STRUCTURAL STUDY OF TULANE VIRUS AND ITS HOST CELL FACTORS AND APPLICATIONS IN CRYO-EM

by

Chen 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 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Wen Jiang, Chair

Department of Biological Sciences

Dr. Richard Kuhn

Department of Biological Sciences

Dr. Andrew Mesecar

Department of Biochemistry

Dr. Markus A. Lill

Department of Medicinal Chemistry and Molecular Pharmacology

Approved by:

Dr. Janice Evans

In dedication to everyone who has helped and believed in me.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. Wen Jiang for his dedicated supervision of my research over the past few years and his great efforts in reviewing the thesis. I am grateful to all my committee members, Dr. Richard Kuhn, Dr. Andrew Mesecar and Dr. Markus A Lill, for the helpful discussions and their gracious support and guidance. Thanks are due also to all my current and previous lab mates, Dr. Rui Yan, Dr. Guimei Yu, Dr. Kunpeng Li, Dr. Manali Ghosh, Dr. Brenda Gonzalez, Frank S. Vago, Dr. Xueyong Xu, Dr. Bharath Sakshibeedu Rajegowda, Daoyi Li, Rejaul Hoq Nayem, Kadir Ozcan, Camila Tanimoto, Hannah Pletcher, Angela Irene Agnew and Samuel R Nemeth for their emotional and technical support in my research projects. I truly appreciate Steve M Wilson for his patience and tremendous help in IT management. I would like to thank Dr. Shishir Poudyal for his guidance in virology experiments. I have enjoyed the research collaborations with Dr. Xi Jiang, Dr. Nicolas Noinaj, Dr. Jiankang Zhu, Dr. Fang Huang, Dr. Arun Bhunia, Dr. Pengwei Li, Dr. Carol Post and Dr. Darci Trader. I am also in great debt to Dr. Thomas Klose for his help in data collection in the Purdue Cryo-EM facility.

My gratitude rightly extends to my family and my friends for their love and support through my graduate life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

a.a.:	amino acid
BS^3:	bis(sulfosuccinimidyl)suberate
BSA:	bovine serum albumin
cryo-EM:`	cryo-electron microscopy
DTT:	Dithiothreitol
ELISA:	enzyme-linked immunoassay
EMDB:	electron microscopy data bank
EMPIAR:	electron microscopy public image archive
FBS:	fetal bovine serum
GO:	graphene oxide
HA:	hemagglutination assay
HBGA:	histo-blood group antigen
kb:	kilobase pair
kDa:	kilodalton
MOI:	multiplicity of infection
nm:	nanometer
nt:	nucleotide
ORF:	open reading frame
PBS:	phosphate buffered saline
PCR:	polymerase chain reaction
PDB:	protein data bank
PFU/ml:	plaque forming units per milliliter
RdRp:	RNA dependent RNA polymerase
SDS-PAGE:	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TV:	Tulane virus
VLP:	virus-like particle
VPP:	volta phase plate

ABSTRACT

Currently, human norovirus is the leading cause of acute gastroenteritis and accounts for most cases of foodborne illnesses in the United States each year. Due to its tissue culture inefficiency, studies of human norovirus have been crippled for more than forty years. Tulane virus (TV) stands out as a suitable surrogate of human norovirus given its high amino acid identity with human norovirus and its well-established cell culture system. It was first isolated from rhesus macaques (*Macaca mulatta*) in 2008 and identified as a novel *Calicivirus* representing a new genus, *Recovirus* genus (Farkas et al., 2008). However, there are still unanswered questions about its infectious cycle and the essential factors for its infection.

In this study, we have obtained a TV variant (the 9-6-17 strain) that has lost the binding ability to the B-type histo-blood group antigen (HBGA), which was proposed to be the receptor of both TV and human norovirus. In the first chapter, we outline how the sequence analysis, structural biology studies, and mutagenesis studies of the 9-6-17 TV strain have shed light on the interaction with its host cell receptor. To investigate the key residues for HBGA binding, we established the full-length infectious clone of the 9-6-17 TV strain. We present a highly selective transformation of serine 367, located in the predicted HBGA binding site, into a lysine residue. Our results advance the understanding of genetic changes in TV required for adaptation to cell culture environments.

Cryo-EM is an awarding winning technique that has been the greatest scientific breakthrough in recent years. It was awarded the Nobel Prize in Chemistry in 2017. Despite the technological advances of the direct electron detector and image processing software, several major roadblocks remain for high-resolution structure determination with cryo-EM. In the later chapters, we explored the most efficient way of using VPP to enhance image contrast, how to tackle the air-water interface problem by encapsulating target protein, how to reach a higher resolution by refining high order parameters, and the helical indexing problem in real space. These technical advances would benefit the whole cryo-EM community by providing convenient tools or insights for future directions.

CHAPTER 1. INTRODUCTION

1.1 Overview of Tulane virus

Tulane virus (TV) is a non-enveloped, positive-strain RNA virus in the *Caliciviridae* family. It was first identified in 2008 and was classified into a new genus –*Recovirus* of the *Caliciviridae* family (Farkas et al., 2008). In 2012, another member of the *Recovirus* genus was found in the stool sample of human diarrhea cases in Bangladesh (Smits et al., 2012) and was named Recovirus Bangladesh/289/2007. Although it is genetically close to the TV, its genome organization is similar to that of norovirus in that their ORF1 and ORF2 overlap with each other, while the ORF1 and ORF2 of TV are separated by a few nucleotides. Human norovirus is the most common foodborne pathogen in the U.S (Qiu et al., 2021). In healthy individuals, it can only cause mild symptoms like vomiting, diarrhea, fever, or muscle aches. However, it can be fatal in immunocompromised individuals, the elderly, and children. Every year, human norovirus deals huge damage to the economy due to health care cost and food recall. Numerous attempts have been made to develop the vaccine for human norovirus. However, most of them failed in the clinical trials due to various reasons. Currently, even the most promising vaccine for human norovirus developed by the company HilleVax, Inc. is still in phase III. As the vaccine development of human norovirus has been lagging due to the lack of an effective culture system, many researchers have turned to find proper surrogates of human norovirus from the *Caliciviridae* family. TV stands out for its close genetic relationship with human norovirus and the fully established culture system. The TV genome is about 6.7 kb in length without the poly(A) tail, and it has three open reading frames (ORFs): ORF1 encodes all nonstructural proteins that are essential for replication and transcription, such as RdRp (RNA dependent RNA polymerase), ORF2 encodes major capsid protein VP1 and ORF3 encodes minor capsid protein VP2. Among all viral proteins, the capsid protein VP1 is of paramount importance because of its essential role in receptor recognition, host cell integration and genome release.

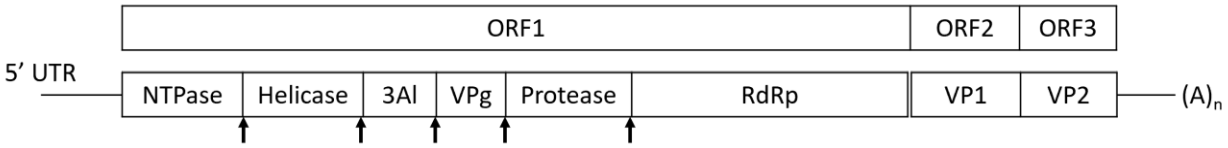


Figure 1.1. The schematic view of Tulane virus genome organization. Arrows point to the viral protease cleavage sites. 3Al is short for 3A-like protein.

1.2 *Caliciviridae* family

Currently, the *Caliciviridae* family consists of 11 genera and 13 species. The genera are listed in Fig. 1.2. They are single positive strain RNA viruses whose genome can be directly translated once released to the host cell. Their capsid diameter ranges from 27 nm to 50 nm. All members in the *Caliciviridae* family are nonenveloped viruses with icosahedral symmetry. The *Norovirus* genus and *Sapovirus* genus are the only two genera that have been reported to cause outbreaks of acute gastroenteritis in humans. The major host of other genera are animals other than humans. The first species identified in the *Recovirus* genus is the Tulane virus (Farkas et al., 2008) which is the object of interest of this study. Recovirus Bangladesh/289/2007 identified from human fecal sample in 2012 (Smits et al., 2012) is the only species in *Recovirus* genus to be able to infect humans.

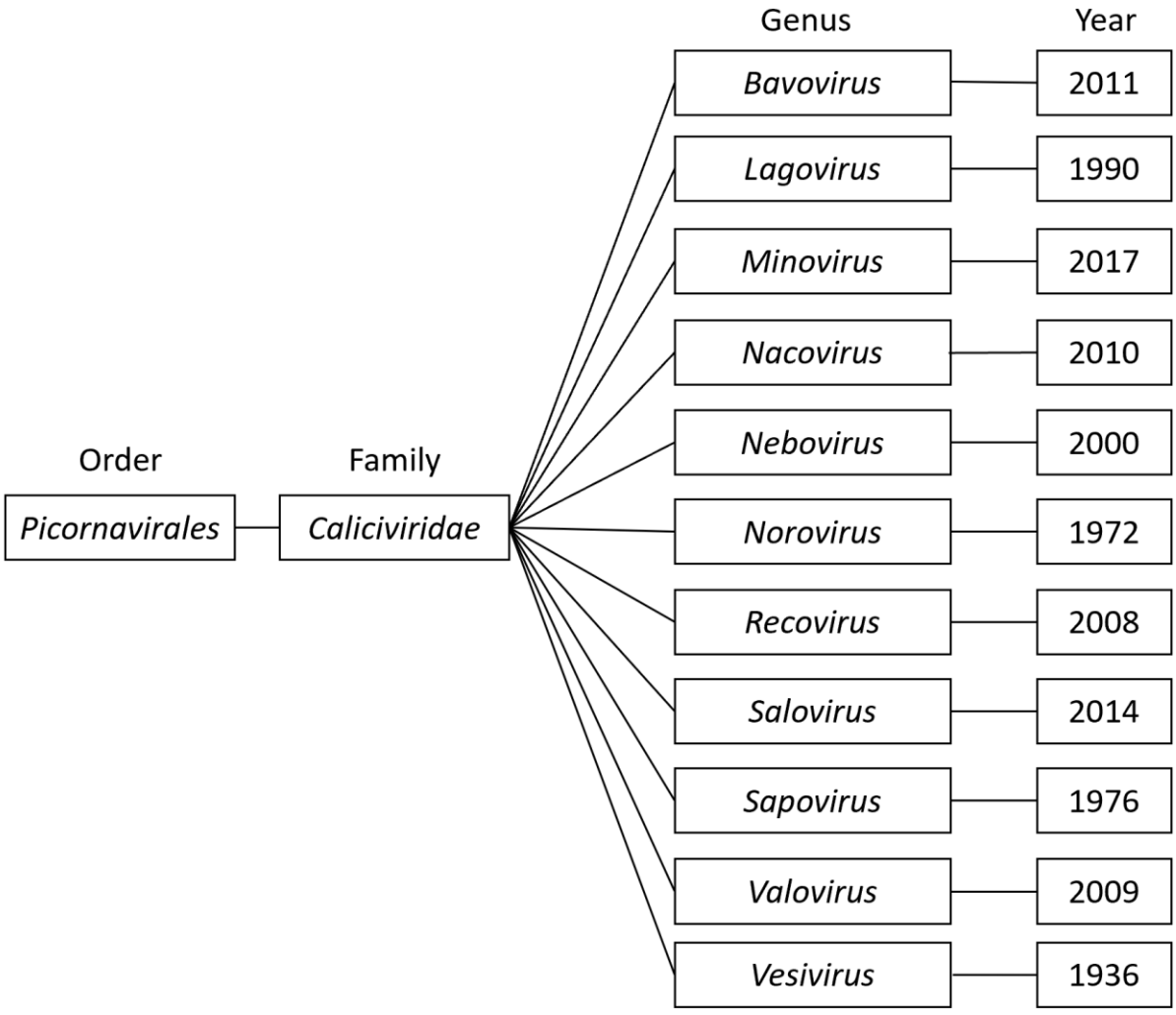


Figure 1.2. The current genus types of the *Caliciviridae* family according to the International Committee on Taxonomy of Viruses (Lefkowitz et al., 2017). The identification year is listed for each genus. References: *Bavovirus* (Wolf et al., 2011); *Lagovirus* (Parra et al., 1990); *Minovirus* (Mor et al., 2017); *Nacovirus* (Day et al., 2010); *Nebovirus* (Akbar et al., 2000); *Norovirus* (Kapikian et al., 1972); *Recovirus* (Farkas et al., 2008); *Salovirus* (Mikalsen et al., 2014); *Sapovirus* (Madeley et al., 1976); *Valovirus* (L'Homme et al., 2009); *Vesivirus* (Traum J., 1936)

1.3 Structure of Tulane virus

The first atomic structure of TV capsid has been published in 2016 by our lab with the antibody-based affinity grid approach (Yu et al., 2016). The structure reveals the characteristic T=3 icosahedral lattice with 90 dimers of capsid protein VP1. Each icosahedral asymmetric unit is composed of three subunits, denoted as subunits A, B and C, with A and B subunits forming a

dimer (A/B dimer) arranged around the five-fold axes of symmetry and the C/C dimer arranged around the two-fold axes. Each VP1 subunit can be further divided into two domains: shell domain (S domain) and protruding domain (P domain). S domain forms the inner spherical shell of the virus capsid. The P domain can be further divided into P1 and P2 sub-domains. The P2 sub-domain is responsible for receptor or antibody binding. VP1 has 534 amino acids with a molecular weight of 57.8kDa. With a high sequence identity over the full-length of the capsid protein, the capsid protein VP1 structure of TV, human norovirus and Norwalk virus adopt a similar fold in both S and P domains (Figure 1.3). Within the S domain, there is an eight-stranded jellyroll fold which is commonly found in virus structures. The L210 in TV at the hinge region connects the S domain and P domain.

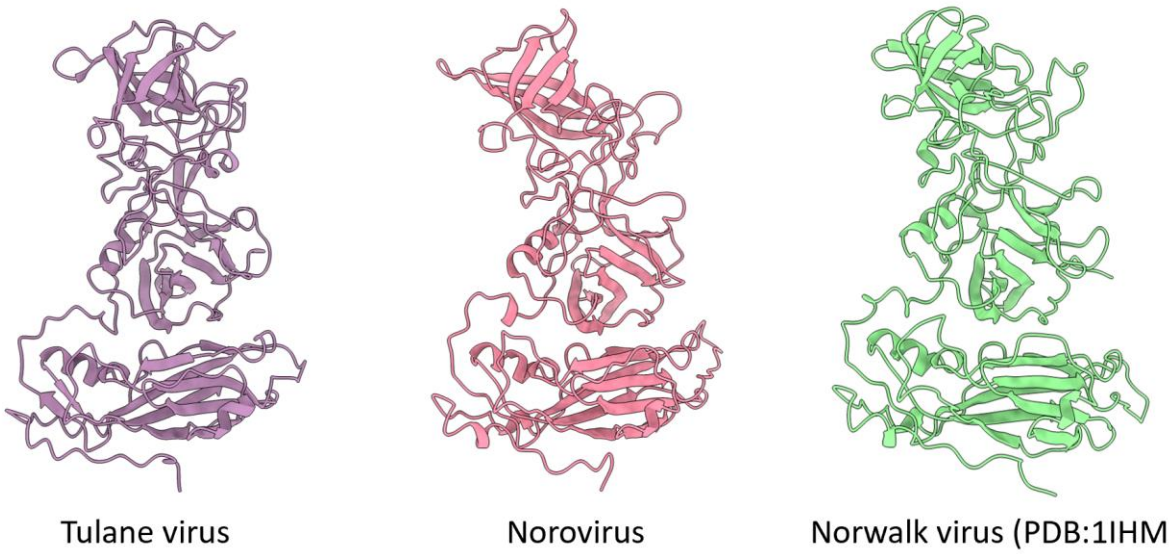


Figure 1.3. The structure comparison of the capsid protein VP1 of Tulane virus (Yu et al., 2016), human norovirus GII.4 (unpublished) and Norwalk Virus (PDB:1IHM).

The minor capsid protein VP2 of TV has 218 amino acids with a molecular weight of 22.84kDa. VP2 has been proven to be critical for the formation of viral capsid and genome release. A study of the Feline Calicivirus (FCV) has found that VP2 can form a funnel-shaped tube at the three-fold axis in the presence of the FCV host cell receptor junctional adhesion molecule A (JAM-A) to facilitate genome release (Conley et al., 2019). VP2 doesn't show up in the first TV density

map. It is still not clear what the number of VP2 in each viral capsid is and its role in the TV life cycle.

1.4 Histo-blood group antigens

Histo-blood group antigens (HBGAs) are oligosaccharide biosynthesized from disaccharide precursors. They are either on the glycoprotein or glycolipids on the cell surface of epithelial tissues and red blood cells or exist in free form in the extracellular matrix, and biological fluids. It consists of ABH antigens and Lewis antigens. ABH antigens are immobilized on the surface of red blood cells, while the Lewis antigens are absorbed onto the cell surface. Noroviruses can be classified into two groups based on the type of HBGAs that they interact with (Huang et al., 2005). One group of noroviruses can bind to ABH antigens but not Lewis antigens. The other group can only bind to the Lewis antigen and H antigens. HBGAs are reported to be the host cell receptor of many viruses, except TV and norovirus. The cell attachment protein VP8* of Rotavirus A (RVAs) recognizes fucosylated glycans (Barbe et al., 2018). HBGA is also the attachment factor of many bacteria, for example, *Pseudomonas aeruginosa*, *Helicobacter pylori*, and *enterotoxigenic Escherichia coli* (ETEC).

1.5 Viruses changing receptors

It is not uncommon for viruses to change their receptors by mutation. Influenza viruses are well-known for their great ability of cross-species transmission. Early in 1984, it has already been discovered that only two mutations in hemagglutinin would enable the chimeric avian influenza A virus with the hemagglutinin of human influenza A virus strain to infect avian (Naeve et al., 1984). Similarly, for the Measles virus which can cause serious disease in small children, a single amino acid change in the hemagglutinin protein can disable its binding ability to its receptor CD64 suggesting the existence of the secondary receptor in the primate B cell (Hsu et al., 1998). The Human Immunodeficiency Virus type 1 (HIV-1) was proved to have multiple coreceptors for infection and the combination of coreceptors is associated with the level of disease progression (Connor et al., 1997). In one study of the influenza C virus, a single point mutation T284I of the hemagglutinin-esterase-function (HEF) protein can allow the virus to grow in the cell line without the receptor that is required for its parent virus (Szepanski et al., 1992).

Similar findings also exist in the *Caliciviridae* family. Previous reports have shown that a single mutation in the capsid protein VP1 can lead to the antibody escape of human norovirus (Lindesmith et al., 2019). Antibody escape is also observed in murine norovirus (MNV) with L386F mutation (Lochridge and Hardy, 2007). Even the post-translational modification of Asp373 in the capsid protein VP1 can attenuate the binding ability of human norovirus with the histo-blood group antigens (HBGAs) (Mallagaray et al., 2019). Interestingly, with multi-sequence alignment, we found that the Asp373 position in human norovirus GII.4 is corresponding to the S367 position in TV which we identified to be the highly selective mutation for virus adaptation.

