer anchoring the virion on the LPS disk. The LPS disk is connected to ΦX174 through five symmetry-related positions at one of the 5-fold vertices of the virus.

A pentameric F complex, from the crystallographic structure of the virus (6), was then fitted into the cryo-EM maps of the emptied and the full particles at the interacting vertex (figure 4.5A) using EMfit (24). Both maps were low-pass filtered to 10 Å resolution. The other vertex on the 5-

fold axis, opposite the bound LPS disk, was also fitted with the F pentamer to measure the influence of LPS binding on the conformation of the F proteins. The average density at the atomic positions of the fitted pentamer, sumF (24), was calculated based on only Cα atoms (Table 4.1). The crystal structure fitted better into the EM map of the emptied ΦX174 particles at the vertices that had no contact with LPS. Thus, the binding of LPS had caused significant conformational changes in the F proteins at the interacting vertex. Given that fitting into full ΦX174 has a smaller sumF value and the local resolution at the connecting region is lower compared to emptied ΦX174, it is reasonable to conclude that this conformational change is more significant in full ΦX174 particles with bound LPS compared to emptied particles.

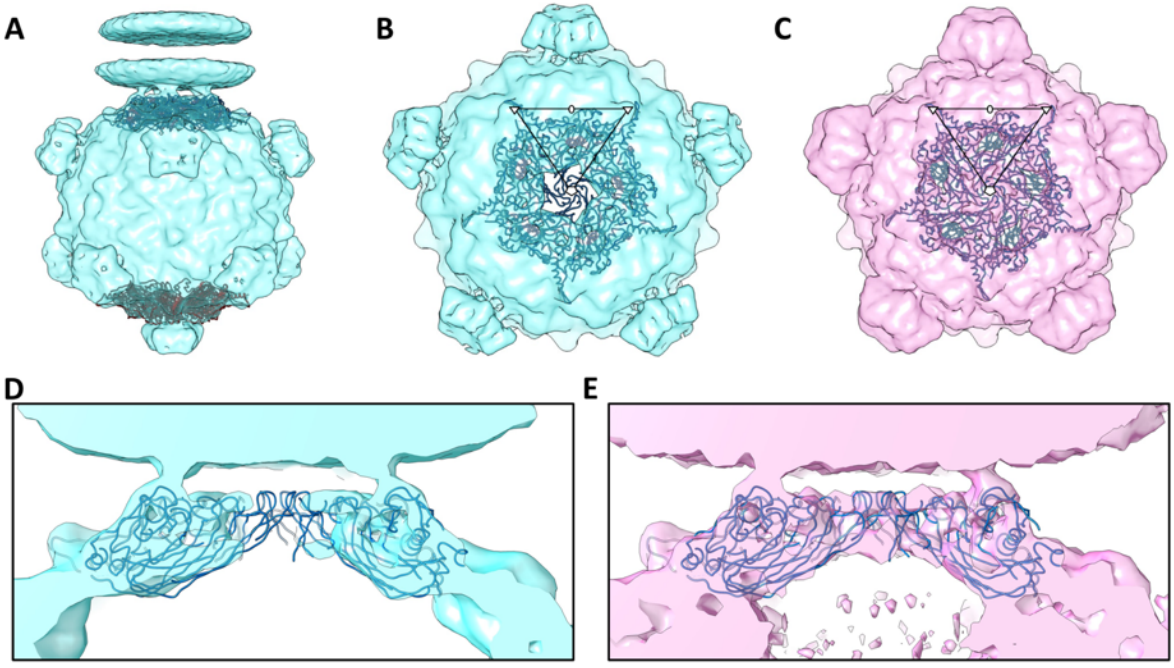

Figure 4.5 Fitting F pentamers into full (pink) and emptied (cyan) ΦX174 when in contact with LPS shows conformational changes at the 5-fold vertices.
A. Side view of fitting F pentamers into opposing vertices on the 5-fold aixs of the reconstruction of emptied ΦX174. B&C. Top views of the 5-fold vertices in contact with LPS after the density of LPS has been removed from the maps. No density corresponds to the HI loops of the F proteins at the vertex in contact with LPS in the emptied virus. An icosahedral asymmetric unit is outlined by a triangle. D&E. Central slices of the maps parallel to the 5-fold axes. Crystal structure of ΦX174 (PDB:2BPA) has been fitted (blue ribbon) to show the change in the 5-fold channel after the genome has been ejected.

Table 4.1 SumF values determined with the EMfit program when fitting the F pentamer into different sites of the maps for the full and emptied particles.

Site being fitted with F pentamer	Emptied	Full
Vertex in contact with LPS	48	41
Vertex on the opposite side	65	59

A "roadmap" (25) that shows the residues of the F proteins present on the viral surface is given here to indicate the position of the density that represents the virus-LPS contact region (figure 4.6). The bulge formed by the EF loop of the F proteins was below the density linking the capsid and the LPS disk. There are positively charged residues Lys118, Arg157 and Lys342 in the contact region, which could form electrostatic interaction with the phosphate group of LPS, while Asn117 and Gln157 may form hydrogen bonds with the hydroxyl group of LPS. This is similar to a crystal structure of a toll-like receptor molecule complex with LPS, in which multiple Lys, Arg , Gln and hydrophobic residues of the receptor mediate the interaction (23).