1.6 Single-particle Cryo-EM

Single-particle cryo-EM is a structure determination technique. To better preserve the sample from the radiation damage of the electron beam, biological samples are plunge-freeze in liquid ethane rapidly to form vitrified ice. Micrographs with 2D projections of the particles with different of orientations are generated in TEM. Based on the central projection theorem, the Fourier transform of the 2D projection is equal to the central slice of the Fourier transform of the 3D object. Thus, the 3D structure in real space can be generated by inverse Fourier transform of the 3D object in Fourier space. In recent years, the number of high-resolution structures determined with cryo-EM has increased dramatically. The Nobel Prize in Chemistry in 2017 was awarded to three pioneers in cryo-EM: Dr. Joachim Frank has built the first single particle image processing software—Spider (Frank et al., 1978); Dr. Jacques Dubochet is the one who has invented the plunge-freezing technique (Dubochet et al., 1988); Dr. Richard Henderson is recognized by his atomic structure of bacteriorhodopsin (Henderson et al., 1990) from 2D crystals early in 1990 and his continuous efforts in technology development in cryo-EM. The resolution revolution that occurred in cryo-EM is largely attributed to the direct electron detector with improved signal-to-noise ratio and the software development that streamlined the process and better utilized the information in a micrograph. However, several major roadblocks remain for high resolution structure determination with cryo-EM. The air-water interface is a major problem that proteins tend to be absorbed onto the air-water interface and get denatured by the surface tension. Some successes were achieved by adding detergent or using the affinity grid to immobilize the particles to the bottom of the grid. However, there is no universal and convenient solution to it so far. Other remaining problems include the flexibility and heterogeneity of the sample which would require

another revolution of the image processing software. The deep learning method is promising in tackling this problem. The radiation damage of the sample by the electron beam would be the hardest problem to solve since the elastic and inelastic scattering are always coupled together. Although the AlphaFold protein structure prediction (Jumper et al., 2021) has achieved near-experimental accuracy, it would still take a few years to validate the predicted structures and accumulate more data in PDB for it to replace the current structural determination process.

In chapter two, we introduced a more efficient way of using VPP to enhance image contrast by using a single position of VPP. It is more advantageous over the current method for it could significantly enhance the data collection efficiency and in the meantime, the results of it deepened our understanding of the effect of volta potential on the image. In chapter three, we designed a novel method to tackle the air-water interface problem by encapsulating the target protein or protein complex into the encapsulin shell. In chapter four, we reviewed the current image processing software and addressed the importance of refining high-order parameters in high-resolution refinement. In the final chapter, we provided a user-friendly and fully automated tool for helical indexing in a vision that anyone can obtain the helical parameters of their reasonable initial map easily without the need of learning about the complicated Fourier-Bessel function and knowing about the range of the helical parameters.

CHAPTER 2. STRUCTURAL ANALYSIS OF TULANE VIRUS AND ITS HOST CELL FACTORS

2.1 Introduction

Human noroviruses (HuNoVs) have been the leading cause of several severe gastroenteritis for many years and are still causing many outbreaks all over the world, often with mutations that alter susceptible populations. Considering the inaccessibility of a cell culture system for human norovirus, Tulane virus (TV) (Farkas et al., 2008) stands out as a valuable surrogate for HuNoVs. TV is the prototype of the *Recovirus* genus in the *Calicivirus* family. This single-stranded positive RNA virus was first detected in the stool sample of Rhesus macaques in 2008 and was found to be cultivable in several monkey kidney cell lines. Compared to murine norovirus (MNV), TV is genetically and structurally closer to the Norovirus genus. More importantly, it has also been reported to be able to interact with the histo-blood group antigens (HBGAs) which have been known as the cellular receptor of HuNoVs. HBGAs are carbohydrates found on the cell surfaces in most epithelial tissues and in secretions (Ravn and Dabelsteen, 2000). HBGAs are also utilized as a mediator of infection by many other human pathogens (Heggelund et al., 2017), for instance, the rotavirus and some types of bacteria like *Pseudomonas aeruginosa*, *Helicobacter pylori*, and *enterotoxigenic Escherichia coli* (ETEC).

Previously, the structural study of TV has been hampered by the low titer that only a few particles can be found under TEM. Therefore, an antibody-based affinity grid method has been developed to enrich the virus particles on the EM grid (Yu et al., 2016). A 2.5 Å resolution TV structure was determined with virus particles captured by the antibody on the grid. However, the virus structure alone does not answer the question of virus-host interactions. In the process of studying the virus receptor complex, a new TV strain (the 9-6-17 strain) was derived in our lab. Confirmed by the experimental results of ELISA (enzyme-linked immunoassay) and HA (hemagglutination assay), it couldn't bind to its host cell receptor, B-type HBGA. This chapter focuses on sequence analysis, structural biology studies, and mutagenesis studies of the 9-6-17 TV strain.

2.2 Material and Methods

2.2.1 Cell culture and purification of Tulane virus

LLC-MK2 cells were cultured with media 199 (Thermo fisher scientific) supplemented with 8% fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin. When the 100 ml cell cultures grew to 90% confluence, it was inoculated with 60 µl Tulane virus stock ($3x10^8$ PFU/ml). After 48h of incubation, the 100 ml culture was collected and used to infect 800 ml of LLC-MK2 confluent cell culture. After 48 hours of incubation, adherent cells were scraped off and the culture was collected and centrifuged at 8000g to remove cell debris. The cell sediments were resuspended and went through three rounds of freeze-thaw cycles to break the cells and release the viruses. Another centrifugation at 8000g was performed to separate the cell debris and the supernatant. The supernatant was collected, combined, and centrifuged at 150,000g for 2 h with the Ti 55.5 rotor (Beckmann, USA). The supernatant was discarded, and the sediment was resuspended in 10 ml PBS overnight at 4°C. The suspension was combined and mixed with CsCl to obtain a solution with a density of 1.34 g/ml by adding 5.11g CsCl (s) to 10 ml suspension. Gradients were formed by centrifugation at 150,000g for 24 hours in an SW 41 Ti rotor using a Beckman Coulter Optima L-90 ultracentrifuge (Beckman, USA). The gradients were fractionated by bottom puncture. SDS-PAGE electrophoresis was performed to identify the fraction with the VP1 band. The fractions that contain virus were combined and concentrated to 250 µl with 0.5 ml 100 kDa centrifugal filter unit (Millipore, MA).

2.2.2 Viral RNA extraction and genome sequencing

The whole TV RNA genome was extracted with QIAamp Viral RNA Mini Kit (Cat No./ID: 52904, QIAGEN, Germantown, MD) according to the manufacturer's protocol. The cDNA library was constructed from the extracted virus RNA using the Illumina TruSeq Stranded Total RNA kit without ribo-depletion. The cDNA library was sequenced using a MiSeq (Illumina, CA, USA) in the Purdue Genomics Core Facility. The genome was assembled from raw data using the SPAdes software and compared with the original Tulane virus sequence (GenBank: EU391643.1). The genome of this 9-6-17 TV will be deposited to the GenBank.

2.2.3 Single-cycle growth curve of 9-6-17 TV and plaque assay

The monolayer of cell culture in a six-well plate was infected with 9-6-17 TV at a multiplicity of infection of 0.043 by covering it with 500 µl virus dilution. After an adsorption period of 2 hours at room temperature, the cell layer was washed with PBS three times to remove the unbound virus. The cell monolayer was then covered with 2 ml of M199 medium with 2% FBS in each well. The six-well plates were incubated in a 37 °C 0.5% CO2 incubator. At each time point (6, 24, 30, 36, 48, 80 hours post-infection), the cell culture media of three wells were collected and frozen at -80 °C. After all time points were collected, the frozen cultures were thawed and clarified with centrifugation. Samples were then plated for plaque assay.

When the cells reached 80% confluency in the six-well plate, cultures were incubated with 250 µl of serial diluted virus for 1.5 hours. Then 2 ml of 1 :1 mixture of 2x M199 medium and 1.2% agarose were added to each well. The plate was rotated to make sure the agarose layer was evenly distributed in the well. After three days, 2ml 0.5% Neutral Red (Sigma, MA) was added and the number of plaques in each well was counted.

2.2.4 Saliva-base ELISA for measurement of HBGA binding

The saliva samples were treated by boiling for 10 min prior to the assay for denaturation of potential antibodies that may interfere with the assay. The expression levels of the A, B, H-type 1, H-type 2, and Lewis's antigens in the saliva were determined previously using anti-H type 1 (BG-4), anti-Leb (BG-6), and anti-Ley (BG-8) (Signet Laboratories Inc. Dedham, MA) MAbs, and anti-H type 2 (BCR 9031), anti-A (BCR 9010), and anti-B (BCRM 11007) MAbs (Accurate Chemical & Scientific Corporation, Westbury, NY). To test HBGA binding by the TV strains, saliva samples diluted with PBS at 1:1000 were coated onto plates at 4 °C overnight and then incubated with a serially diluted TV preparation. The salivary HBGA bound TVs were detected by mouse anti-TV serum (1:3500) and then by an HRP conjugated goat anti-mouse IgG (1:5000). The color signal was developed with the TMB (3,3',5,5'-Tetramethylbenzidine) and OD was read at a wavelength of 450 nm.

2.2.5 Cryo-EM sample preparation and data acquisition

The purified 9-6-17 virus (2 μg/ml) was applied to the graphene oxide coated Quantifoil grid (Ted Pella Inc, USA), then frozen with Cryoplunge 3 system (Gatan, USA). The frozen samples were imaged on a 300 kV Titan Krios electron microscope (Thermo Fisher Scientific, USA) equipped with a post-GIF K2 summit camera mounted on a 20 eV slit Quantum energy filter (Gatan, USA). Three datasets were collected for 9-6-17 TV. Detailed data collection conditions are listed in Table 2.1.

2.2.6 Image processing

The movies were aligned with Motioncor2/1.0.5 (Zheng et al., 2017) and 1.5x binned. Subsequently, the CTF determination of dose-weighted micrographs was performed with CTFFIND4 (Rohou and Grigorieff, 2015). The particles were picked by cisTEM (Grant et al., 2018) and imported into cryoSPARC (Punjani et al., 2017) for 2D classification, ab-initial reconstruction, and homogeneous refinement with icosahedral symmetry. The particle numbers retained in each step are listed in Table 2.1. After obtaining 3.15 Å resolution from cryoSPARC, we further refined the map to 2.63 Å resolution with JSPR (Guo and Jiang, 2014) based on the 0.143 criteria of "gold standard" Fourier shell correlation of the maps built from half-datasets. The local resolutions were evaluated using RELION/3.0 (Zivanov et al., 2018). The EM density map of 9-6-17 TV will be deposited to EMDB.

2.2.7 Model refinement

The PDB model of the previous TV structure was fitted into the cryo-EM maps of the 9-6-17 TV with *Chimera* (Pettersen et al., 2004). The asymmetric unit was cut out from the entire virus map and refined with *Rosetta* (Wang et al., 2016) and the GNU parallel system (Tange et al., 2018). The model with the best FSC score was selected for further analysis. The model statistics are shown in Table 2.1.

Table 2.1 9-6-17 TV EM data collection and refinement statistics

	9-6-17 TV	Without DTT	With DTT
Data Collection			
Date	5-12-2018	9-14-2018	
Grid	Lacey grid coated with graphene oxide	Quantifoil grid coated with graphene oxide	
Voltage (kV)	300		
Total dose ($e^-/Å^2$)	23	25	
Nominal Magnification	81,000	105,000	
Number of Movies	969	503	771
Physical Pixel Size (Å)	1.73	1.384	
Intended Defocus (μm)	0.8-2.2	1.5-2.5	
Number of frames	35	30	
Image processing			
Particles picked	48,505	20,260	132,461
Number of particles after 2D classification	27,187	18,397	44,239
Final number of particles	11,645	16,777	44,239
Resolution (Å)	3.2	2.73	2.63
Refinement			
Map CC	0.71	0.76	0.79
All-atom clashscore	8.99	3.38	4.29
Rotamer outliers (%)	0.00	0.23	0.23
Ramachandran plot			
Outliers (%)	0.20	0.59	0.00
Allowed (%)	8.70	9.09	8.30
Favored (%)	91.11	90.32	91.70

2.2.8 Expression and purification of TV VLP

The viral RNA was extracted from the 9-6-17 TV with the QIAamp Viral RNA Mini Kit from QIAGEN (52904). A single-strand cDNA was generated with the ProtoScript® II First Strand cDNA Synthesis Kit from NEB (E6560S). To generate the pFastBac vector for TV VLP, the fragment containing the sequence of both VP1 and VP2 was PCR amplified from the first strand cDNA and gel extracted. The purified PCR product was inserted into the linearized pFastBac Dual expression vector via In-Fusion reaction. The reconstructed vector was verified by both sequencing (Genewize, USA) and PCR. The K367S mutation was introduced to this pFastBac-VP1VP2

vector. The Bac-to-Bac baculovirus expression system (Invitrogen) was used to express TV VLP. Baculovirus generation and expression in insect cells were performed as described by the manufacture's User Guide.

The infected cells were collected 72 hours post-infection. The cells were lysed completely with a 40ml dounce homogenizer in ice-cold PBS. The lysed cells were centrifuged in a JA-10 rotor at 8000g for 30 min at 4 °C. The resuspension was pelleted through a 20% sucrose cushion in a Beckman SW 28 Ti rotor at 100 kg for 1 hour. The pellet was resuspended in 0.5 ml PBS overnight at 4 °C. The resuspension was applied on top of the OptiPrep (Sigma-Aldrich, Inc., USA) density gradient (OptiPrep layers from bottom to top: 0.5ml 54%, 2ml 45%, 2ml 40%, 2ml 30%, 2ml 20%, 2ml 15%) in a SW 41 Ti rotor and centrifuged at 100 kg for 4 hours at 4 °C. The fractions were collected by bottom puncture and analyzed by SDS-PAGE.

2.2.9 Identification of 9-6-17 TV host cell receptor with mass spectrometry

300 μl 3.4 μg/ul purified virus sample was incubated with 300 μl of 10 mg/ml Pierce™ NHS-Activated magnetic beads (Catalog. No. 88827, Thermofisher Scientific, USA) or FG-NHS magnetic beads (Nacalai tesque) at room temperature for two hours. The reaction was quenched by adding 1mL of 1M Tris 150mM NaCl pH 7.7 to the beads and incubating at RT for 30min. The virus conjugated magnetic beads were washed with 1ml PBS three times.

1ml virus conjugated magnetic beads with 600 μl PBS was applied to one T-25 flask with confluent cell and incubated at 4 °C for two hours. The crosslinking reagent BS^3 was added to the reaction mixture to a final concentration of 5 mM and incubated for 30 min at room temperature. 1ml 1M Tris pH 7.5 was applied and incubated for 15 min at room temperature to quench the crosslinking reaction. The adherent cells and magnetic beads were scraped off the flask and washed with 1ml ice-cold PBS twice. The mixture was resuspended in the lysis buffer (1% NP-40, 50mM Tris pH 7.7, 150mM NaCl, 20 ul protease inhibitor) and incubated for 30 minutes at room temperature, with vertexing at 10-minute intervals. The receptor-bound magnetic beads were washed with ice-cold PBS three times and stored at -80°C until ready to use. The negative control sample was prepared with the same procedure described above without adding any virus.

The mass spectrometry analysis was performed by Purdue proteomics facility with ESI-Orbitrap.

2.2.10 Construction of the full-length cDNA clone of 9-6-17 TV strain (icTV-9-6-17)

The cDNA fragments F1 and F2 were successfully synthesized by the company Twist Bioscience (South San Francisco, CA) and cloned into high copy plasmid pTwist Amp High Copy respectively. The F1 contains a T7 promoter sequence upstream of the 5' end of the TV sequence. A poly$(A)_{30}$ sequence was introduced by PCR to the 3' end of F2 before the restriction enzyme site. A 186 bp antisense sequence with T7 promoter was inserted into the backbone of the F1 vector to tackle the toxicity problem. The F2 fragment was cut out with the restriction enzymes, AfIII and KpnI. The vector containing the F1 fragment was linearized with the same restriction enzymes. The linearized F1 vector and the F2 fragment were gel purified and in vitro ligated with the In-Fusion HD Cloning Plus Kits (Takara Bio USA, lnc.) at 50 °C for 30 min then placed on ice. The infusion products were transfected into DH5α cell line according to the manufacturer's protocol. The plasmids were extracted and verified with sequencing and PCR.

2.2.11 Recovery of isogenic recombinant TV variants K367 and S367

To recover recombinant 9-6-17 TV from the infectious cDNA clone (icTV-9-6-17), we used the chemical transfection reagent TransIT®-mRNA (Mirus, WI, USA) to transfect the 9-6-17 TV mRNA with or without K367S mutation into LLC-MK2 cells. The transfected cells developed cytopathic effects (CPEs) on day 1 after transfection. After two days of transfection, all the cells were dead. The P0 infected cultures were collected and used for amplification. The P2 infected culture was collected and used for virus purification. The virus purification step is the same as above.

2.2.12 Virus quantification with qRT-PCR

The real-time qRT-PCR was performed with the Luna® Universal One-Step RT-qPCR Kit (NEB) with the extracted viral RNA. Reactions, including a standard curve, were generated per the manufacturer's instruction. The one-step RT-PCR primers are listed in Table 2.2.

Table 2.2 Primers for cloning and sequencing

Primers for cloning K367S mutation	Forward primer	Reverse primer
9-6-17 TV K367S	CCTACCAAAGGACCA TCTACATTAGTGACC CATCCCTGG	CCAGGGATGGGTCAC TAATGTAGATGGTCCTTT GGTAGG
Primers for one-step RT-PCR		
TV 4457F-4572R	GCTGTCTCTTCACTCTCC TTG	CCAGGTTGTGTAGCACTA GATAC
Primers for Sanger sequencing of VP1 VP2		
Segment 1	GCTCACGCTAAGTTTGAA CC	GCACGTTGCGTAACTGG
Segment 2	AGAGGATCCCAAAATAA TGCACTGCCAATTTG	CCCAAGCTTTTATCTAAA GACAACTGCTGTCTG
Segment 3	CCTGCTGATGGTTATTTC AG	GCTACATGGGCTTAGGGT CTCACT

2.2.13 HBGA-triggered genome release measured by negative stain

Three types of HBGA were used for the HBGA-triggered conformational dynamics of Tulane virus: B-Trisaccharide (Sigma® B1422), blood type B (tri)-PAA-biotin (GlycoTech® 01-085) and blood type B (tri)-sp-biotin, sp=(CH2)3NHCO(CH2)5NH- (GlycoTech® 02-085). Purified 9-6-17 Tulane viruses were incubated with 1mg/ml synthetic type B HBGA in a humid chamber for different lengths of time at 37°C, followed by negative stain TEM imaging. Briefly, 3ul incubated sample was applied to glow discharged grid and incubated for 1min. The three drops of 2% PTA (phosphotungstic acid) were applied to the grid and blotted away sequentially. The negative stain grids were examined under the 200kV TEM Tecnai T20 (Thermofisher Scientific, USA).

2.2.14 Disulfide bond detection with SDS PAGE

Homemade protein loading buffer is made up of 100mM Tris-Cl pH8.6, 4% SDS, 20% glycerol, and 200mM DTT. The protein loading buffer without the reducing reagents contains the same substances but without DTT. Purified virus samples were incubated in 2x loading buffers for 30 min before loading on the gel.

2.3 Results

2.3.1 Sequence analysis of the TV strains

The whole-genome sequencing of the 9-6-17 TV strain was performed with extracted viral RNA of the purified virus. Compared to the wild-type TV (GenBank: EU391643), the first eight nucleotides at the 5' end of the 9-6-17 TV genome are missing, and 31 nucleotide mutations are identified in the 9-6-17 TV which is about 0.45% of the whole genome length (Table 2.3). Among the 18 amino acid mutations resulting from the 31 nucleotide mutations, eight amino acid mutations were located in VP1 which is responsible for the receptor binding in the infection process. The eight mutations in VP1 are N3S, N284H, F334V, A335E, A343T, S367K, I451M, and R452C.

Table 2.3 Statistics of the 9-6-17 TV sequence

	9-6-17 TV
Sequence length (bp)	6701
Number of nucleotide mutations in total	31
Number of nucleotide mutations in ORF2	9
Number of nucleotide mutations in ORF3	7
Number of amino acid mutations in total	18
Number of amino acid mutations in VP1	8
Number of amino acid mutations in VP2	5

We obtained the 11-25-12 TV strain from our collaborator as the earliest strain cultured from the wild-type TV. We designed three sets of primers to amplify three segments of DNA covering the entire sequence of VP1 and VP2 for sequencing. To purify this virus, it was passages two rounds in the LLC-MK2 cell line. We sequenced the VP1 and VP2 regions of the 11-25-12 TV strain before and after the amplification. We performed multisequence alignment of the wild-type TV, the 11-25-12 TV, the 11-25-12 TV after amplification, and the 9-6-17 TV (Fig. 2.1). It is not surprising to see mutations in 9-6-17 TV that already existed in the 11-25-12 TV strain. Although only one mutation existed in VP1 of the 11-25-12 strain, the VP2 sequence of the 9-6-17 and 11-25-12 strains are almost identical (Fig. 2.2). It is reasonable to hypothesize that the 9-6-17 strain is probably further evolved based on the 11-25-12 strain. With only two passages, the 11-25-12 TV amplified sequence has obtained four mutations in VP1 (E72D, T271I, F334V, A343T) in which the F334V and A343T also existed in the 9-6-17 TV sequence. It suggests that

it might be the same driving force for these two mutations in the established cell culture environment. Most surprisingly is that the isoleucine at position 367 in 11-25-12 TV has mutated back to serine in the amplified sequence, while in the 9-6-17 TV, the position 367 has mutated into lysine.

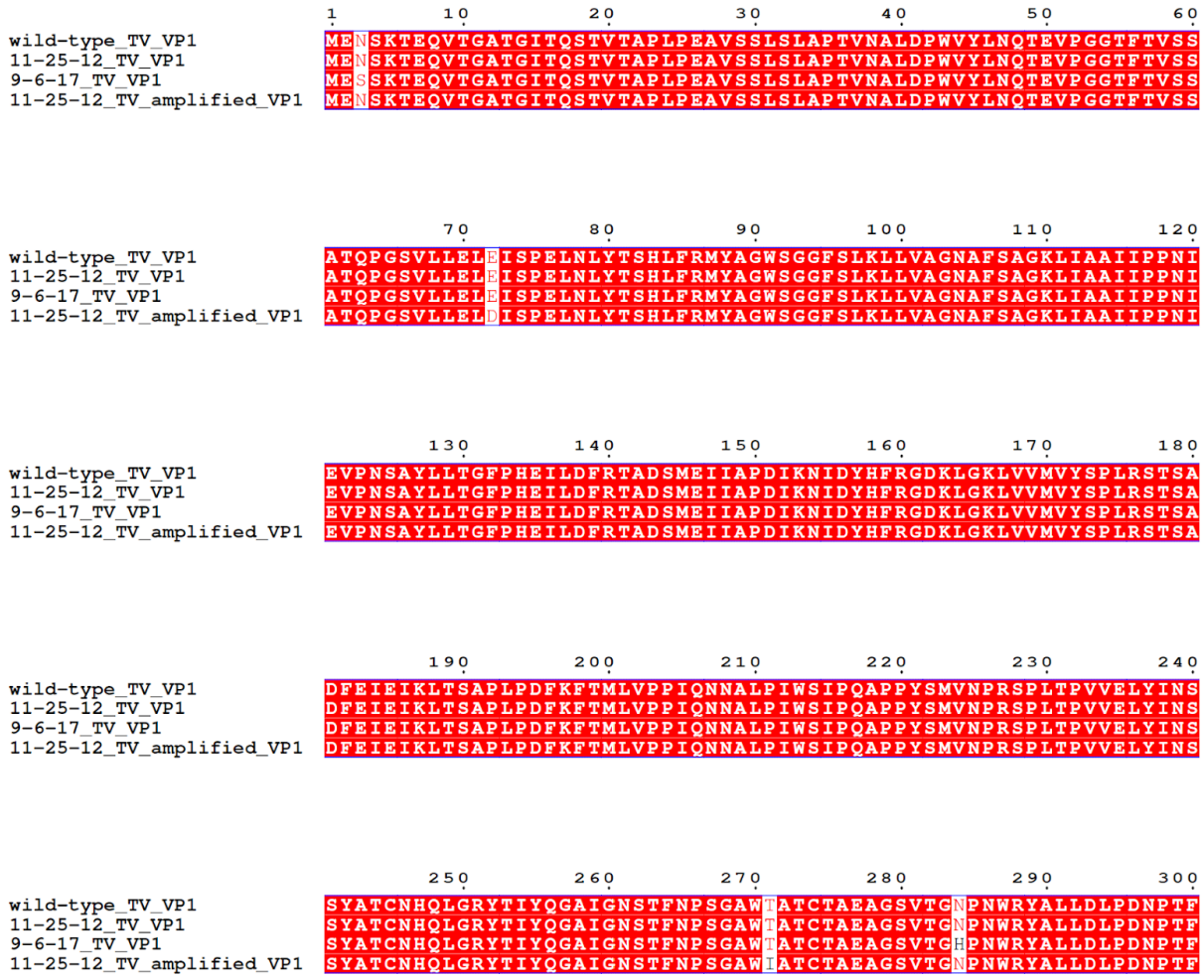


Figure 2.1. Sequence alignment of VP1 of the wild-type TV, the 11-25-12 TV, the 11-25-12 TV amplified, and the 9-6-17 TV strains. The wild-type TV sequence was obtained from the GenBank under the accession number EU391643. The sequence alignment was performed by Clustal Omega (Madeira et al., 2019) and displayed with ESPript 3.0 (Robert et al., 2014).

Figure 2.1 Continued

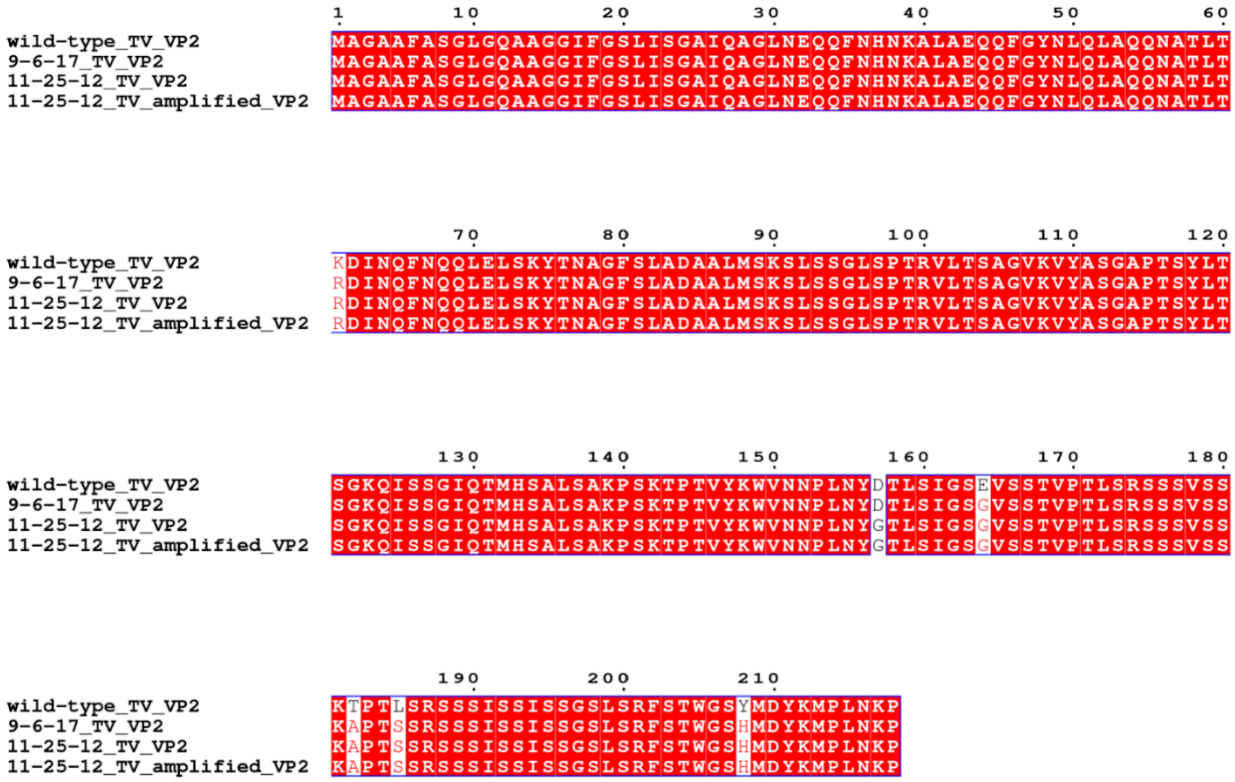


Figure 2.2. Sequence alignment of the minor capsid protein VP2 of the wild-type TV, the 11-25-12 TV, the 11-25-12 amplified TV, and the 9-6-17 TV strains. There are six mutation sites in VP2. Among them, K61R, E164G, T182A, L185S are found in all strains except the wild-type, suggesting their predominance and the potential phenotypic effects. The D157G mutation only exists in the 11-25-12 TV and its amplified version.

To distinguish adaptive mutations from random mutations, we examined the VP1 sequence of other ten *Recovirus* strains. We compared the amino acid identity at each mutation site in VP1 in these *Recovirus* strains and listed them in Table 2.4. Among the eight mutation sites, only A343T is a conservative substitution. Nine of the homolog proteins have alanine at position 343 but there is one homolog protein that has threonine. For position 3 of VP1, six homolog proteins in the same genus have the same amino acid serine as in the 9-6-17 TV instead of asparagine in the wild-type TV. It suggests that this mutation in 9-6-17 TV is probably evolutionarily favored over the residue asparagine at position 3 since it is more frequently found in other *Recovirus* strains. The same situation applies to N284H and F334V. For the N284H mutation, it has the

structural basis that it can form π–π stacking interaction with the tryptophan at position 287 as shown in Fig. 2.3.

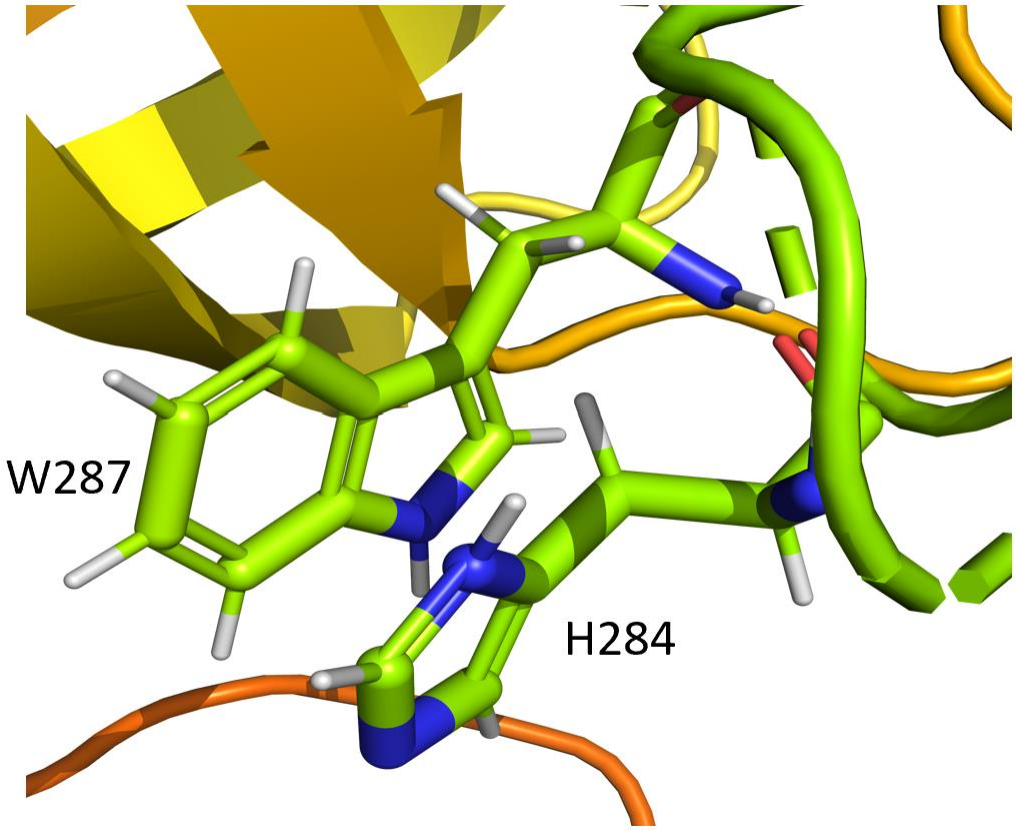


Figure 2.3. The interaction of H284 and W287 in the 9-6-17 TV structure.

At position 335, all the ten *Recovirus* sequences are aspartic acid, while the 9-6-17 TV has a glutamic acid replacing alanine at that position. The 335 position is at the top end of the P domain. It is near the dimer interface but is not directly involved. The side chain of Asp^{335} protrudes out and is not interacting with neighboring residues. Therefore, this residue is very likely to involve in receptor binding. Since both aspartic acid and glutamic acid are negatively charged, it must be essential to have negatively charged residues at this position.

The S367K and the conjunct mutations I451M and M452C are unique in our 9-6-17 TV. It is not observed in any other *Recoviruses*. Interestingly, the 11-25-12 TV has isoleucine at position 367 as the other four virus sequences. After the amplification, it mutated back into serine. Serine is a polar residue, while isoleucine is hydrophobic. They may have the steric effect on HBGA binding at position 367.

Table 2.4 The residue identity at the 8 mutation sites in VP1 of 10 *Recovirus* homologues in *Calicivirus* family

a.a. position	a.a. in wild-type TV	a.a. in 9-6-17 TV	10 *Recoviruses*
3	N	S	S (AHE37931.1;AHE37934.1;AHE37928.1;AHE37919.1; AHE37916.1; AHE37913.1), K (AHE37922.1;AFV48067.1;AHE37925.1;AHE37947.1)
284	N	H	H (AHE37919.1; AHE37916.1; AHE37913.1), Q (AHE37925.1; AHE37947.1; AHE37931.1; AHE37934.1; AHE37928.1), N (AFV48067.1), K (AHE37922.1)
334	F	V	V (AHE37922.1;AFV48067.1;AHE37925.1;AHE37947.1;AHE37919.1;AHE37916.1;AHE37913.1), T (AHE37931.1; AHE37934.1; AHE37928.1)
335	A	E	D (AHE37922.1; AFV48067.1; AHE37925.1; AHE37947.1; AHE37919.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1)
343	A	T	A (AHE37922.1; AFV48067.1; AHE37925.1; AHE37947.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1), T (AHE37919.1)
367	S	K	S (AHE37919.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1), I (AHE37922.1; AFV48067.1; AHE37925.1; AHE37947.1)

Table 2.4 Continued

451	I	M	I (AHE37922.1; AFV48067.1; AHE37925.1; AHE37947.1; AHE37919.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1)
452	R	C	R (AHE37922.1; AFV48067.1; AHE37925.1; AHE37947.1; AHE37919.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1)

References: (AHE37922.1; AHE37925.1; AHE37947.1; AHE37919.1; AHE37916.1; AHE37913.1; AHE37931.1; AHE37934.1; AHE37928.1) (Farkas et al., 2014); AFV48067.1 (Handley et al., 2012).

2.3.2 HBGA-triggered genome release

In our previous study, we have observed HBGA-triggered genome release of the Tulane virus. After on-grid incubation of purified TV with synthetic type B HBGA (100 μg/ml, GlycoTech® 02-085) for 30-60 minutes at room temperature, a significantly increased number of empty particles were observed in the negative stain images compared to the untreated sample. Some protruded particles were observed, and we hypothesized that they represent the intermediate state of TV genome release. Therefore, we performed the same procedures with the purified 9-6-17 TV sample. However, no obvious morphology change was observed. The majority of the particles are intact with only less than one percent of the particles are empty which is identical to the untreated sample. We increased the incubation time to 24 hours at room temperature and took 24 micrographs of the negative stained grid with the pixel size of 1.4 Å. 111 virus particles were picked manually. The representative negative stain image is shown in Fig. 2.4A. Both the 2D class averages (Fig. 2.4B) and the reconstructed map showed no difference from the untreated virus sample. We have also plotted the 1D radial intensity curve to gauge the diameter of the particles and the percentage of the genome inside. The 1D profiles of the experimental group were almost identical to that of the untreated sample.

We have explored different experimental conditions by varying the incubation time (0.5, 1, 1.5, 2, 3, 24, 48 hours), the amount of virus, the types of HBGAs (B-Trisaccharide, type B (tri)-PAA-biotin, type B (tri)-sp-biotin), the temperature (4°C, 22°C, 37°C). However, no significant differences were observed.

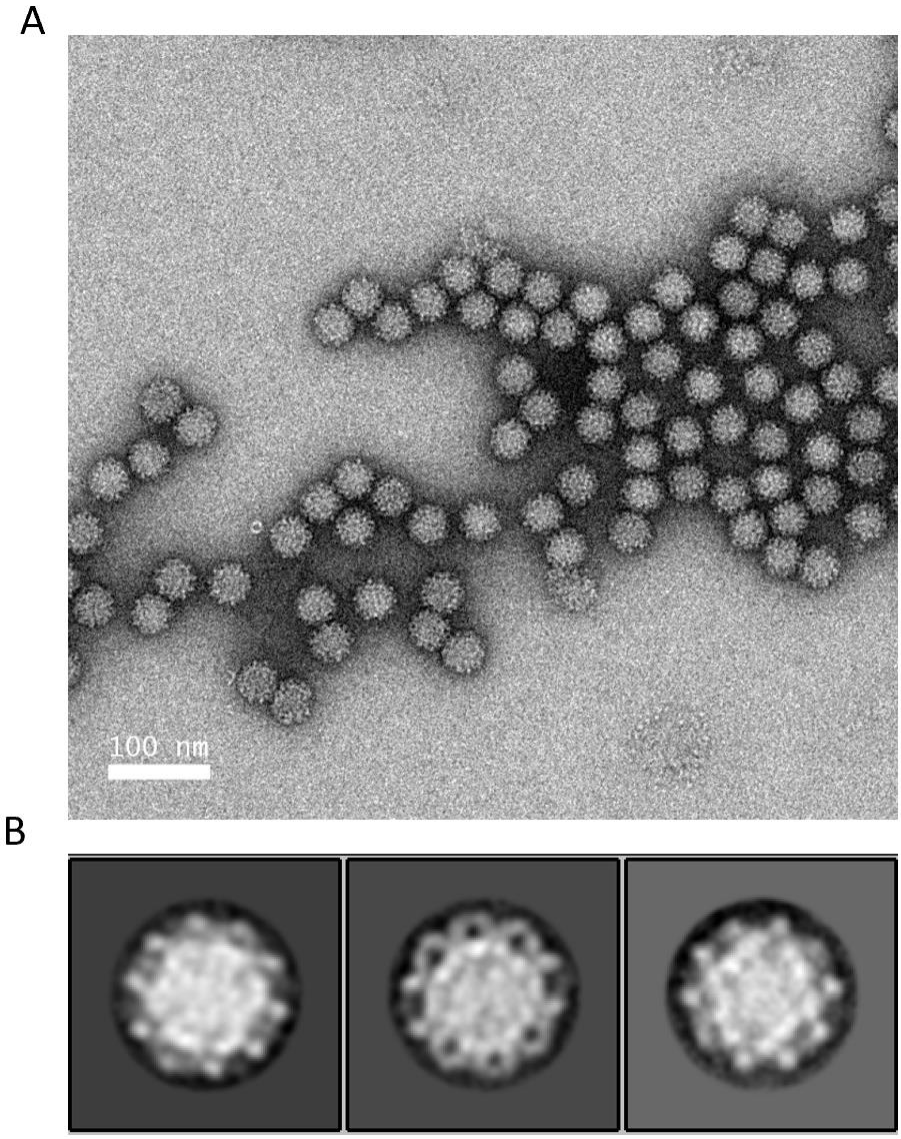


Figure 2.4. HBGA-triggered genome release measurement by negative stain. (A) The representative negative stain image of 9-6-17 Tulane virus. The scale bar is 100 nm. (B) Three 2D class averages are showing packed genome inside. They have the same diameter as the nontreated virus.