A five-residue-long hydrophobic region was identified by aligning F proteins from *Microviridae* members representing the three major evolutionary clades (figure 4.7). This region is located on the "bulge" formed by the EF loop and buried under the surface of the virus. When the G spike dissociates from the capsid, the interaction that has confined the conformation of the EF loop no longer exists. Thus, this hydrophobic region has a potential to reorient and interact with the lipid component of LPS.

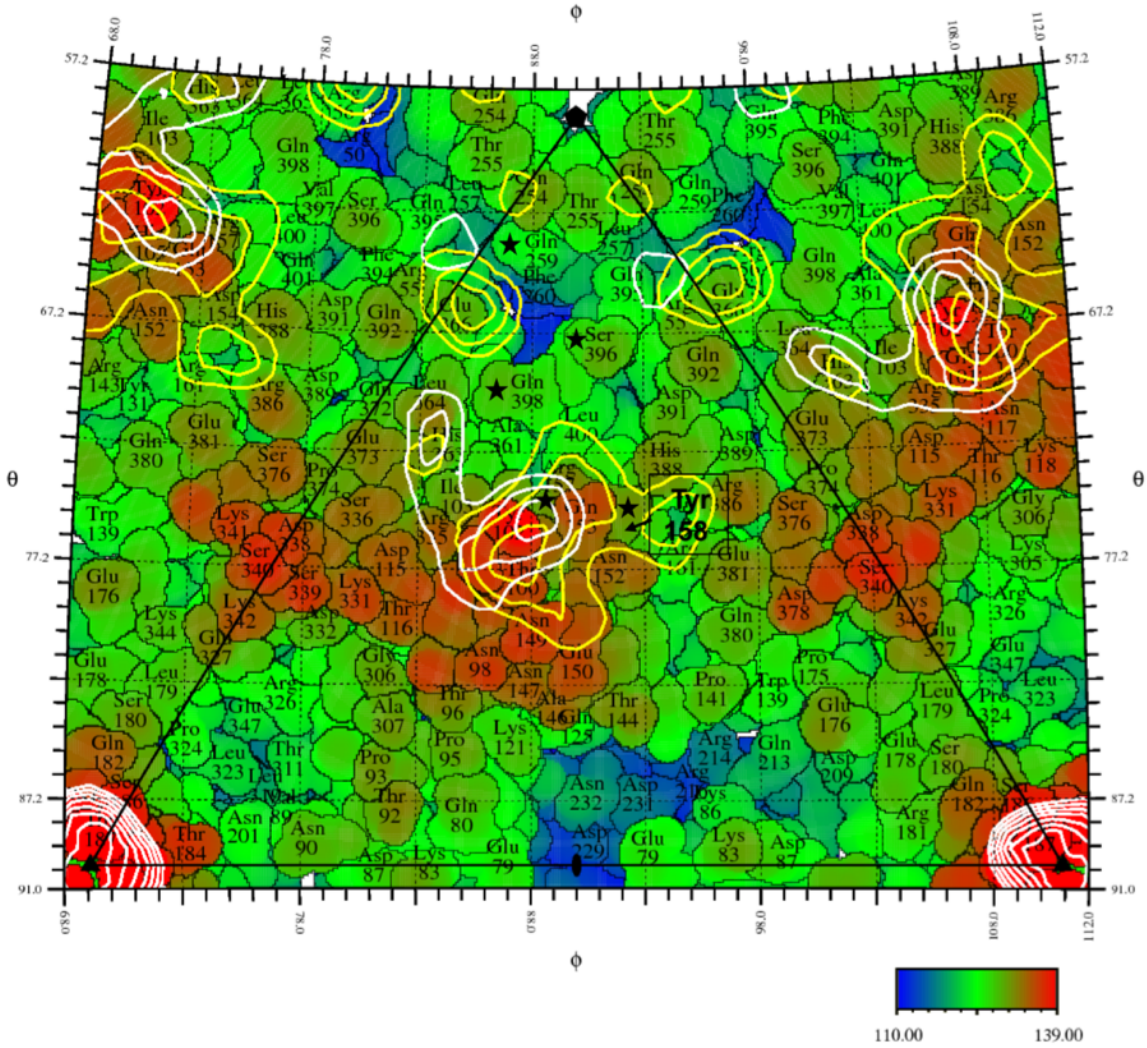

Figure 4.6 Residues on the F proteins that may be involved in the interaction with LPS. A "bulge" (mainly colored in red) formed by the long loop insertion between βE and βF of the capsid protein F is located underneath the density that connects the capsid to the LPS disk. A roadmap representing the surface of F capsids within one asymmetric unit (black triangle), colored by distance (Å) to the center is shown. The numbers on the side represent the angles of the corresponded spot on the surface relative to the 2-fold axis (black ellipse at middle bottom) which is normal to the plane of this page. The contour lines show the density of the reconstructions that is right above the capsid and connect it to the LPS disk (white: emptied, yellow: full). The residues that are involved in the interaction with G proteins in native ΦX174 are labeled with black five-pointed stars. ΦX174's attachment to LPS seen in this study is closely related to the dissociation of the G spike at the same vertex.

Figure 4.7 Sequence alignments of EF loops in the F proteins of homologues phages from *Microviridae*.
A conserved hydrophobic region (magenta) was identified. There is also an overlap between conserved residues (green) and residues that are sticking out on the surface of the "bulge". These residues may play important roles in the LPS binding.