2.3.3 Structure determination of the 9-6-17 TV

Three EM datasets were collected for 9-6-17 TV. The first dataset is with opposite contrast that the virus particle is brighter than the background (Fig. 2.5A). We hypothesis that it was caused by the residue iodixanol (OptiPrep) in the buffer that wasn't fully dialyzed away. Iodixanol is the density gradient that we used in the final purification step. It has several iodine atoms in its structure (Fig. 2.5B) that are able to scatter electrons stronger than virus particles. Therefore, the residue iodixanol in the buffer created a darker background in the EM image. Despite the opposite contrast, the reconstruction was performed as normal, and the map contrast can be easily inverted by multiplying -1 to the map pixel intensities (Fig. 2.5C). In the end, a 3.2 Å resolution map was generated and the atomic model was derived from it.

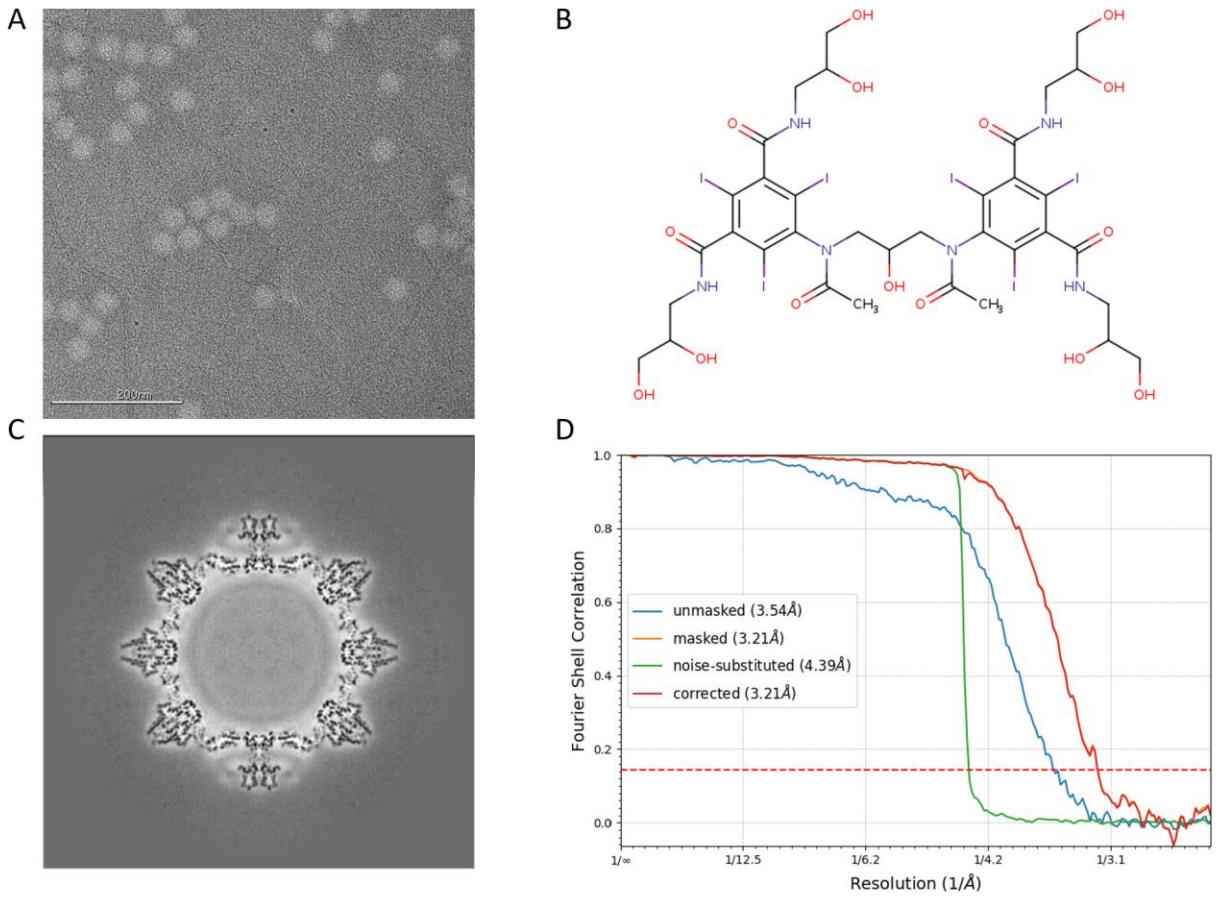


Figure 2.5. The first dataset of 9-6-17 TV. (A) The representative micrograph shows the opposite contrast due to the iodixanol in the background. The scale bar is 200 nm. (B) The chemical structure of iodixanol (OptiPrep). (C) The central section of the reconstructed map before contrast inversion. (D) The FSC curve of the final reconstruction.

The other two datasets were designed to explore the effects of the Cys^{452}-Cys^{452} disulfide bond in the virus structure. To reduce this disulfide bond, we treated one group with DTT while the other one serves as the control group. These two datasets were using the same batch of viruses and are collected on the same day. The datasets with and without DTT have reached 2.63 and 2.73 Å resolution respectively. However, the reconstructed maps from these two datasets are almost identical in that there is no obvious difference in the disulfide bond density. These two datasets are of similar quality. Therefore, here we only presented the cryo-EM image and 2D class averages of the dataset without DTT treatment (Fig. 2.6).

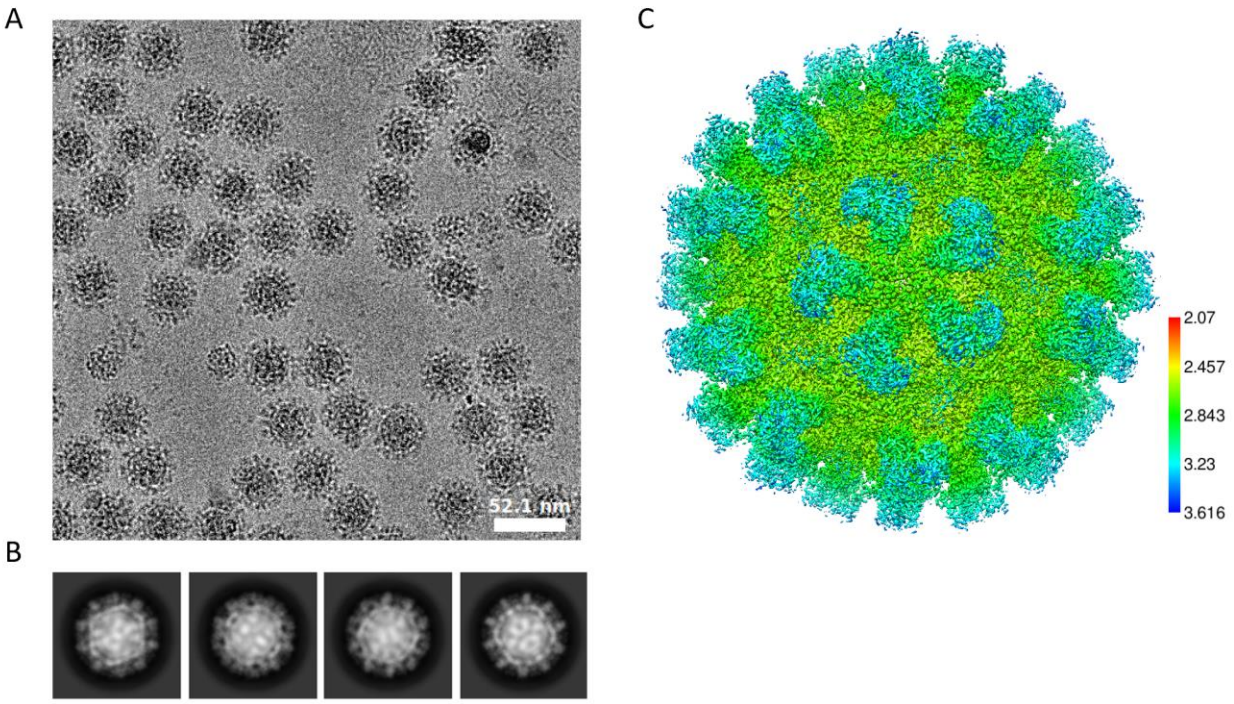


Figure 2.6. The 9-6-17 TV cryo-EM dataset in which the virus sample wasn't treated with DTT. (A) Representative image without low pass filter showing high contrast virus particles. The scale bar is 52.1nm. (B) The 2D class averages show different virus orientations with clear spikes on the virus surface. (C) The local resolution map of the 2.73 Å structure was reconstructed from this dataset.

The gold standard Fourie Shell Correlation (FSC) curves of these two datasets are shown in Figure 2.7A. The DTT treated dataset has yielded a map of 2.63 Å resolution. Based on the estimated local resolution (Fig. 2.7 B), the resolution of the S-domain shell has an average resolution of 2.52 Å, while the P domain regions are more flexible that are round 2.86 Å resolution.

Except for the first 20 residues at the N-terminal of subunit A and B, all residues in both subunit A and B were fitted into the EM density. Almost all residue densities at the N-terminal are visible in subunit C. Two representative volumes of density are displayed in Fig 2.7 D overlapping with the atomic model. Compared to the no reducing reagent map, the DTT treated map has a large blub of extra density at the top of the P domain near residue His284 and Pro285 as shown in Fig. 2.7E. We are not certain of its identity. The same batch of viruses was used for structure determination with and without reducing reagent. But only the map reconstructed from the DTT treated dataset has this large extra density. Therefore, it must be a result of the DTT addition. However, the DTT molecule itself is too small to encompass the whole volume of the extra density.

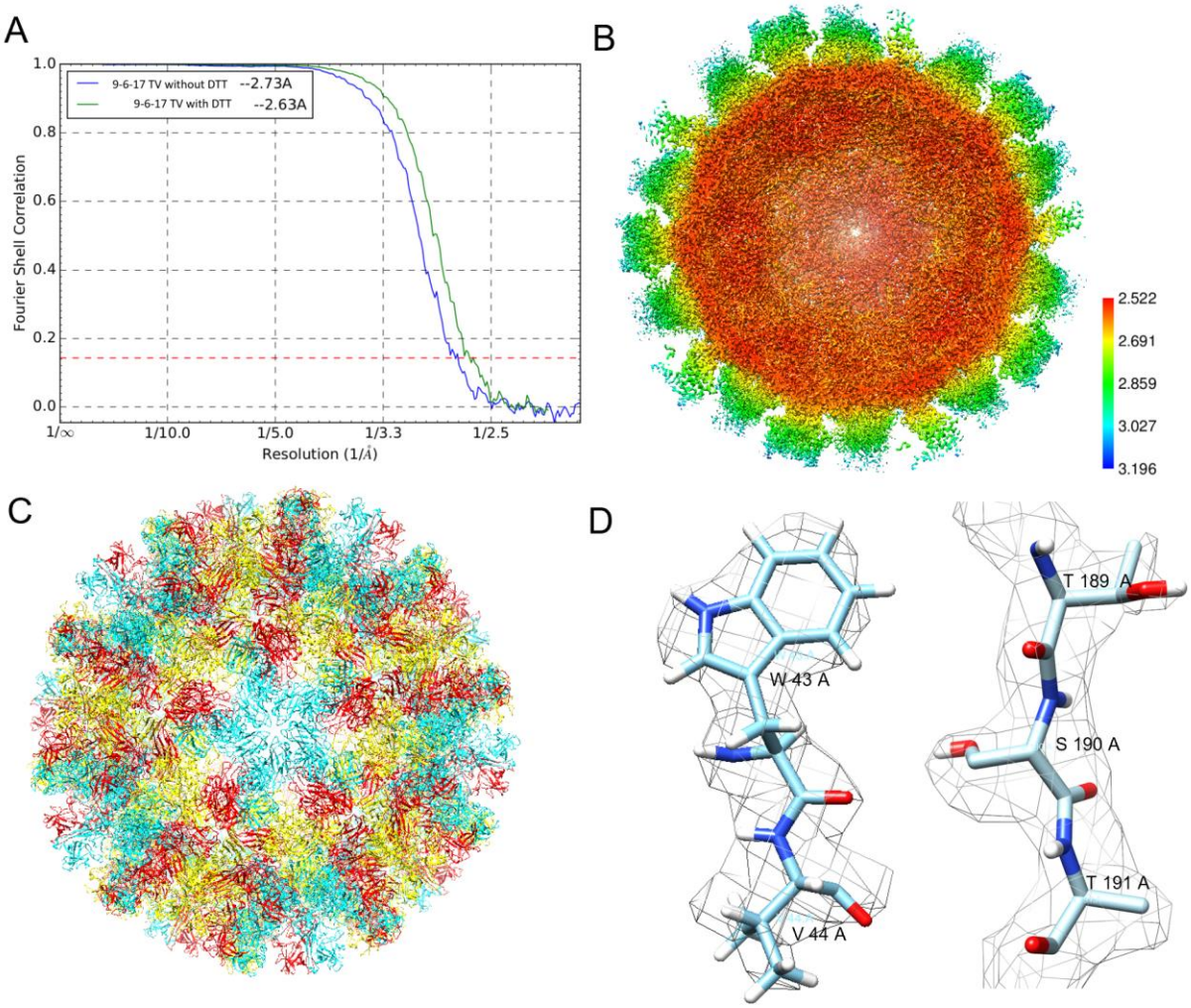
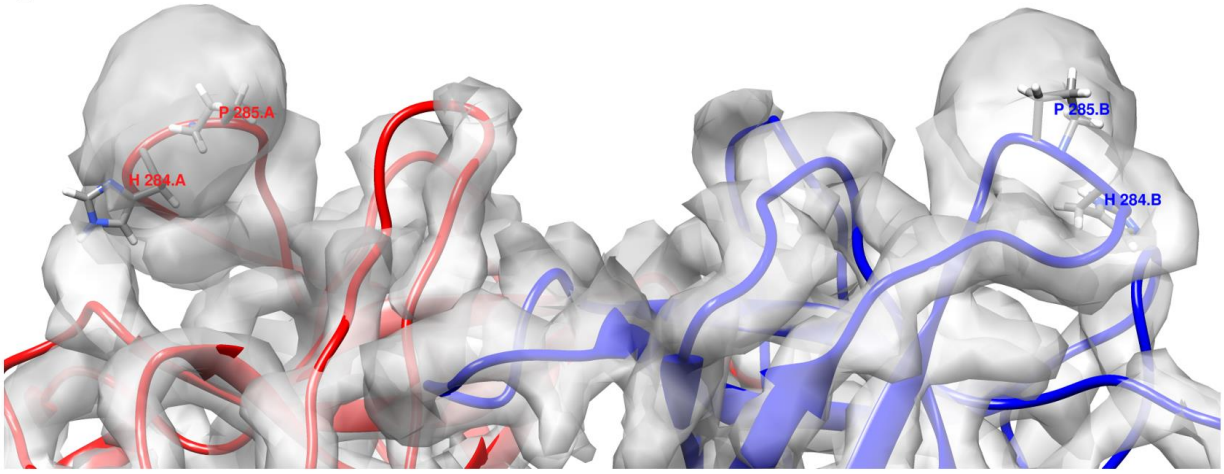


Figure 2.7. The 9-6-17 TV datasets with and without DTT treatment. (A) The gold-standard Fourier shell correlation (FSC) curves of these two datasets. (B) The cross-section of the reconstructed map with DTT treatment shows the estimated local resolution. (C) The ribbon diagram of the full capsid shows the T=3 icosahedral organization with subunit A (blue), subunit B (red) and subunit C (yellow) colored respectively. (D) Two segments, a.a. 43-44 and a.a. 189-191 are selected to present the quality of the electron density map. The electron density map is overlaid on the final refined structure of subunit A. (E) Extra density near His284. It is observed on all subunits. The color scheme of chain A and chain B is consistent with the whole virus model shown in C.

Figure 2.7 Continued

E



We refined the 9-6-17 TV model for all three datasets based on the first atomic model of TV. They are identical except for some variations at the loop regions. We choose the model of the DTT treated dataset to highlight the location of the eight mutation sites in VP1 (Fig. 2.8). The dimer of subunits A and B were refined as a whole to account for the effect of the disulfide bond on the dimer structure. Surprisingly, the distance of the two subunits in a dimer doesn't change much but it is also close enough for the two cysteines to form the disulfide bond. The insertion is showing the four mutation sites presented at the top of the P domain.

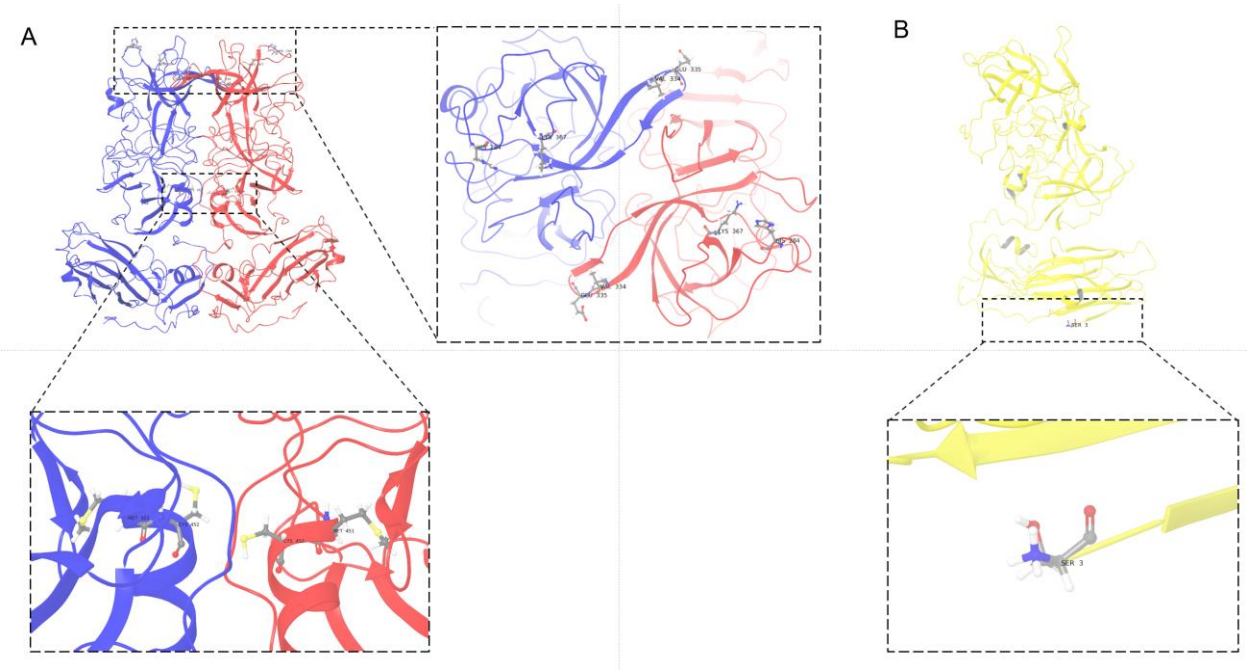


Figure 2.8. The atomic model of 9-6-17 TV DTT treated. (A) The refined model of the AB dimer with seven of the eight mutation sites (except for N3S) was displayed and labeled. The insertion on the right is showing the 90-degree tilted view of the top of the P domain containing 4 mutation sites. Bottom insert, the close-up view of the two mutations at the dimer interface. (B) The refined model of subunit C with the mutation N3S labeled.

Among the 18 mutation sites of the whole genome, 8 of them are responsible for the changes in the capsid protein VP1. One of the mutation sites of VP1 (N3S) is at the very beginning of the N-terminal ends, while all the other 7 mutation sites are located on the P domain of VP1. The density of amino acid position 3 is only visible in subunit C (Fig. 2.8B). Both subunits A and B start from amino acid position 20. It indicates that the N-terminal arms of the subunit A and B are more flexible than that of the subunit C. However, it is hard to conclude if this flexibility is associated with any biological functions, such as involving in the virus genome packing or interacting with VP2. The side-chain densities in the EM map at the mutation site N3S, N284H, F334V and A335E, A343T, S367K, I451M, and R452C largely agree with the mutated amino acid identity as shown in Fig 2.9.