A major difference between the reconstructions of full and emptied ΦX174, both complexed with LPS, lies at the channel along the 5-fold axis at the interacting vertex. The other long F protein loop HI is located in this region. In the apo state, two gates separate the genome from the outside environment at the 5-fold vertices. The first gate is the major spike protein G and residing 35 Å underneath the first gate, the second gate is formed by residues Gly252 to Gln260 of the major capsid protein F. The diameter of the second gate is only about 4 Å (6). In emptied ΦX174, loop 252-260 is disordered and the "gate" is more open with a bigger diameter of 30 Å (figure 4.8). However, in the reconstruction of full ΦX174 complexed with LPS, which likely represents an intermediate in the genome delivery pathway, the "second gate" remains closed, keeping the genome inside the capsid.

The extra electron density inside full ΦX174 particles is not evenly distributed. A radial, cylindrical cavity is found close to the vertex that interacts with the LPS, with a length of approximately 110 Å and a diameter of 80 Å (figure 4.8). The dissociation of the G spike protein pentamer and alterations in the structure of the F protein pentamer may result in the release of the internal pressure. Consequently, the DNA in this region may be less condensed, which might be the reason for the observed cavity in the reconstruction. Alternatively, a complex between the H protein and the ssDNA genome may form in this region. Since protein has a smaller density than DNA, the electron density in this volume might be somewhat lower.

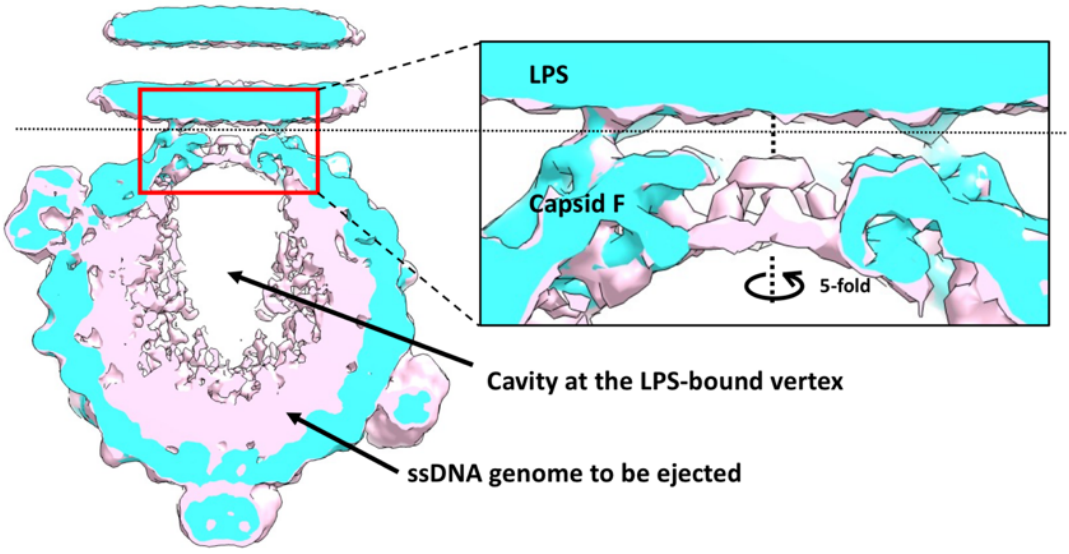

Figure 4.8 A close look at the differences between full (pink) and emptied (cyan) ΦX174 around the 5-fold axis.
Left: overlaid central slices of the reconstructions of full (pink) and emptied (cyan) ΦX174 along the 5-fold axis; right: a zoom-in view of the red window. Full ΦX174 has an unevenly distributed internal density which accounts for the genome and the H proteins. The cavity is very likely a result of averaging less homogeneous structures.

The full ΦX174 complexed with LPS described here is likely an intermediate during the infection. It has undergone significant structural changes at the interacting vertex but still contains the genome. The portion of ΦX174 particles in this state was relatively low, suggesting that this might be a metastable conformation of ΦX174. The existence of this intermediate implies that ΦX174 pauses transitorily after losing the G spike so that the ssDNA genome can be rearranged for the subsequent ssDNA ejection.

4.3.4 ΦX174's interaction with LPS-containing liposomes

In an attempt to capture any intermediate state of ΦX174 virion during DNA ejection, LPS molecules were incorporate into multilamellar vesicles (MLVs) consisting of PC, PE and PG at 1:1:0.25. These LPS-containing MLVs were than mixed with native ΦX174 for cryo-electron tomographic study. Tilt series from -60° to 60° were collected in a dose symmetric way, the alignment of each tilted image and the final reconstruction of tomographic volumes were carries out in IMOD (26). ΦX174 particles were observed to adsorb on the membrane bilayers of liposomes in multiple structural states (figure 4.9).

One of the issues accoutered was the reconstituted liposomes were very sensitive to radiation damage. Initially the data was recorded at an angular step of 2°, but aerification due to radiation damage was seen at high tilts. The step size between each tilt was then increased to 3°.

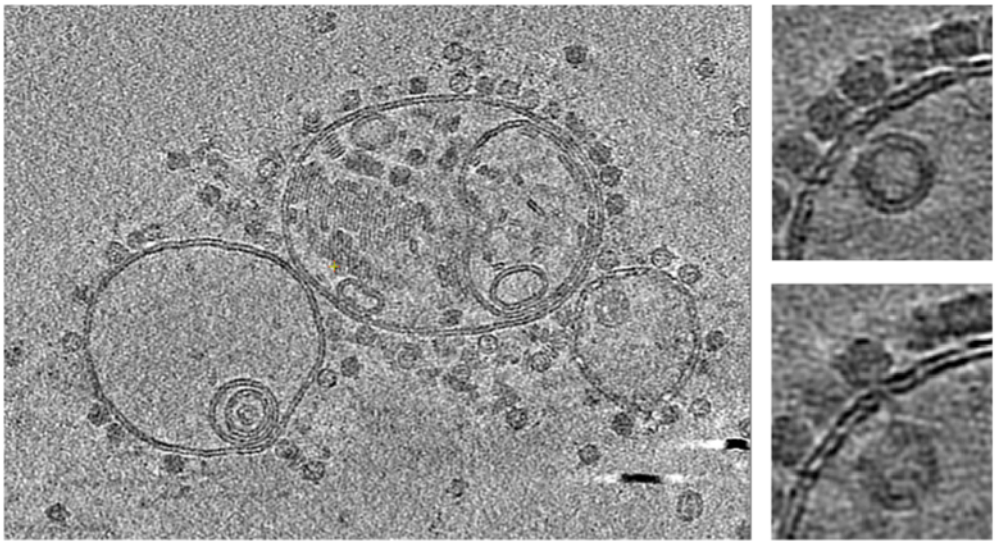

Figure 4.9 Cryo-ET of ΦX174 interacting with LPS-containing liposomes.
Sections of the reconstructed electron tomograms with the z-axis (direction of incoming electron beam) perpendicular to the plane of the image.

Due to the missing wedge, low SNR and imperfect stage tracking, the electron tomograms were extremely noisy. To further characterize different structural states, 3D classifications and sub-tomogram averaging were performed in DYNAMO (27). Two major classes, full and empty particles, were identified. The sub-tomogram averages showed that ΦX174 in both states are binding to liposomes through one of the twelve 5-fold vertices, and the G spike at the interface is missing. Besides full and empty particles that resemble those observed in single particle reconstructions, a minor class consisting of only 84 sub-tomograms was identified, with a continuous density extruding from the ΦX174 capsid and penetrating the liposomal membranes (figure 4.10C&F).

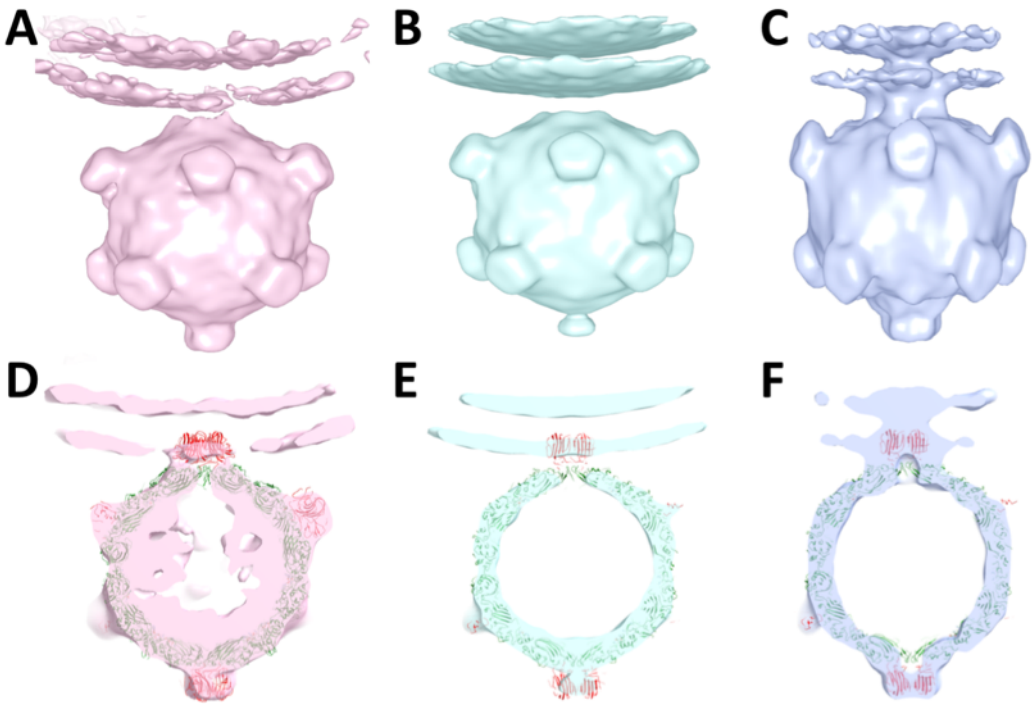

Figure 4.10 Sub-tomogram averages of liposome-bound ΦX174 particles.
A-C. Side views of sub-tomogram averages of full (pink), empty (cyan), and a minor class (purple) ΦX174 particles associated with liposomal membranes. D-E. Side views of central slices of the same averages. The structure of native ΦX174 (PDB-2BPA) is fitted to show the relative position of the F capsid proteins (green) and the G spike proteins (red).

4.4 Discussions

Our findings, combined with results from previous studies (9, 12–14, 17, 19, 28, 29), suggest a pathway for the ΦX174 infection (Fig. 5). The initial contact with the host cell is made by one of the twelve G protein spikes. This spike then dissociates from the capsid leaving a structurally altered F protein pentamer to maintain viral attachment with LPS. When the G proteins dissociate from the capsid, the F proteins on this vertex gain more conformational freedom, particularly the EF and HI loops. The change in the "bulge" formed by the EF loop enables the formation of a stable contact between the capsid and the host cell's membrane. Meanwhile, the change in the HI loop opens the gate at the special vertex, and prepares the virus for DNA translocation. At the same time, changes in the conformations of the H proteins and the ssDNA presumably occur inside the capsid to facilitate genome ejection through the vertex that is in contact with LPS.