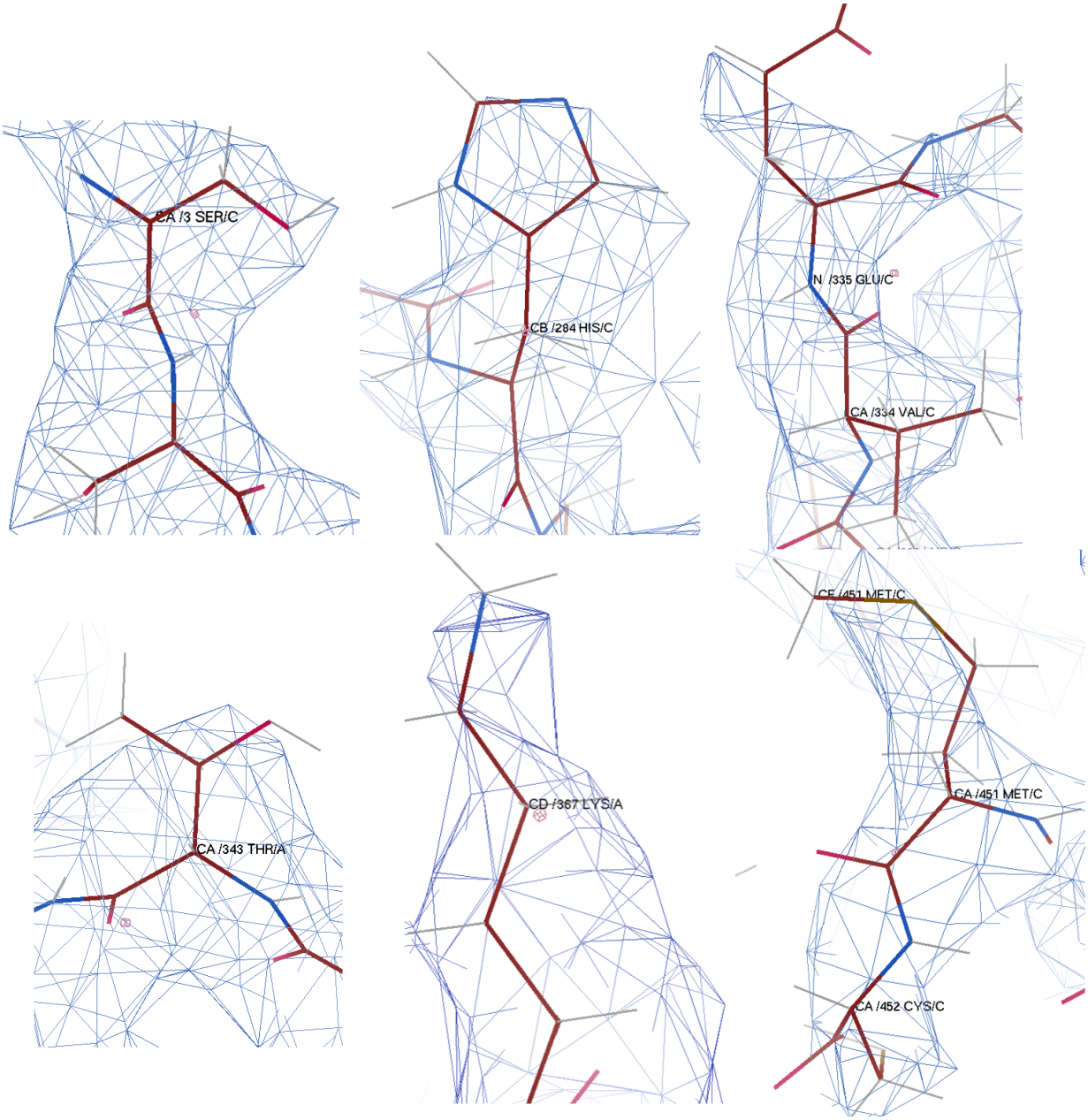


Figure 2.9. The electron density at the eight mutation sites in VP1 for the DTT treated dataset. From left to right, top panel: N284H, F334V, and A335E; bottom panel: A343T, S367K, I451M, and R452C. The S^3 density is from subunit C.

2.3.4 Disulfide bond formed from mutation stabilize dimer interaction

When we examined the density map of the mutate TV, we found a bridge of density connecting the density of A/B and C/C dimers at the position of Cys^{452} (Fig. 2.10B). This density is entirely consistent with a Cys^{452}-Cys^{452} disulfide bond within all the dimers. It is reasonable to hypothesize that this mutation is preserved to stabilize the dimerization of the P domains. To prove the dimerization of VP1 in the existence of the disulfide bond, we performed SDS PAGE with and without the reducing reagents (DTT and beta-ME) in the loading buffer and observed the dimer band at around 120kD molecular weight in the without DTT sample (Fig. 2.10A). However, the EM density of the disulfide bond in the DTT retreated map is still strong even at a low threshold. It indicates that the on-grid 3mM DTT 15min treatment is probably not enough to reduce the disulfide bond of all viruses.

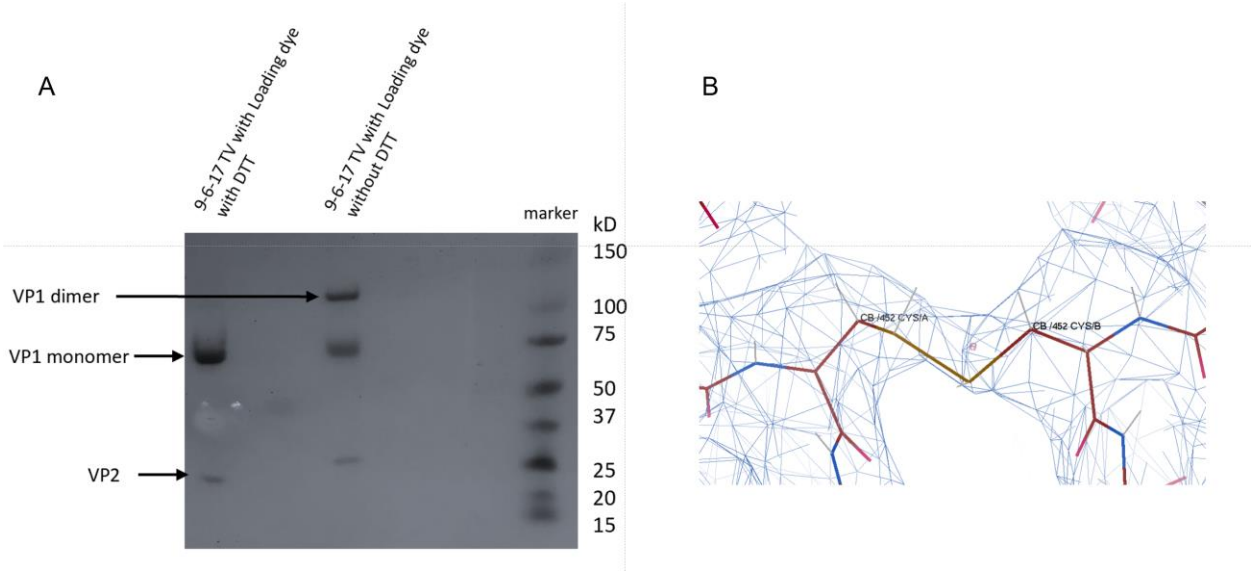


Figure 2.10. The disulfide bond formed by C452-C452. (A) The SDS-PAGE result of 9-6-17 TV with or without DTT treatment. Without reducing agents, the 9-6-17 TV showed a dimer band at around 120kD position, while the one with reducing agents doesn't have the dimer band. (B) The density of the disulfide bond of C452 at the dimer interface.

2.3.5 Elongated extra density in the hydrophobic pocket in P dimer

An elongated extra density (Fig 2.11) was found between the AB subunit near residues P425, V341, and M352. It has a volume of 93.71 $Å^3$ and a surface area of 219.4 $Å^2$. The pocket around it is mostly composed of hydrophobic residues (F329, V341, I380, M382, P425) with only one charged residue D342, and one polar residue T432. Picornaviruses are known to have a pocket factor at the five-fold canyon to stabilize the virus capsid and once the virus binds to its receptor, the pocket factor is dislodged and induced genome release (Plevka et al. 2013). However, it has not been reported in *Caliciviruses* before. The pocket factor of human enterovirus 71 (EV71) was modeled as lauric acid. We tried to fit the unmodified lauric acid ($C_{12}H_{24}O_2$) into the extra density. It has the exact length of the extra density. Only that the extra density seems to have C2 symmetry while the lauric acid doesn't have symmetry. This extra density exists in both A/B and C/C dimers which means it is likely that its C2 symmetry appearance is not an artifact of image processing. Because for icosahedral reconstruction, the asymmetric unit was processed as a single unit. The A/B dimer is within the asymmetric unit so that their dimer interface would not be affected by the symmetrization process when applying icosahedral symmetry.

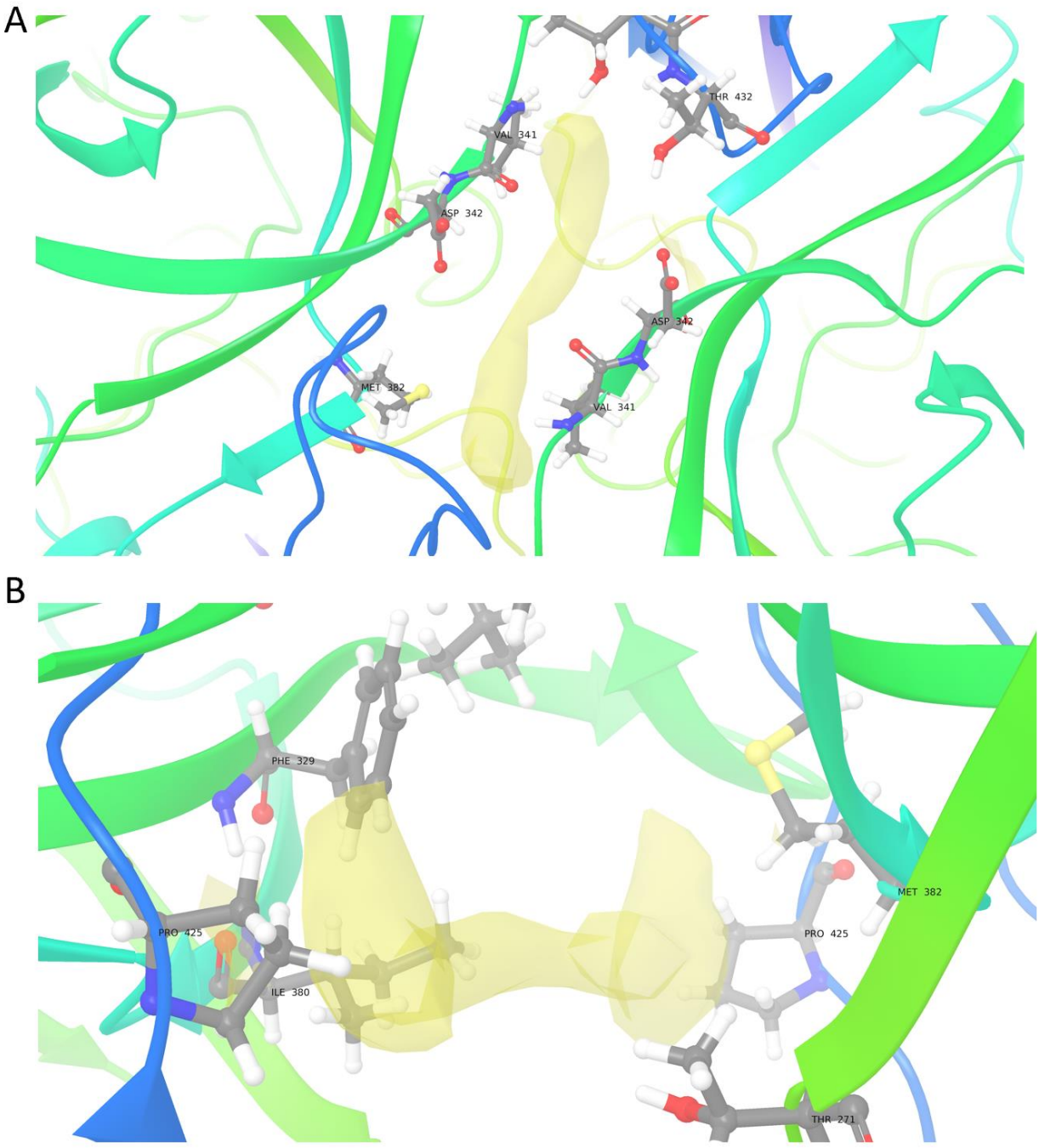
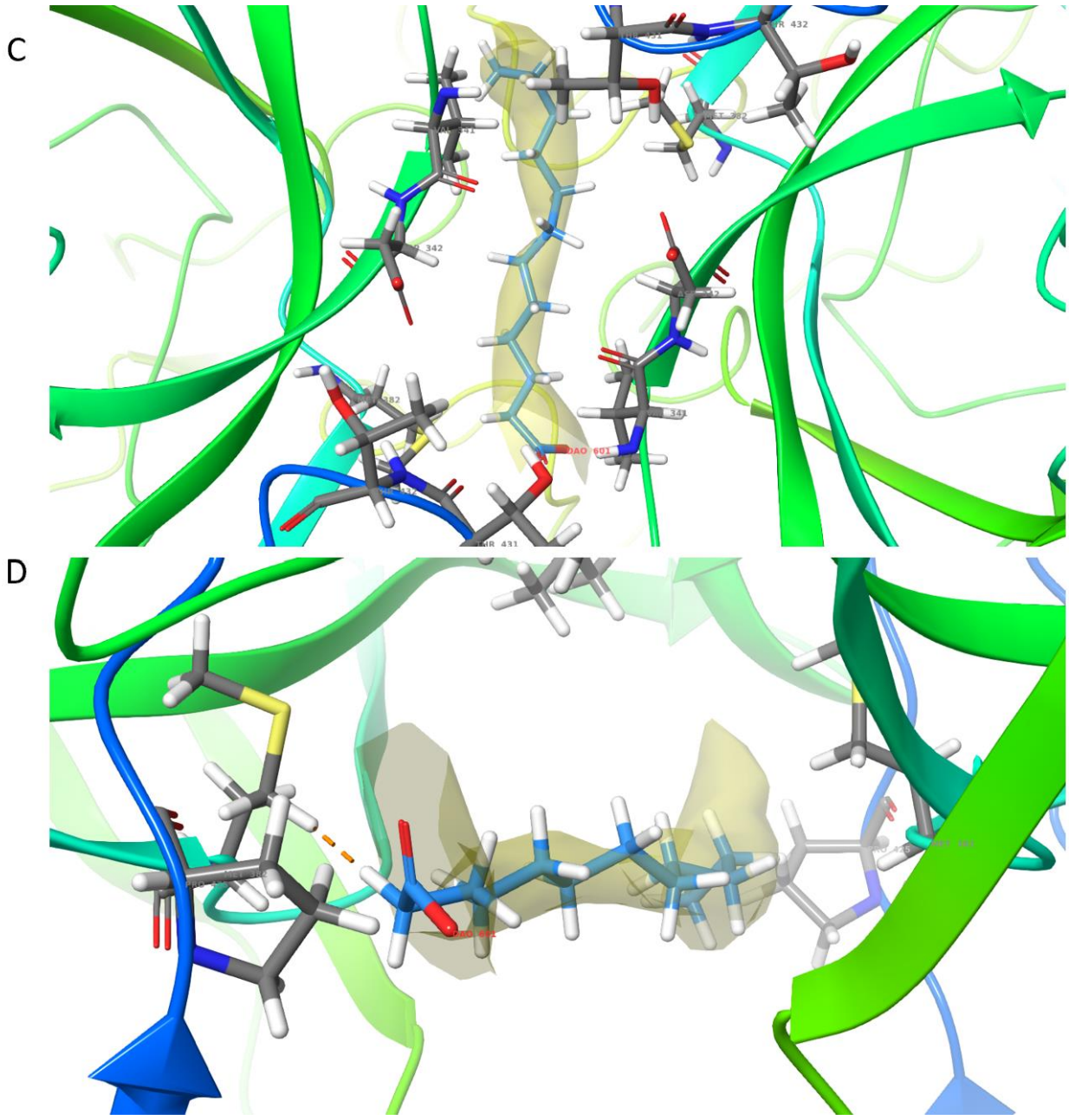


Figure 2.11. Extra density in the hydrophobic pocket of the two dimers showing in top view (A) and side view(B). Lauric acid is fitted into the extra density, top view (C), side view (D).

Figure 2.11 Continued



2.3.6 9-6-17 TV has lost the binding ability to the type B antigen

To compare the binding ability of the 9-6-17 TV and the wild-type TV to HBGA, we performed the saliva binding assay with the saliva with type B saliva sample (OH68) and type O saliva sample (OH7), representing B type and O type HBGA. It has been verified previously that the Tulane virus can strongly bind with type B saliva samples (Zhang et al., 2014). The type O saliva sample was used as a negative control since it has been reported to be the phenotype that TV doesn't interact with. Different types of saliva sample were first applied to the plate and the serial diluted virus samples were added subsequently to measure the binding ability. The absorbance signal should be corresponding to the amount of virus that binds with the saliva sample on the plate. According to Fig. 2.12A, the A450 of the wild-type TV with the type B saliva sample decreases as the concentration of the virus sample decreases. In contrast, the absorbance signal of the 9-6-17 TV doesn't change coordinately with the concentration of the virus sample applied. And the absorbance signal is similar to the level of the negative controls. We cannot definitely conclude from ELISA that the mutation sites identified are directly responsible for HBGA binding. Because mutations can affect the HBGA binding indirectly by affecting the overall conformation of the virus capsid protein.

The 9-6-17 TV has higher yield compared to the wild-type TV. The growth curve of the wild-type TV only reached 10^6 PFU/ml before the plateau, while the 9-6-17 TV is one order higher. It reached 10^7 PFU/ml (Fig. 2.12B) 24 hours post-infection. It is expectable that 9-6-17 TV is better adapted to the cell culture system and has higher amplification efficiency than the wild-type TV.