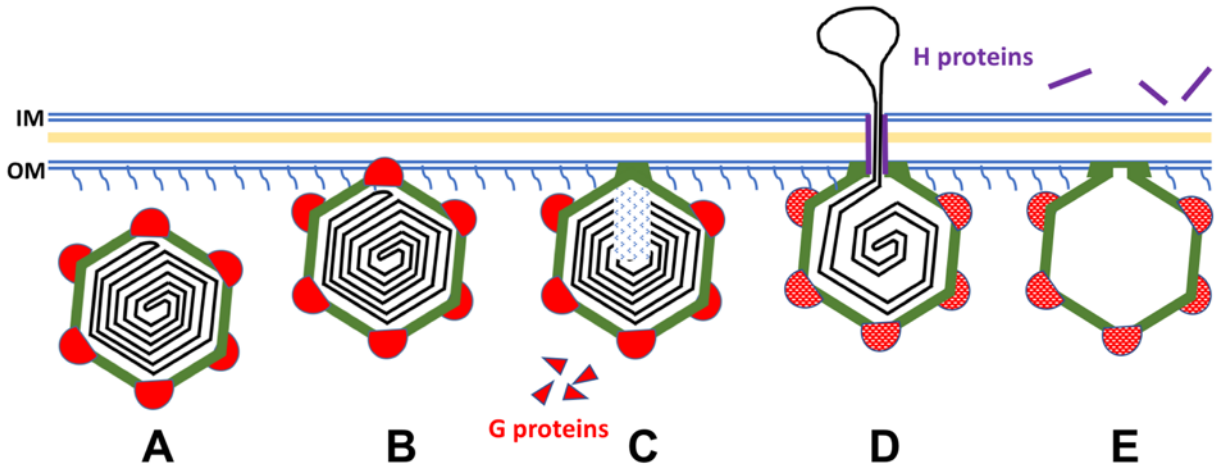

Figure 4.11 A proposed model for ΦX174 DNA ejection.
A&B. Native ΦX174 recognizes its host carrying specific LPS on the surface hypothetically through one of the G spikes (red). C. The interaction destabilizes the G spike and causes its dissociation, leaving the loops on the surface of F proteins (green) maintain the interaction with the cellular wall of host. D. The dissociation of G proteins causes subsequent conformational changes in the F proteins (green) and H proteins (purple). The F and H proteins cooperate in translocating the genome (black) across the cellular wall of the host. E. The exit used by DNA remains open after ejection. H proteins are ejected along with the genome. F proteins still interact with the outer membrane.

In a crystallographic study, McKenna et al. found that the G proteins make minimal contact with the capsid (6). Thus, it is likely that LPS can induce the dissociation of the G spike by disrupting the interaction between the G spikes and the F proteins. A further question is how ΦX174 penetrates the peptidoglycan and the inner membrane of the host after the dissociation of the G proteins. A partial answer is that the H proteins may form a tube-like conduit that serves to translocate the ssDNA genome (9, 30). It may be relevant that ΦX174 particles congregate at membrane adhesions sites, or Bayer's patches, where the inner and outer membranes appear to merge (31). Although the H-tube is not seen in this structure, we might expect that the H tube would be right behind the G spike in contact with the LPS. But as long as the H tube is still mixed up with the genome, we are unlikely to recognize it in any reconstruction as was suggested (9, extended data figure 2). Furthermore, it is unknown whether the H proteins assemble into tubes at the time of viral assembly or are assembled later during ejection. After the genome is injected, the H-tube disappeared in the cytoplasmic space (9), and it has been shown that the H proteins are transported inside the cell along with the genome (32). These results are confirmed by the

reconstruction of emptied ΦX174 present in this study. No tubular structure is observed associated with the emptied capsid.

4.5 Methods and Materials

4.5.1 Amplification and purification of ΦX174

6-8 liters of E. coli C122 were grown in TK broth (1.0 % tryptone, 0.5% KCl) to 1.0 X 108 cells/ml at 37°C. Prior to infection at multiplicity of infection of 10-5, MgCl2 and CaCl2 were added to respective concentrations of 10 mM and 5 mM. Infections were incubated for approximately 5 hr until the clear majority of cells had lysed. Cultures were refrigerated overnight to allow phage to attach to cell debris, which was concentrated by centrifugation. The resulting pellet was resuspended in 32 ml of 50 mM Na2B4O7 / 3.0 mM EDTA, and shaken at 4°C overnight to elute attached particles. Debris was removed by centrifugation, and the supernatant was layered atop CsCl gradients made in 50 mM Na2B4O7 / 3.0 mM EDTA and spun as previously described (33). Virion bands were pulled and dialyzed against 100 mM NaCl, 5.0 mM EDTA, 6.4 mM Na2HPO4, 3.3 mM KH2PO4 (pH 7.0), and concentrated down to 0.8 ml. 0.2 ml were loaded atop a 5-30% sucrose (w/v) gradient made in 100 mM NaCl, 5.0 mM EDTA, 6.4 mM Na2HPO4, 3.3 mM KH2PO4 (pH 7.0) and spun at 192,000 × g for 1 hour. Gradients were divided into approximately forty 125 µl fractions and analyzed by UV spectroscopy (OD280).