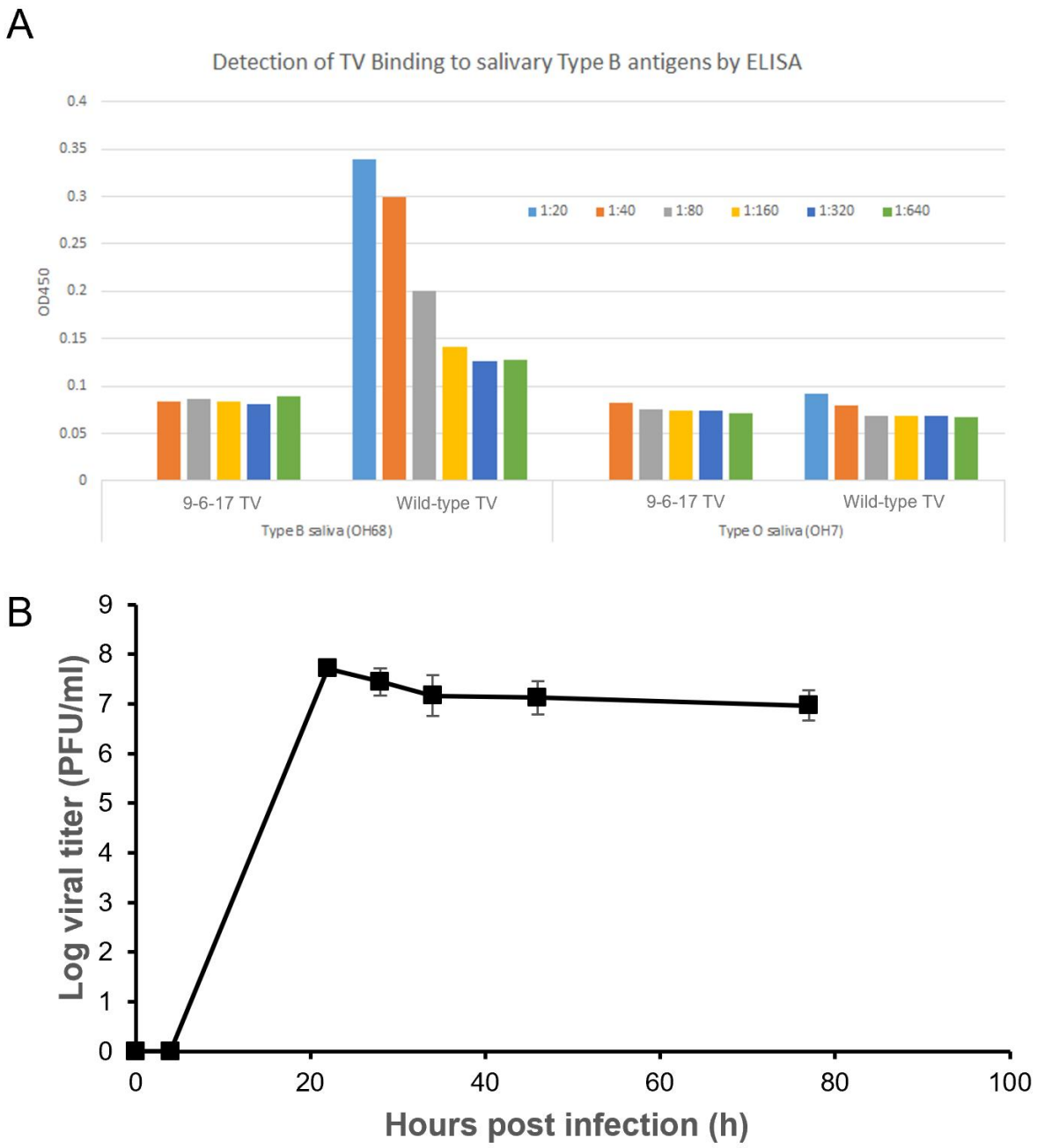


Figure 2.12. Saliva-based ELISA for measurement of HBGA binding of TV and the growth curve of 9-6-17 TV variant. One type B saliva sample (OH68) and one type O saliva sample as negative control were used for testing the change of TV binding to blood type B antigen. The titer of the 9-6-17 TV is 10^9 PFU/ml. The initial concentration in the ELISA starting at 1:40 was $2.5x10^7$ PFU/ml. The wild-type TV sample was from the PBS dialyzed Pool of wild-type TV peak fractions (F7 and F8) after CsCl density gradient centrifugation. The wild-type TV sample was prepared in 2016 and the virus titer was estimated at $3x10^8$ pfu/ml based on the plaque assay. The virus inoculum was generated in 2015. The initial concentration in the ELISA starting at 1:20 was $1.5x10^7$pfu/ml. (B) The growth curve of 9-6-17 TV in a time span of 2-77 hours after infection.

2.3.7 Identification of the 9-6-17 TV host cell receptor with mass spec

Given the ELISA result that the 9-6-17 TV strain doesn't bind to HBGA, we designed an experiment to identify the host cell receptor for this virus strain with mass spectrometry. The workflow is illustrated in Figure 2.13. We first immobilize the virus to the magnetic beads taken advantage that there are 13 surface exposed lysine per dimer of 9-6-17 TV which can react with the N-hydroxy-succinimide (NHS) functional group on the magnetic beads. Then incubate the virus bound magnetic beads with the cell in 4 °C to allow the virus to interact with its receptor. After 2 hours, the crosslinking reagent BS3 is applied to crosslink the virus with all surrounding proteins. Finally, the cells are lysed with detergent and magnetic beads are collected and washed multiple times before used for mass spec.

Initially, the Pierce™ NHS-Activated Magnetic Beads from Thermofisher Scientific was used in this experiment. However, more than one thousand proteins were identified by mass spec. We reason that it is because of nonspecific binding to the surface of the magnetic beads. Therefore, we also tried out the FG-NHS beads from Nacalai tesque (Toyoko, Japan) which is covered by a polymer layer to reduce nonspecific binding to the beads. The number of proteins identified with the FG-NHS beads is indeed less than those of the Pierce™ NHS-Activated Magnetic Beads. The negative control samples were generated with the same protocol but without virus addition. The database containing all plasma membrane proteins of the *Macaca mulatta* organism was used to screen out possible targets from the identified proteins. The sample information is listed in table 2.5. Only three membrane proteins are consistently observed in all parallel samples: Caveolin (Uniprot ID: F7C1V2), Cadherin 6 (Uniprot ID: F6YS32), and 40S ribosomal protein SA (Uniprot ID: F6U6B5). However, these three proteins were also observed in the negative control samples. It is worth to note that not all membrane proteins are properly annotated in the Uniprot database. We have tried to use membrane protein prediction tools to predict if an uncharacterized protein is a membrane protein. However, it is hard to examine each identified protein individually. We conclude that we failed to find the protein receptor for 9-6-17 TV for the negative control samples identified the same set of proteins that are identified in the samples with viruses. The number of samples tested, and the number of proteins identified in each sample is listed in Table 2.5.

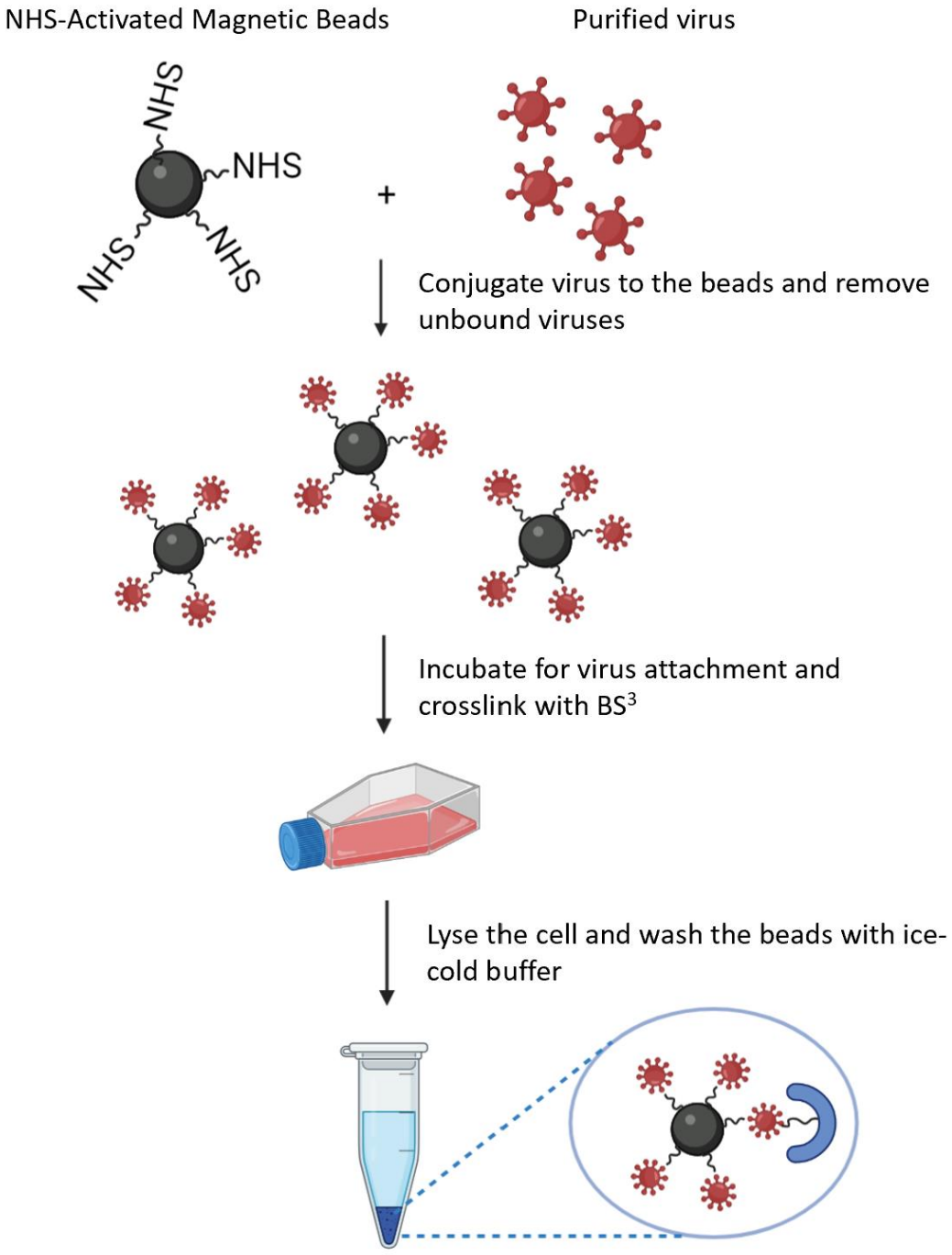


Figure 2.13. Workflow of the receptor identification of 9-6-17 TV. The NHS-Activated magnetic beads were used to conjugate the purified virus to it. The virus-conjugated magnetic beads were then incubated with the cell at 4 °C to allow the virus to attach to the cell receptor. The crosslinking reagent BS^3 was applied for 30 min before being quenched by high concentration Tris. Then the cells are lysed, and magnetic beads were washed multiple times. The resuspension of magnetic beads was subjected to trypsin digestion for mass spec.

Table 2.5 The number of identified proteins in all test samples.

Sample ID	Magnetic beads	The amount of 9-6-17 TV (mg)	Number of identified proteins
1	300ul 10mg/ml Pierce-NHS	1.02	1262
2	300ul 10mg/ml Pierce-NHS	1.02	2538
3	300ul 10mg/ml Pierce-NHS	1.02	906
4	100ul 10mg/ml Pierce-NHS	0	936
5	100ul 10mg/ml Pierce-NHS	0.34	747
6	50ul 20mg/ml FG-NHS	0	56
7	50ul 20mg/ml FG-NHS	0.34	476
8	50ul 20mg/ml FG-NHS	0	547

2.3.8 Design a full-length cDNA of the 9-6-17 TV (icTV-9-6-17)

We initially attempted to use the reverse genetic system to clone full-length cDNA sequence with the extracted viral RNA from 9-6-17 TV. With poly T primer, the first strand cDNA is obtained with reverse transcription. Two sets of primers were designed to amplify two segments F1 (1-3530) and F2 (3531-6743) (Figure 2.14 A). We engineered a T7 promoter at the upstream end of the virus genome and a poly$(A)_{30}$ tail at the downstream end of F2. However, the F2 segment couldn't be amplified efficiently that it always yields a faint band. The ligation of the two segments with the infusion method failed after multiple attempts. Then we synthesized two vectors containing the F1 and F2 sequences respectively (Figure 2.14 B). The restriction site in the viral genome AfIII was used for ligation of the two segments. However, the ligation of the synthesized F1 and F2 into a full-length vector still failed. At this point, we realized that the polyprotein of TV might be toxic to bacteria. Because of that, even a trace amount of polyprotein expression would lead to the death of the bacteria used for cloning. To tackle the toxicity problem, we adopted the

antisense strategy to inhibit the translation of the toxic protein by inserting a complementary sequence of the toxic region to the vector backbone. We inserted a 186bp sequence complementary to a segment in the polyprotein region into the vector. The full-length cDNA clone of the 9-6-17 TV (icTV-9-6-17) was established successfully and further validated by sequencing and single enzyme digestion.

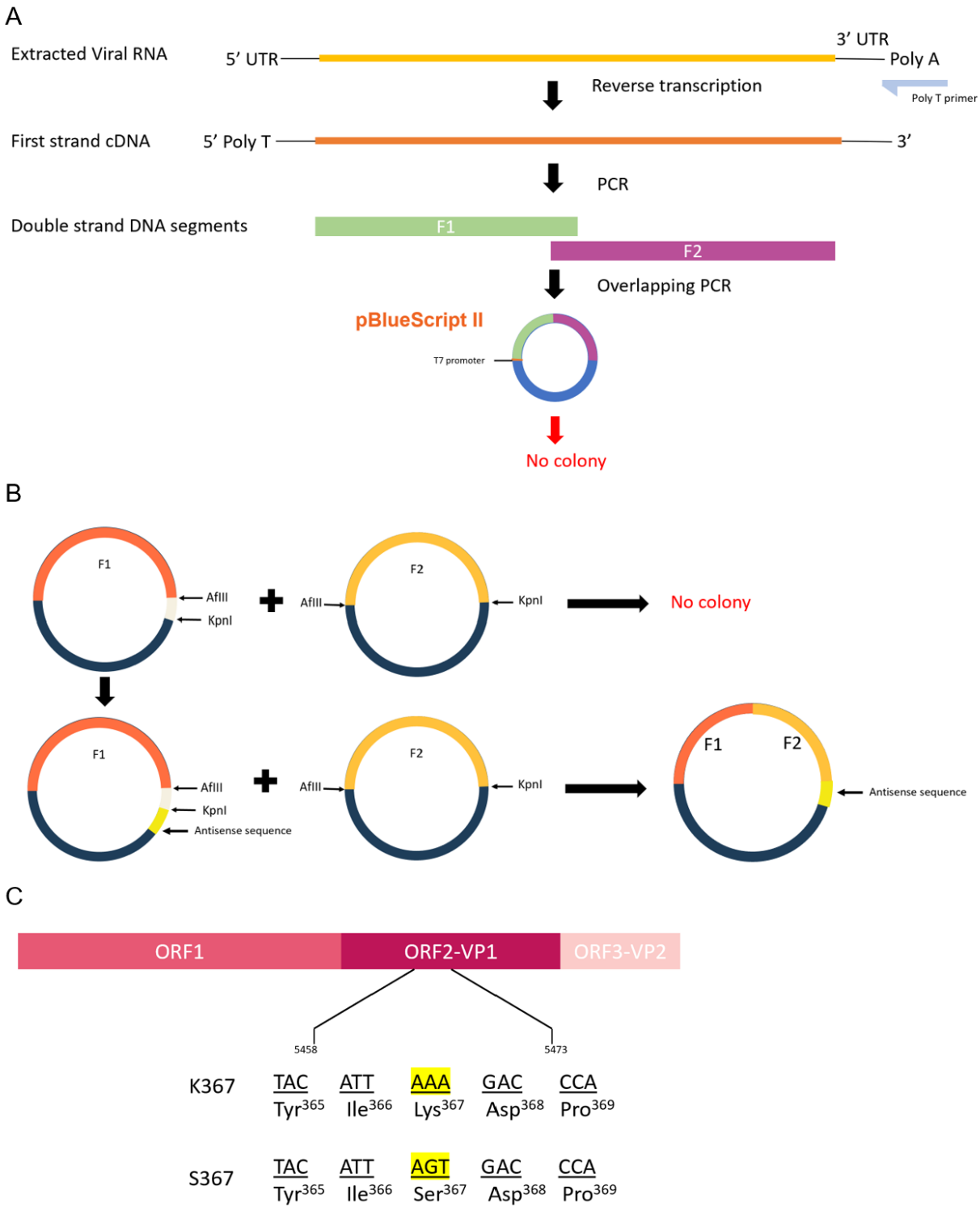


Figure 2.14. Assembly of a full-length 9-6-17 TV infectious cDNA Clone. (A) In vitro assembly of an infectious cDNA clone of TV with reverse transcription. (B) In vitro assembly of an infectious cDNA clone with antisense strategy. (C) Construction of K367 and S367 virus. A two-nucleotide substitution was introduced to produce the spike K367S substitution in the infectious cDNA clone of TV.

The in vitro transcription generated mRNA of full-length 9-6-17 TV, followed by in vivo transfection to transfect the mRNA into LLC-MK2 cell. 48 hours post-transfection, cytotoxic effect was obvious. The P0 virus was collected after 72 hours and used for amplification. The amplified virus was able to propagate in the cell culture. In the meantime, site-directed mutagenesis was performed to mutate the lysine at position 367 into serine. Two isogenic viruses K367 and S367 were generated and purified with density gradient centrifugation. They were used to assess with hemagglutination assay to justify if restoring K367 to Serine would be able to restore the binding ability of 9-6-17 TV to HBGA. However, both two viruses showed the negative results in the hemagglutination assay with B-type red blood cells which indicates that they are not binding to B-type HBGA (Figure 2.15). Sequencing of the extracted viral RNA confirmed the S367 virus still has a serine at position 367 after two passages of cell culture. It suggests that reverse-engineering the Lys at position 367 to serine is not enough to restore its HBGA binding ability. The HBGA binding site of human norovirus is at the interface of the two subunits which involves multiple amino acids to stabilize the binding. Given their close relationship, TV is likely to adopt a similar strategy that binding to HBGA requires the coordination of multiple residues in the P domain.

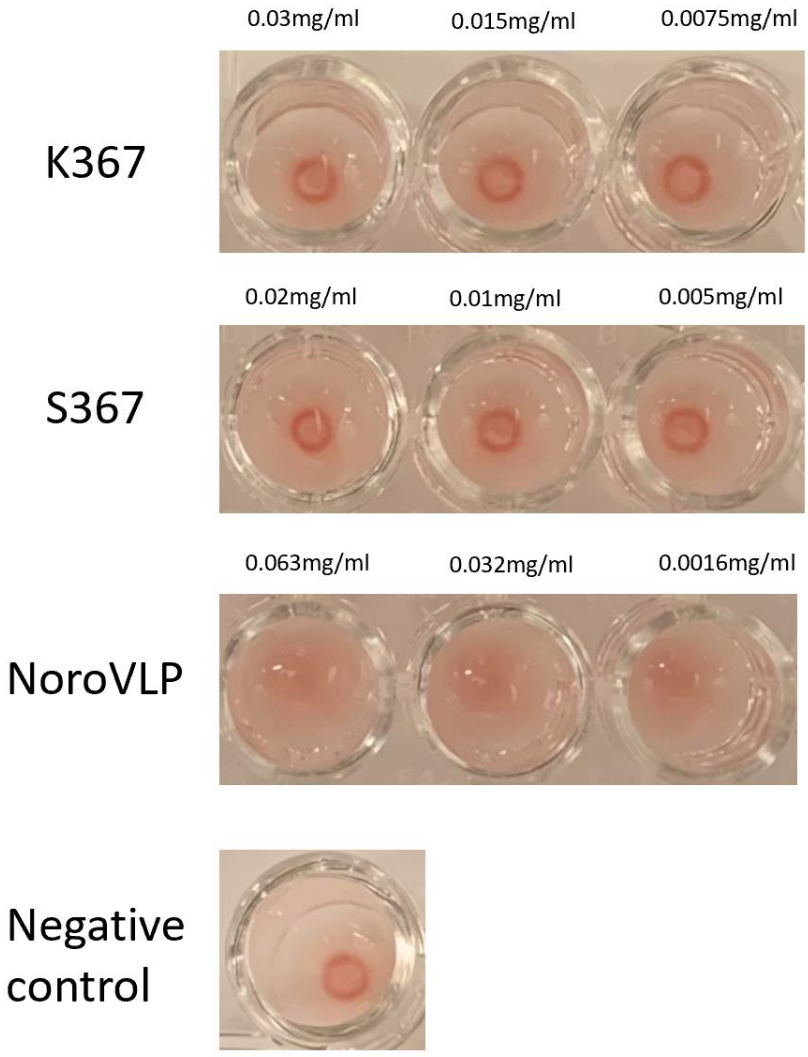


Figure 2.15. The hemagglutination assay of K367 and S367 isogenic viruses. Purified viruses were serially diluted as the concentration indicated above. The negative control was performed with PBS, while the positive control was with norovirus VLP.