4.5.2 ΦX174 DNA ejection assay.

Concentrated ΦX174 in 0.06 M NH4Cl / 0.09 M NaCl / 0.1 M KCl / 0.1 M Tris-HCl (pH 7.4) / 1.0 mM $MgSO_4$ / 1.0 mM $CaCl_2$ was mixed with 5mg/ml stock LPS dissolved in Tris-HCl (pH8.0) buffer to achieve desired concentrations. $CaCl_2$ was added to a concentration of 5.0 mM. ΦX174 and Tris-HCl buffer without LPS was used as the negative control. The final concentration of ΦX174 was about 108-109 pfu/ml. The mixtures including the negative control were then kept at 33°C for 20 min. After 1, 3, 5, 8, 12 or 20 minutes, the mixtures that contained LPS were diluted into 3.0 mM EDTA to terminate the DNA ejection of the particles. Diluted samples were tittered as previously described (33). The dilution procedure was optimized for each batch of sample so that the number of plaque on a plate was between 10 and 500. The plaque counts of three replicates were averaged for each condition to establish the relative infectivity.

4.5.3 Small-angle X-ray Scattering

SAXS data were collected using 11.18 keV X-rays at the Cornell High Energy Synchrotron Source G1 station. Scattering profiles (0.006–0.26 Å-1) were collected with a sample-to-detector distance of 2.05 m (measured using an Ag-Behenate standard) onto a PILATUS 200K (Dectris) detector. SAXS profiles were normalized using the transmitted beam through a semi-transparent molybdenum beamstop. All samples were oscillated during data collection to reduce the effects of radiation damage inside a quartz capillary (2 mm diameter, 10 μm walls, Hampton Research) mounted in a custom-built temperature-controlled holder. Multiple 10 s exposures were acquired and averaged to improve the signal-to-noise ratio. All SAXS images were analyzed using MATLAB (MathWorks). Scattering profiles were azimuthally averaged about the beam center. Buffer scattering was measured before and after each sample and averaged before being subtracted from the sample scattering. For time-resolved experiments, profiles are shown at the indicated time after the addition of LPS to the sample. Sample conditions examined: 30 μL samples at 0.05-0.1 mg/mL ΦX174, 0.15 mg/mL LPS, 0.06 M NH4Cl2, 0.09 M NaCl, 0.1 M KCl, 1 mM MgS04, 1 mM CaCl2, 0.1 M Tris-HCl pH 7.4 at 33°C. The same trends were observed but with faster rates at elevated temperatures (37°C) and slower rates with lower LPS concentrations (0.01 and 0.05 mg/mL).

4.5.4 Cryo-Electron Microscopy

Cryoplunge 3 system was used to prepare the cryo samples. 5 μl of purified ΦX174 at 10^{11} PFU/ml was incubated with 1 μl of 1 mg/ml LPS from Salmonella enterica TV119 (Sigma-Aldrich) at 33°C for 1 minute. Aliquots of 3 μl of the mixture were loaded onto lacey carbon grids (400 mesh; Ted Pella Inc.). The grids were blotted by filter paper for 5 seconds and plunged into liquid ethane for cryo-EM inspection. Forty consecutive frames of ΦX174 reacting with LPS embedded in vitreous ice were recorded using a Titan Krios TEM (FEI) operated at 300 kV and equipped with a Gatan K2 Summit direct electron detector (3,838×3,710). All of the frames were collected automatically in counting mode using Leginon (34) at a magnification of 18,000 and with a defocus range of 1 to 3 μm, which generated a pixel size of 1.62 Å per pixel.

4.5.5 Image Processing

A total of 2908 frame stacks were collected. Beam induced motion was corrected by realigning the frames within one stack using the program MotionCorr (35). The version of MotionCorr that was used here had been modified by Prof. Wen Jiang to take three consecutive frames instead of one to calculate the correlation coefficient. CTF values were estimated for each motion-corrected image using CTFFIND3 (36). From the 2908 images, 11,845 particles were manually selected using *e2boxer.py* in the EMAN2 package (37). The particle images were 2x binned and subjected to non-reference 2D classification using RELION (22). Subsequently the particles in the classes that failed to generate a good average projection of ΦX174 were discarded, leaving 7002 emptied particles and 2384 full particles. Two rounds of masked 3D classifications were then used (Fig. S3) to select particles more homogeneous in the region where the LPS disk is connected to ΦX174. This mask was generated using the program Chimera (38) and *relion_mask_create* from RELION to identify the region of interest. The 3D classification and subsequent refinement were done in RELION. The initial model was generated using *jspr* (39) from a smaller dataset containing 800 particles. The initial model used for that reconstruction was a 50 Å low-passed-filtered map of native icosahedral ΦX174.

For the reconstruction of emptied ΦX174, after 2 rounds of 3D classification a subset of 4226 particles was selected and a map at 8.9 Å resolution was achieved. If all 7025 particles from 2D classification were used, the final resolution would be 8.2 Å, but with less continuous density in the stem region. This suggests that the stem region is more flexible and heterogeneous. However, 3D classification did not help improve the map of full ΦX174, probably due to the smaller number of particles to start with. All 2384 full particles selected based on 2D classification were used for the final reconstruction, which generated a map with a resolution of 10.2 Å judged by the "Gold Standard" criterion. The maps of the full and emptied particles have been deposited with the EMDB accession numbers EMD-8862 and EMD-7033, respectively.

4.5.6. Fitting of the pentameric F protein crystal structure into the EM maps.

One F protein pentamer was selected from the crystal structure of the whole virus (PDB:2bpa). The cryo-EM maps of full and emptied ΦX174 were both low-passed to 10 Å resolution. EMfit (24) was then used to fit the pentamer into different vertices of these low-passes maps. SumF

values (a measurement of quality of the fit) were calculated based on only the Cα atoms. The coordinates of the fitted pentamer have been deposited to Protein Data Bank.

4.5.7 The preparation of LPS- containing liposomes.

The protocol for incorporating LPS into liposomes was modified based on a previously described "dry method" (40). Mixture of 5 mg DOPC, 5 mg DOPE and 1.25 mg DOPG (Avanti Polar Lipids) was dried and exposed to high vacuum for 4 hours to remove any residual chloroform. Under nitrogen stream, 0.5ml lipopolysaccharides (Sigma-Aldrich, L6016) from *Salmonella enterica* at 5 mg/ml (20 mM Tris-HCl, pH 8.0) was used to resuspend the phospholipids. Three cycles of freezing and thawing of the solution in liquid nitrogen was performed. The solution was then placed in a speed vacuum overnight to remove water. Dried mixture was resuspended again and vertexed at 50°C. Then the solution was sonicated in water bath for 15 minutes.

4.5.8 Cryo-electron tomography

Purified ΦX174 virions (10^{12} PFU) were mixed with same volume of LPS-containing MLVs prior to freezing, with 10 nm gold fiducials added. Grids were vitrified immediately (within 2 minutes) as described in section 4.5.4 using C-flat holey carbon grids (Electron Microscopy Sciences, CF-2/2-4C).

Tilt series from -60° to +60° were collected using a Titan Krios microscope (FEI) operated at 300 kV onto a K2 summit detector (Gatan) with an angular step size of 3° and a defocus of -3 μm. Automated data collection was realized by SerialEM (41). The nominal magnification was 11000 and the pixel size was 1.3 Å. The dose rate used was 8 e-/pixel/s, and 10 of 150 ms frames were collected for each tilt image. The total dose of an entire tomogram series was about 71 $e^-/Å^2$.

Motion correction was performed for the frames within one tilt by motioncor2 (35). The tilt series were aligned using the gold fiducial markers and a total of 18 tomographic volumes were reconstructed using IMOD (26).

4.5.9. Sub-tomogram extraction and averaging

Liposomes-bound ΦX174 particles were selected and sub-tomogram volumes were cropped using DYNAMO (27). The average of eight manually aligned particles was used as an initial model

and all sub-tomograms were aligned. The aligned volumes were classified using a PCA (principle component analysis) based algorithm that was available in DYNAMO. Angular refinements were performed within each of the classes and averages were calculated. The maps were then inverted and low-pass filtered to 20 Å for structural interpretation.

4.6 Chapter Acknowledgements

This chapter is reprinted from "Sun, Y., Roznowski, A. P., Tokuda, J. M., Klose, T., Mauney, A., Pollack, L., ... & Rossmann, M. G. (2017). Structural changes of tailless bacteriophage ΦX174 during penetration of bacterial cell walls. Proceedings of the National Academy of Sciences, 114(52), 13708-13713." Sections 4.3.4, 4.5.7, 4.5.8, 4.5.9 and 4.6, however, have not been published.

Helpful discussions with Yue Liu and Lei Sun were appreciated. We also appreciate help from Weifeng Shang and Srinivas Chakravarthy at Advanced Photon Source (APS) sector 18 as well as Richard Gillilan and Arthur Woll at Cornell High-Energy Synchrotron Source (CHESS). We thank Sheryl Kelly for help preparing this manuscript. Use of APS, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. CHESS is supported by the National Science Foundation (DMR01332208) and the National Institutes of Health/National Institute of General Medical Sciences. The research was supported by NSF grant MCB-1515260 to MGR, NSF grant MCB-1408217 to BAF, U.S. Department of Agriculture Hatch funds awarded to the University of Arizona and the BIO5 Institute (BAF). This work supported by an NSF Graduate Research Fellowship to A.M. under grant DGE-1650441.

4.7 References

1. Yap ML, et al. (2016) Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc Natl Acad Sci U S A* 113(10):2654–9.
2. González-García VA, et al. (2015) Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. *J Biol Chem* 290(16):10038–44.
3. Taylor NMI, et al. (2016) Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533(7603):346–52.
4. Hu B, Margolin W, Molineux IJ, Liu J (2013) The bacteriophage t7 virion undergoes extensive structural remodeling during infection. *Science* 339(6119):576–9.
5. Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1–24.
6. McKenna R, et al. (1992) Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications. *Nature* 355(6356):137–143.
7. Ilag LL, Incardona NL (1993) Structural Basis for Bacteriophage ΦX174 Assembly and Eclipse as Defined by Temperature-Sensitive Mutations. *Virology* 196(2):758–768.
8. Roux S, et al. (2012) Evolution and Diversity of the Microviridae Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. *PLoS One* 7(7):e40418.
9. Sun L, et al. (2014) Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. *Nature* 505(7483):432–5.
10. Andres D, et al. (2010) Carbohydrate binding of Salmonella phage P22 tailspike protein and its role during host cell infection. *Biochem Soc Trans* 38(5):1386–9.
11. Brown DT, MacKenzie JM, Bayer ME (1971) Mode of host cell penetration by bacteriophage phi X174. *J Virol* 7(6):836–46.
12. Inagaki M, et al. (2000) Characterization of the binding of spike H protein of bacteriophage phiX174 with receptor lipopolysaccharides. *J Biochem* 127(4):577–83.
13. KAWAURA T, et al. (2000) Recognition of Receptor Lipopolysaccharides by Spike G Protein of Bacteriophage ΦX174. *Biosci Biotechnol Biochem* 64(9):1993–1997.
14. Inagaki M, et al. (2003) Different contributions of the outer and inner R-core residues of lipopolysaccharide to the recognition by spike H and G proteins of bacteriophage phiX174. *FEMS Microbiol Lett* 226(2):221–7.