2.3.9 Insect cell expression of Tulane virus virus-like particle (VLP)

To assess the binding of HBGA and TV capsid protein, we also tried a different approach by expressing Tulane virus capsid proteins in insect cells to form virus-like particles. A construct with VP1 and VP2 sequence was cloned from the extracted viral RNA and confirmed by sequencing (Fig. 2.16 A). After the bacmid with VP1 and VP2 sequence was generated, we also confirmed its sequence by sequencing. The site-directed mutagenesis of K367S was performed on the pFastBac vector and sequence confirmed before transfection into DH10Bac.

The bacmids with or without K367S mutation were used for transfection into insect cell Sf9 to generate baculovirus. After transfection, enlarged cells are observed and the number of viable cells is reduced suggesting successful baculovirus production. The P0 baculoviruses were amplified in small culture two passages. 1L culture of insect cells were harvested 72h post-infection of P2 virus. The SDS PAGE result (Fig. 2.16 B) of the whole cell lysate of the culture with P2 baculovirus is very similar to that of the uninfected cell lysate. There is a band at 60kD position in the P2 cell lysate but is not dominant. Further purification with density gradient didn't show a 60kD band in the expected density fractions.

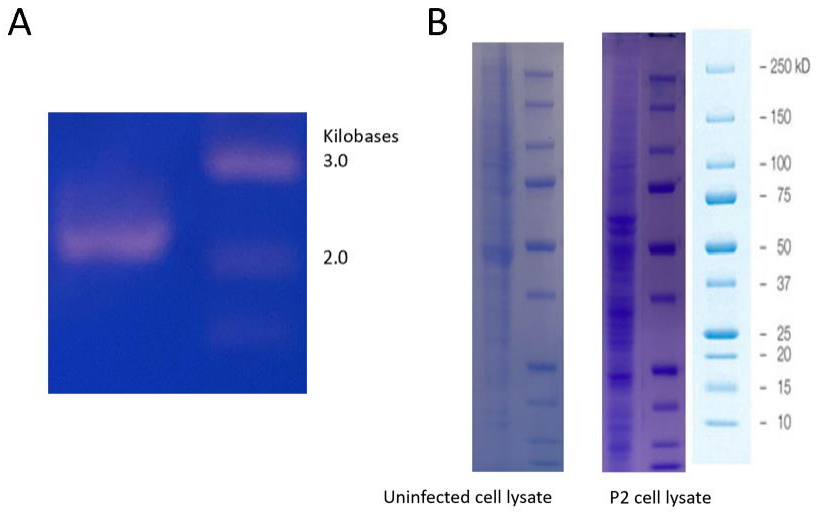


Figure 2.16. Insect cell expression of TV VLP. (A) After obtaining baculovirus, PCR with a set of primers to amplify the VP1 and VP2 sequence yielded the expected band at 2kb. (B) The SDS PAGE of cell lysate with and without baculovirus infection showed similar bands but no band at 60kD for VP1 expression.

To rule out other possibilities, we have asked the company Genscript (Piscataway, NJ) to perform the same insect cell expression experiment. A His-tag was engineered to the N-terminal of VP1. The result that they have obtained is identical to our results that no dominant bind at 60kD in the whole cell lysate (Fig. 2.17A). They have also confirmed the expression of VP1 with western-blot (Fig. 2.17B). Based on the SDS PAGE result, the 60kD band intensity in the supernatant of cell lysate is weaker than that in the whole cell lysate, it indicates that the VP1 expressed in insect cell could have solubility issue. The solubility issue can also be the main reason responsible for the low yield of VP1. The accumulated insoluble VP1 proteins would become a burden to the host cell and could potentially influence the cellular metabolism. The affected insect cell might also initiate apoptosis. With reduced viability, the amount of VP1s expressed in the end would not be much.

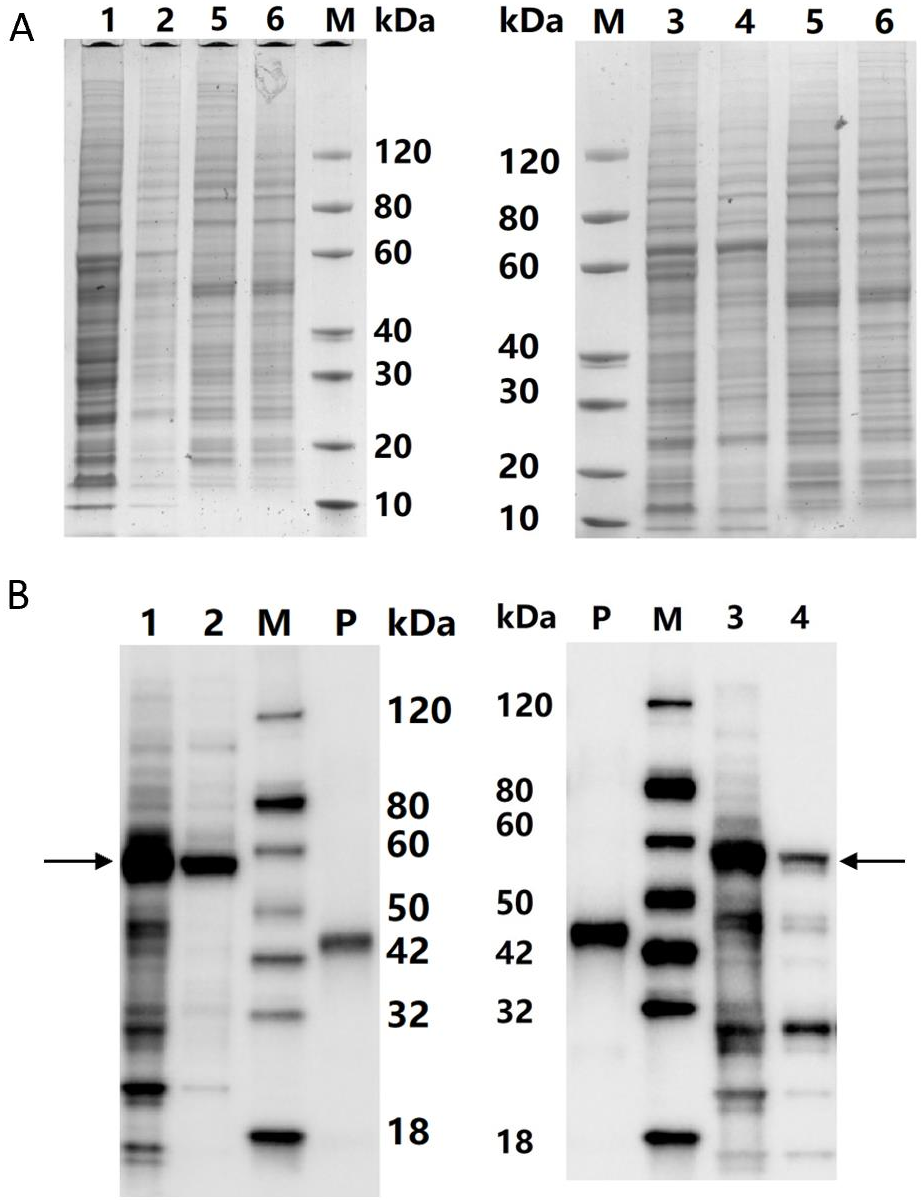


Figure 2.17. Insect cell expression of TV VLP. (A) The SDS PAGE result of P1 and P2 generation expression. (B) The western-blot result of P1 and P2 generation expression of VP1 and VP2. The His-tag was engineered at the 5' end of the VP1. Lane 1: Whole cell lysate of P1 culture. Lane 2: Supernatant after removing sediment of P1 culture. Lane 3: Whole cell lysate of P2 culture. Lane 4: Supernatant after removing sediment of P2 culture. Lane 5: Whole cell lysate of the negative control sample. Lane 6: Supernatant after removing sediment of the negative control sample. Lane P: Positive control.

2.3.10 Assessing HBGA binding affinity of 11-25-12 TV strain

The 11-25-12 TV strain was obtained from our collaborator. Before infection, 140ul infected cell culture was used to extract the viral RNA from this stock. We designed three sets of primers to amplify three segments of DNA covering the entire sequence of VP1 and VP2 (Fig. 2.18). Although only one mutation existed in VP1 of the 11-25-12 strain, the mutations in VP2 are largely the same as the ones in the 9-6-17 TV. It is reasonable to hypothesize that the 9-6-17 strain is probably further evolved on the basis of the 11-25-12 strain. After inoculation, the cytotoxic effect appeared after 48 hours. After being amplified within two passages and concentrated, the viral titer was measured by plaque assay to be around 10^6 to 10^7 PFU/ml. Two-fold series dilution was performed for hemagglutination assay. However, the result is still negative. Sequencing of the extracted viral RNA after amplification showed the four mutations in VP1 remained the same.

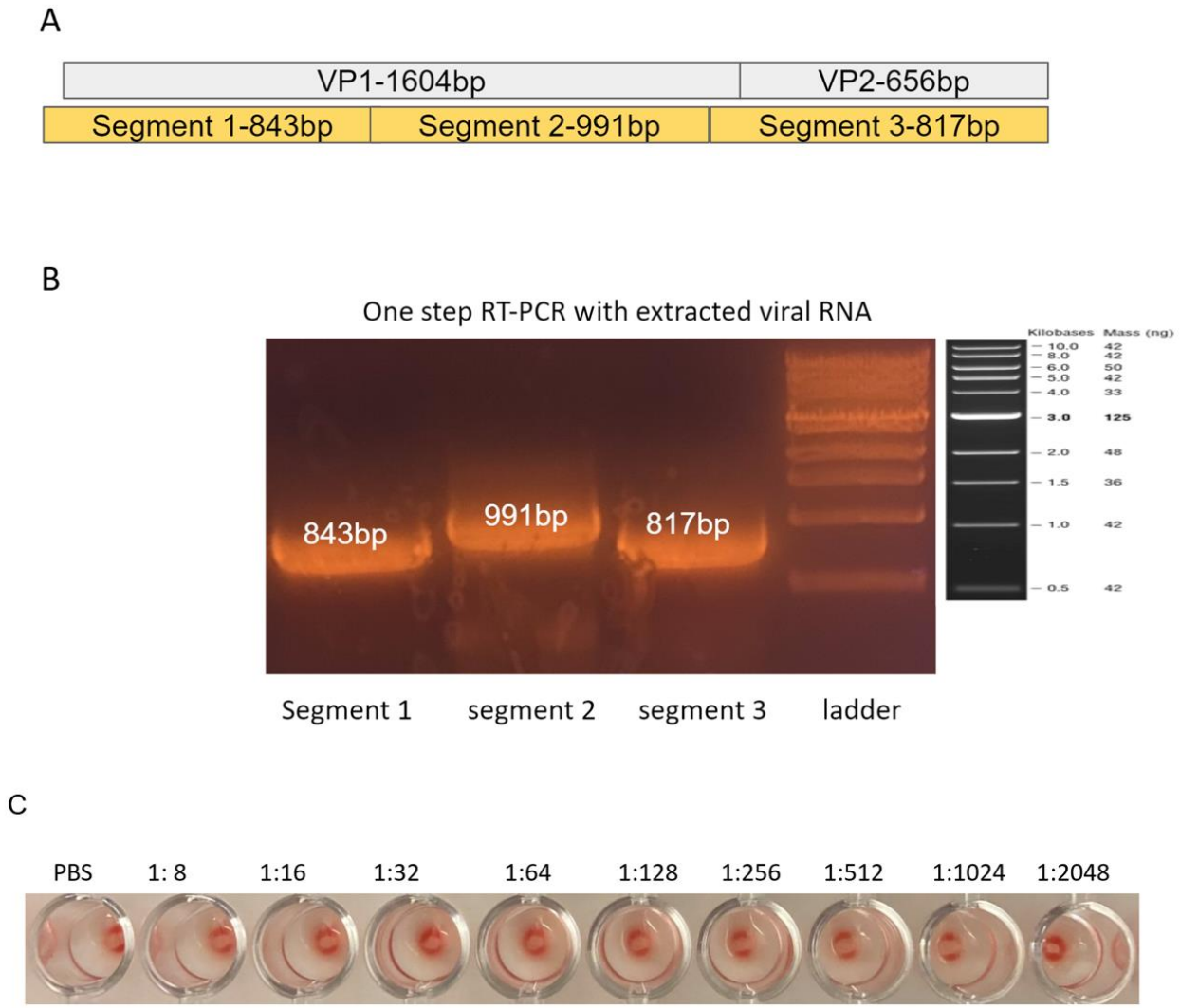


Figure 2.18. Sequencing of 11-25-12 TV stock. (A) Schematic view of the sequencing primer design to cover the entire sequence of VP1 and VP2. (B) Amplified segments from extracted viral RNA with One-step RT-PCR. (C) Hemagglutination assay of purified 11-25-12 strain. Each well contains 50ul 0.1% B-RBC. 1ul of PBS or serial dilution of the viruses was added and incubated for 1 hour at room temperature.

2.4 Discussion

Virus evolution for interspecies transmission has long been in the center of the spotlight. In contrast, virus evolution in the cell culture system has been largely underestimated. Instead of random mutations at random positions, this study shows that adaptive, evolutionally favored mutations related to the receptor interaction and the virus infectivity could be generated and preserved in the cultivated Tulane virus. Previous works have demonstrated that TV could recognize all the B-types and the A type-3 HBGAs (Zhang et al., 2015). In this study, we presented a new TV strain (the 9-6-17 TV strain) that has lost the binding ability to its original receptor the B-type HBGA. Sequence analysis and structure determination were performed to analyze the mutation sites in the major capsid protein VP1. Multisequence alignment of VP1 proteins in ten other *Recovirus* strains showed the prevalence of each mutation in these viruses. Three of the eight mutations in VP1 (N3S, N284H, F334V) are more commonly existed in the Recoviruses compared with the wild-type TV sequence. Suggesting that these mutations are probably also favored in the evolutionary process in vivo. Based on the high-resolution electron density map, the mutated residue H284 is positioned to form a π- π interaction with the nearby residue W287. Four of the mutations (A335E, S367K, I451M, R452C) in 9-6-17 TV were not observed in the other *Recovirus* strains studied. The A335E mutation is unique in that all other *Recovirus* strains have the negative residue aspartic acid at this position suggesting the necessity of having a negatively charged residue here. Interestingly, based on the atomic structure of VP1, the side chain of E335 doesn't interact with any neighboring residues. It is located at the top of the P domain near the dimer interface but is pointing outward. A highly selective negative charged residue at this position is very likely to be involved in the receptor binding of TV. The R452C mutation at the dimer interface resulted in the formation of a disulfide bond Cys^{452}-Cys^{452} that links the P domain of two subunits in a dimer. The existence of this disulfide bond is supported by the SDS PAGE result of the virus without reducing reagent in the loading buffer has revealed a dimer band at around 120kD. Only one mutation (A343T) is considered to be a conservative substitution that might not affect the VP1 stability or function.

Three EM datasets were collected for 9-6-17 TV and yielded three high-resolution structures. The first dataset was imaged with opposite contrast due to the presence of the density gradient iodixanol, which could serve as a representative case demonstrating the feasibility of high-resolution reconstruction with the opposite contrast. The DTT treated dataset reconstructed a

map with a large blob of extra density near the residues H284 and P285. The electron density of the disulfide bond Cys^{452}-Cys^{452} is well preserved and is found in both A/B and C/C dimers. This disulfide bond formed at the dimer interface could potentially stabilize the dimer structure and might also result in receptor change. An elongated extra density is observed in a hydrophobic pocket in the VP1 dimer. It is located between the surface of the P domain and the disulfide bond that links the two subunits. It seems to have C2 symmetry. Lauric acid has the same length of the extra density but doesn't match the symmetry.

A study of human norovirus (Mallagaray et al. 2019) demonstrated that the binding of the glycan to human norovirus P dimers follows a simple single-site binding model in which the post-translational modification of a single amino acid N373 can affect HBGA recognition, as demonstrated by NMR. The sequence analysis shows that the N373 residue is corresponding to the S367 position which has mutated to lysine in the 9-6-17 TV sequence. It is tempting to hypothesize that the mutation from serine to lysine at the amino acid position 367 is the key residue for HBGA recognition in TV.

However, we have found that the S367 is not the only important residue required for HBGA binding. Two isogenic viruses were engineered to investigate whether reverse-engineering the lysine back to serine in the 9-6-17 TV would be able to restore its HBGA binding ability. The results of the hemagglutination assay suggest that neither the K367 nor S367 viruses could bind to B-type red blood cells. The HBGA binding site of human norovirus is at the interface of the two subunits which involves multiple amino acids to stabilize the binding. Given their close relationship, TV is likely to adopt a similar strategy that binding to HBGA requires the coordination of multiple residues in the P domain.

The finding that the 9-6-17 TV strain that has completely lost the binding ability to HBGA also helped our understanding of the function of HBGA in the virus infection process. HBGA has been recognized as the putative receptor of both human norovirus and TV for a long time. However, except for the binding affinity assays, there is no fully established theory of how it mediates the virus cell entry. Some norovirus strains have been found not to bind to any types of HBGAs (Almand et al. 2017). These results suggested the presence of potential alternative receptors for TV. Given the higher yield of the 9-6-17 TV, the infection via the secondary receptor is probably more efficient in the cell culture environment. The characterization of our adaptive TV variant provided new evidence for this theory.

As of now, no proteinaceous receptors have been identified for TV. Therefore, we have designed a magnetic bead-based affinity capture experiment to identify the receptor of 9-6-17 TV with mass spectrometry. A large number of proteins were identified. However, the same sets of proteins were also found in the negative control sample. It indicates that the majority of the proteins identified were captured either by binding to the surface of the magnetic beads or randomly crosslinked to the polymer chains of the magnetic beads.

In summary, we demonstrated the important aspects of structural and sequence analysis of the new 9-6-17 TV strain as well as laying the groundwork for future studies of the TV and receptor binding. Together these data suggest the important contribution of the amino acid residues mutated in the 9-6-17 TV strain to HBGA recognition and virus adaptation. Considering this fast evolution of the Tulane virus in the cell culture system that it has totally abandoned the original receptor, it alerts us to the possibility of receptor change for human norovirus. The constant outbreak of norovirus over the years is indicating numerous passages when the virus spread among people. Some of the variants of norovirus have already been found to be able to escape from the antibody of the current type by adopting a single mutation (Lindesmith et al., 2018). It reminds us that the receptor of the norovirus variant may already change. Therefore, intensive studies of the receptor of norovirus new variants still wait to be conducted.

2.5 Acknowledgement

This chapter is in collaboration with Dr. Xi Jiang in Cincinnati Children's hospital. The ELISA experiment was performed by Pengwei Huang.

CHAPTER 3. SUB-3 Å APOFERRITIN STRUCTURE DETERMINED WITH FULL RANGE OF PHASE SHIFTS USING A SINGLE POSITION OF VOLTA PHASE PLATE

This chapter is from a publication (Li et al., 2019) in collaboration with other scientists across different institutions. Their contributions are outlined in the journal website.

3.1 Abstract

Volta Phase Plate (VPP) has become an invaluable tool for cryo-EM structural determination of small protein complexes by increasing image contrast. Currently, the standard protocol of VPP usage periodically changes the VPP position to a fresh spot during data collection. Such a protocol was to target the phase shifts to a relatively narrow range (around 90°) based on the observations of increased phase shifts and image blur associated with more images taken with a single VPP position. Here, we report a 2.87 Å resolution structure of apoferritin reconstructed from a dataset collected using only a single position of VPP. The reconstruction resolution and map density features are nearly identical to the reconstruction from the control dataset collected with periodic change of VPP positions. Further experiments have verified that similar results, including a 2.5 Å resolution structure, could be obtained with a full range of phase shifts, different spots of variable phase shift increasing rates, and at different ages of the VPP post-installation. Furthermore, we have found that the phase shifts at low resolutions, probably related to the finite size of the Volta spots, could not be correctly modeled by current CTF model using a constant phase shift at all frequencies. In dataset III, severe beam tilt issue was identified but could be computationally corrected with iterative refinements. The observations in this study may provide new insights into further improvement of both the efficiency and robustness of VPP, and to help turn VPP into a plug-and-play device for high-resolution cryo-EM.