15. Suzuki R, et al. (1999) *Specific interaction of fused H protein of bacteriophage ΦX174 with receptor lipopolysaccharides* doi:10.1016/S0168-1702(98)00145-2.
16. Bull JJ, et al. (1997) Exceptional Convergent Evolution in a Virus. *Genetics* 147(4).
17. Young LN, Hockenberry AM, Fane BA (2014) Mutations in the N terminus of the ΦX174 DNA pilot protein H confer defects in both assembly and host cell attachment. *J Virol* 88(3):1787–94.
18. Incardona NL (1974) Mechanism of adsorption and eclipse of bacteriophage ΦX174. 3. Comparison of the activation parameters for the in vitro and in vivo eclipse reactions with mutant and wild-type virus. *J Virol* 14(3):469–78.
19. Incardona NL, Tuech JK, Murti G (1985) Irreversible binding of phage ΦX174 to cell-bound lipopolysaccharide receptors and release of virus-receptor complexes. *Biochemistry* 24(23):6439–46.
20. Doore SM, et al. (2015) The Kinetic and Thermodynamic Aftermath of Horizontal Gene Transfer Governs Evolutionary Recovery. *Mol Biol Evol* 32(10):2571–2584.
21. Svergun DI, Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep Prog Phys* 66(03):1735–1782.
22. Scheres SHW (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519–530.
23. Park BS, et al. (2009) The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* 458(7242):1191–1195.
24. Rossmann MG, Bernal R, Pletnev S V. (2001) Combining Electron Microscopic with X-Ray Crystallographic Structures. *J Struct Biol* 136(3):190–200.
25. Xiao C, Rossmann MG (2007) Interpretation of electron density with stereographic roadmap projections. *J Struct Biol* 158(2):182–187.
26. Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer Visualization of Three-Dimensional Image Data Using IMOD. *J Struct Biol* 116(1):71–76.
27. Castaño-Díez D, Kudryashev M, Arheit M, Stahlberg H (2012) Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J Struct Biol* 178(2):139–151.

28. Incardona NL, Selvidge L (1973) Mechanism of adsorption and eclipse of bacteriophage phi X174. II. Attachment and eclipse with isolated Escherichia coli cell wall lipopolysaccharide. *J Virol* 11(5):775–82.
29. Inagaki M, Wakashima H, Kato M, Kaitani K, Nishikawa S (2005) Crucial role of the lipid part of lipopolysaccharide for conformational change of minor spike H protein of bacteriophage /X174. doi:10.1016/j.femsle.2005.08.014.
30. Sun L, Rossmann MG, Fane BA (2014) High-resolution structure of a virally encoded DNA-translocating conduit and the mechanism of DNA penetration. *J Virol* 88(18):10276–9.
31. Bayer ME, Starkey TW (1972) The adsorption of bacteriophage ΦX174 and its interaction with Escherichia coli; a kinetic and morphological study. *Virology* 49(1):236–256.
32. Jazwinski SM, Marco R, Kornberg A (1975) The gene H spike protein of bacteriophages phiX174 and S13. II. Relation to synthesis of the parenteral replicative form. *Virology* 66(1):294–305.
33. Fane BA, Hayashi M (1991) Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage ΦX174. *Genetics* 128(4).
34. Suloway C, et al. (2005) Automated molecular microscopy: The new Leginon system. *J Struct Biol* 151(1):41–60.
35. Li X, et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584–590.
36. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
37. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
38. Pettersen EF, et al. (2004) UCSF Chimera: A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
39. Guo F, Jiang W (2014) Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses. *Methods in Molecular Biology (Clifton, N.J.)*, pp 401–443.
40. A procedure for the efficient incorporation of wild-type lipopolysaccharide into liposomes for use in immunological studies (1988) *J Immunol Methods* 114(1–2):197–205.

41. Mastronarde DN (2003) SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microsc Microanal* 9(S02):1182–1183.

VITA

I was born in Anshan, a city in the northeast of China known for steel production. I attended Anshan No. 1 Middle School from where I was admitted by Tsinghua University in 2009. When I applied for Tsinghua University, I wished to study automotive engineering but ended up in the School of Life Sciences. In Tsinghua, I had the honor to have Yigong Shi as my class advisor and I began to develop love for biological sciences. After all, a living creature is similar to a car, only far more sophisticated.

During the time in Tsinghua, I became interested in structural biology because of its accuracy and versatility. I then came to Purdue in 2013 and joined Prof. Michael G. Rossmann's lab to study structures of different viruses. I have been mainly studying virus-host interactions using cryo-electron microscopy.