3.2 Introduction

Transmission electron microscope (TEM) images of biological samples are known to have low contrast due to the weak scattering of the electron beam. Traditionally, the phase contrast in TEM is realized by collecting under-focused images. A phase plate that can increase the image contrast has been a long-sought device in the cryo-EM field. While the physical principle of phase

plates has been well-established since 1942 (Zernike, 1942) and its application to light microscopy succeeded quickly, the implementation for TEM has proven to be very difficult. Several different approaches, such as the Zernike phase plate with a small hole in a thin carbon film (Danev and Nagayama, 2001) and the Boersch phase plate with an electrostatic lens (Schultheiß et al., 2006), were explored to generate the 90° phase shift between the scattered and unscattered beam. However, these approaches were hampered by contaminations, short-lifetime spans, low resolutions, *etc*.

The Volta Phase Plate (VPP) (Danev et al., 2014) is the first phase plate design that has demonstrated both high-resolution and robustness, and it has quickly become a popular commercial product since its invention in 2014. It is also made of a thin layer of amorphous carbon but without a hole in the center. VPP utilizes the phase shifts to the central, unscattered beam based on the Volta potential of the spot on carbon film where the intense central beam passes. The mechanism of the Volta potential self-induced by the central beam has not been fully understood, although it was hypothesized to be caused by the beam-induced desorption of the water molecules from the carbon surface (Danev et al., 2014; Hettler et al., 2018). Despite the quick success of VPP as demonstrated by several near-atomic resolution structures solved using VPP data (Danev et al., 2017; Fan et al., 2017; Liang et al., 2017), it is also recognized that VPP is currently still challenging to set up and far from a plug-and-play device. Efforts are still being made to improve its performance and user experience.

The current VPP workflow changes VPP position periodically during image acquisition and limits the phase shifts to a relatively narrow range around 90°. Such a procedure was based on the observations of increased phase shifts associated with accumulating dose (Danev et al., 2014) and increased image blur associated with accumulating dose and phase shifts (Danev et al., 2017). Furthermore, the power-spectra based CTF fitting methods currently used in the cryo-EM field cannot distinguish a phase shift ϕ from ϕ+180°. Large phase shifts are thus considered undesirable for image collection and the procedure employing a periodic change of VPP position has been widely adopted. In this study, we investigated the correlations of phase shifts and image/3D reconstruction quality, with a varying number of movies imaged using a single VPP spot, to test if it is feasible to obtain high-resolution 3D reconstructions using a single VPP position. Surprisingly, these tests have consistently resulted in 3D reconstructions at near-atomic resolutions even with phase shifts larger than 180° and more than 400 movies taken using a single VPP position. This

study may lead to a better understanding of VPP phase shifts and suggest the potential of using a single VPP position to image a large number of movies for more efficient data collection with VPP.

3.3 Material and Methods

3.3.1 Sample preparation and grid screening

The human ferritin light chain (FLT) was expressed in *Escherichia coli*, purified, and assembled into ferritin homopolymers (24-mers) as described previously (Baraibar et al., 2009). Protein concentration was determined with Bradford assay with bovine serum albumin (BSA) as a standard. The recombinant ferritin sample was incubated with 1% thioglycollic acid, pH 5.5, and 2,2'-bipyridine to remove iron. It was further dissolved in 100 mM Hepes buffer (pH 7.4) that was treated with Chelex ion exchange resin (Bio-Rad) to remove transition metals often found in aqueous buffers as contaminants. Subsequently, a 3 μl aliquot sample of 0.05 mg/ml concentration was applied to a holey UltraAufoil grid (R1.2/1.3, 300 mesh) (for dataset I, II and III) that was precoated with graphene oxide and 0.01% chitosan (Sigma). For dataset IV, Quantifoil holey carbon grid (R0.6/1, 400 mesh) coated with graphene oxide was used. Cryo-EM grids were prepared using a CP3 plunger (Gatan, CA, USA) with 80% humidity, ashless filter paper and blotting time 7-9 s. The cryo-grids were screened using a CM200 microscope or Talos F200 microscope.

3.3.2 Cryo-EM data collection

The grids prepared with the optimized freezing condition were then imaged using a Titan Krios electron microscope (Thermo Fisher Scientific) equipped with a VPP. The on-plane alignment of VPP was performed after correcting objective lens astigmatism and coma-free alignment. Images were recorded using a Gatan K2 Summit detector in super-resolution mode. The images were collected using *Leginon* software (Suloway et al., 2005) with a constant intended underfocus of 500 nm. To minimize the changing of objective lens current, eucentric Z-height was corrected once at SQ mode and twice at HL mode, then further corrected at each of the Focus node in *Leginon*. The magnifications were calibrated using the 2.13 Å resolution diffractions of graphene oxide substrate as the internal standard. Detailed imaging conditions for the four datasets are listed in Table 3.1. For dataset I and dataset II, every time after the VPP had advanced to a new

position, the VPP was charged for 120 s to build up an initial phase shift of ~20 degrees. For dataset III and IV, precharging was not performed. All four datasets are available from the EMPIAR database (EMPIAR-10263).

Dataset I and II: The datasets I and II were collected using the same imaging condition and the same sample grid except that the VPP position was set to change every 30 movies in dataset I, while the VPP position remained the same in dataset II. Both datasets were imaged about six months after the installation of the VPP.

Dataset III and IV: These datasets were collected in two separate sessions, after about six and eight months respectively after datasets I and II using the same Titan Krios electron microscope but with the K2 camera now mounted on a GIF Quantum LS Imaging Filter with 20 eV slit width and the nominal mag at 130k. Dataset III was collected with the VPP position changed every 90 to 144 movies. All 461 movies in dataset IV were collected with a single VPP position.

Table 3.1 Refinement and Model Statistics

	Dataset I	Dataset II	Dataset III	Dataset IV
Data Collection				
Voltage (kV)	300	300	300	300
Dose (e/Å^2)	35	35	30	30
Nominal Magnification	22,500	22,500	130,000	130,000
Energy filter	-	-	+	+
VPP age post-installation (months)	6	6	12	14
VPP Positions	18	1	6	1
Number of Movies	629	164	735	461
Particles picked	642,775	159,409	500,613	279,190
Particles after 2D classification	155,639	96,631	375,811	193,853
Particles after 3D classification	72,521	96,631	304,638	73,314
Pixel Size (Å)	0.658	0.658	0.535	0.535
Number of frames	50	50	40	40

Table 3.1 Continued

Intended Defocus (μm)	0.5	0.5	0.5	0.5
Beam size (μm)	0.966	0.966	0.931	0.801
Electric charge per movie (nC)	0.41	0.41	0.33	0.24
Refinement				
Resolution (Å)	2.94	2.85	2.51	2.93
Map CC	0.82	0.79	0.80	0.79
All-atom clashscore	3.97	3.61	3.47	2.31
Rotamer outliers (%)	2.05	2.05	0	0.72
Ramachandran plot				
Favored (%)	95.71	95.71	100	99.35
Outliers (%)	0	0	0	0

3.3.3 Image processing

Dataset I: The 629 raw movies were aligned and dose-weighted with *Motioncor2/1.0.5* (Zheng et al., 2017). Subsequently, *Gctf-v1.18* (Zhang, 2016) and *CTFFIND4* (Rohou and Grigorieff, 2015) were used to determine the CTF parameters using the movie averages without dose weighting. Figure 3.8A shows the *CTFFIND4* estimated phase shift and maximum resolution of each micrographs. 642,775 particles were selected with *RELION/2.1* autopicking (Scheres, 2012). After the first round of 2D classification, 328,691 particles were retained. Another round of 2D classification was applied to remove more junk particles with 155,639 good particles selected. 72,521 particles were retained after 3D classification and were used in the final reconstruction. The number of particles in different phase shift ranges before and after 2D/3D classification are shown in Figure 3.9 A. The *RELION/2.1* result at 3 Å was further refined to 2.9 Å using *JSPR* (Guo and Jiang, 2013) that includes refinement of defocus, astigmatism, beam tilt, magnification, and magnification distortion (Liu et al., 2016; Yu et al., 2016).

A subset of 164 micrographs with phase shift in the range of 72 to 108 degree were selected as the "around 90 subset" from the dataset I. After 2D classification, 102,353 of 131,689 particles were retained. 71,732 particles in the class with correct ferritin structure were selected after 3D classification and were used in the final reconstruction. The *RELION/2.1* refined result at 3.2 Å resolution was further refined to 2.94 Å using *JSPR*.

A random subset of 164 micrographs were also selected from dataset I. The total number of particles after autopicking was 136,697. After 2D classification, 106,718 particles were retained. 89,376 particles in the 3D class with correct structure were refined to 3.33 Å with *RELION/2.1* and further refined to 2.9 Å with *JSPR*.

Dataset II: The image processing procedures of the dataset II are the same as that for the dataset I. The difference is that dataset II has 164 movies from which 159,409 particles were auto-picked using *RELION/2.1*. After two rounds of 2D classification in *RELION/2.1*, 96,631 particles were retained. After 3D classification, all the 3D classes contain the correct structure. Thus, all of the 96,631 particles were further refined to 3.27 Å resolution with *RELION/2.1*. The same set of particles were further refined to 2.87 Å with *JSPR*. In the second round of image processing with bin-2 particles, the total number of 142,079 particles were subjected to two rounds of 2D classification which resulted in 110,347 particles in retained classes. After 3D classification, the

two classes with the correct structure were combined (90,813 particles) and refined to 3.3 Å with *Relion/2.1* and further refined to 2.88 Å with *JSPR* (Fig. 3.10).

Dataset III: After aligning the movies with Motioncor2/1.0.5, the dose-weighted movie averages were imported to *cisTEM* (Grant eFt al., 2018) to perform CTF fitting using *CTFFIND4*. 500,613 particles were automatically picked. After two rounds of 2D classification in cisTEM, 375,811 particles were retained and exported to *RELION/2.1* for 3D classification. Two classes of 304,638 particles with correct structure were selected and refined to ~6 Å with *RELION/2.1*. The Fourier shell correlation (FSC) curve showed severe CTF artifacts that was later found to be caused by a large degree of beam tilt. After refining beam tilt and other parameters using *JSPR*, the final resolution was significantly improved to 2.51 Å.

Dataset IV: The motion correction and CTF estimation were performed with *Motioncor2* and *CTFFIND4* respectively. To correct the 180° errors of phase shift assignments for the second half of the 461 movies taken using a single VPP spot when the phase shifts have increased beyond 180 degree, we manually added 180° to the *CTFFIND4* determined values of these micrographs to unwrap the phase shifts. 279,190 particles were selected and then subjected to 2D classification using *cryoSPARC/v2* (Punjani et al., 2017). A total of 79,226 particles was then exported to *RELION/3.0* for 3D classification and auto-refinement that resulted in 4.25 Å resolution. After *JSPR* refinement, the reconstruction was improved to 3.12 Å resolution. We then reprocessed this dataset by high-pass filtering all the particles at 90 Å using EMAN2 (Tang et al., 2007) *e2proc2d.py* and the *filter.highpass.gauss* processor. The 2D and 3D classification steps were repeated for the high-pass filtered particles in *RELION/3.0*. 193,853 particles were retained after 2D classification of high-pass filtered particles. Using 73,314 particles retained after 3D classification, we got 3.94 Å resolution with *RELION* 3D autorefine which was then improved to 2.93 Å resolution with *JSPR* refinement.

3.3.4 Model refinement

The initial atomic model of FTL PDB: 2ffx (Wang et al., 2006) was fitted into the cryo-EM maps using *Chimera* (Pettersen et al., 2004) and refined using *Phenix* real space refinement (Afonine et al., 2012). The *phenix.mtriage* program was used to calculate the FSC of the atomic model and the density maps. All maps and models were displayed with *Chimera*.

3.4 Results

3.4.1 A single position of Volta phase plate is able to acquire enough data for high-resolution reconstruction

Here we used the apoferritin particles as the test sample for all the datasets presented in this paper (Table 3.1). Dataset I of 629 movies was collected with VPP advanced to a new spot after every 30 movies, while dataset II of 164 movies only used a single spot on VPP. Since the two datasets were imaged with the same sample grid in the same image condition, the particle distribution and particle density are about the same for the two datasets. Using the same single particle reconstruction procedure, the dataset II, and the dataset I both produced a 3D reconstruction at ~2.9 Å resolution.

The trajectory of *Gctf*-measured phase shifts as a function of time in dataset I is shown in Figure 3.1A. Since the VPP was programed to advance to a new position after collecting 30 movies, a large number of cycles of phase shift increments were observed. In contrast, the phase shifts of dataset II show a very different trajectory (Fig. 3.1 B). The phase shifts kept increasing with two regions of different slopes. The trajectory began with a fast increase region until the phase shifts have reached ~110 degrees and then the phase shifts increased at a much reduced rate. Such a biphasic increase is consistent with previous observations (Danev et al., 2014). The phase shift histograms clearly showed the dramatically different distributions of phase shifts of these two datasets. While the phase shifts are mostly less than 110 degrees for the dataset I (Fig. 3.1 C), the phase shifts of dataset II are dominated by a peak around 130 degrees (Fig. 3.1 D), a range that is not recommended by the standard VPP protocol. The histograms of defocus distributions of both datasets are shown in Figure 3.2 A, B. It is obvious that the two distributions are similar with both having a peak at ~0.6-0.7 μm defocus although the histogram of dataset II is skewed to larger defocuses (Fig. 3.1 B) compared to that of dataset I (Fig. 3.1A). The *CTFFIND4*-estimated figure of merit distribution of the two datasets are presented in Fig. 3.2 C and Fig. 3.2 D. Dataset II has a comparable mean value and the spread of figure of merit values as those of dataset I.

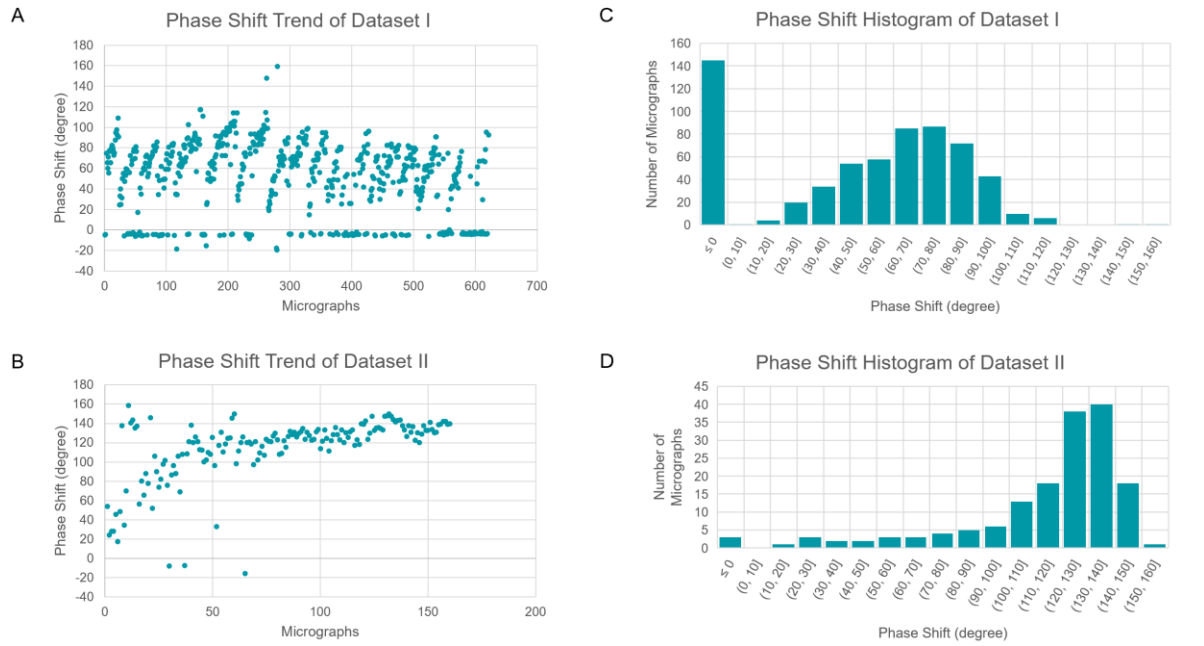


Figure 3.1. Statistics of phase shifts. (A and B) The time-course of GCTF-determined phase shifts for dataset I (A) and dataset II (B). (C and D) The phase shift distribution of the micrographs of dataset I (C) and dataset II (D).

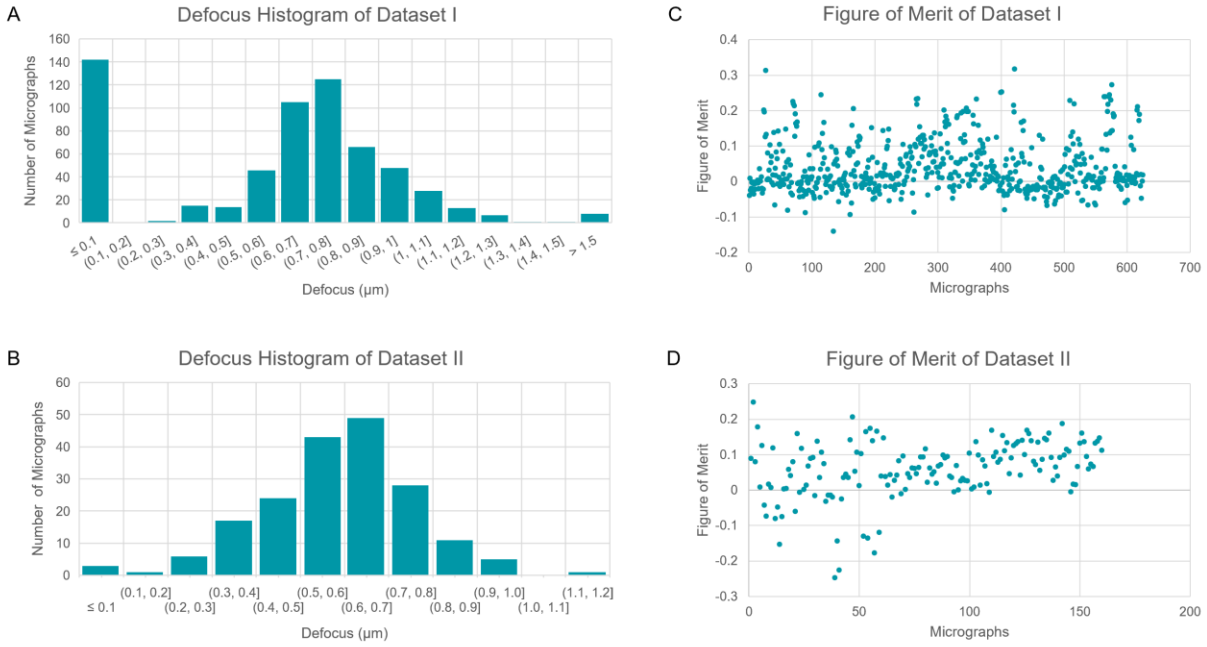


Figure 3.2. Statistics of defocuses. (A and B) The defocus distribution of dataset I (A) and dataset II (B). (C and D) The CTFFIND4-measured figure of merit distribution of dataset I (C) and dataset II (D).

Large phase shifts caused by large, accumulated dose on one VPP spot were previously found to strongly correlate with increased CTF artifacts and increased image blur (Danev et al., 2017). However, such an effect was not obvious for the micrographs in dataset II. The micrographs with phase shifts ranging from 72 to 144 degrees have comparable visual quality (Fig. 3.3 A). The 2D class averages of the particles showed fine structural features (Fig. 3.3 B). Since image "blur" is rather subjective, we used the high-resolution limit reported by CTF fitting to quantitatively represent image blur - lower resolution limit means more blurring (Fig. 3.8). From the plot of the *CTFFIND4*-determined maximum resolution as a function of the phase shift of dataset II (Fig. 3.8), except for a small number of outliers probably due to fitting errors, there was no obvious correlation between image blur and increased phase shifts.

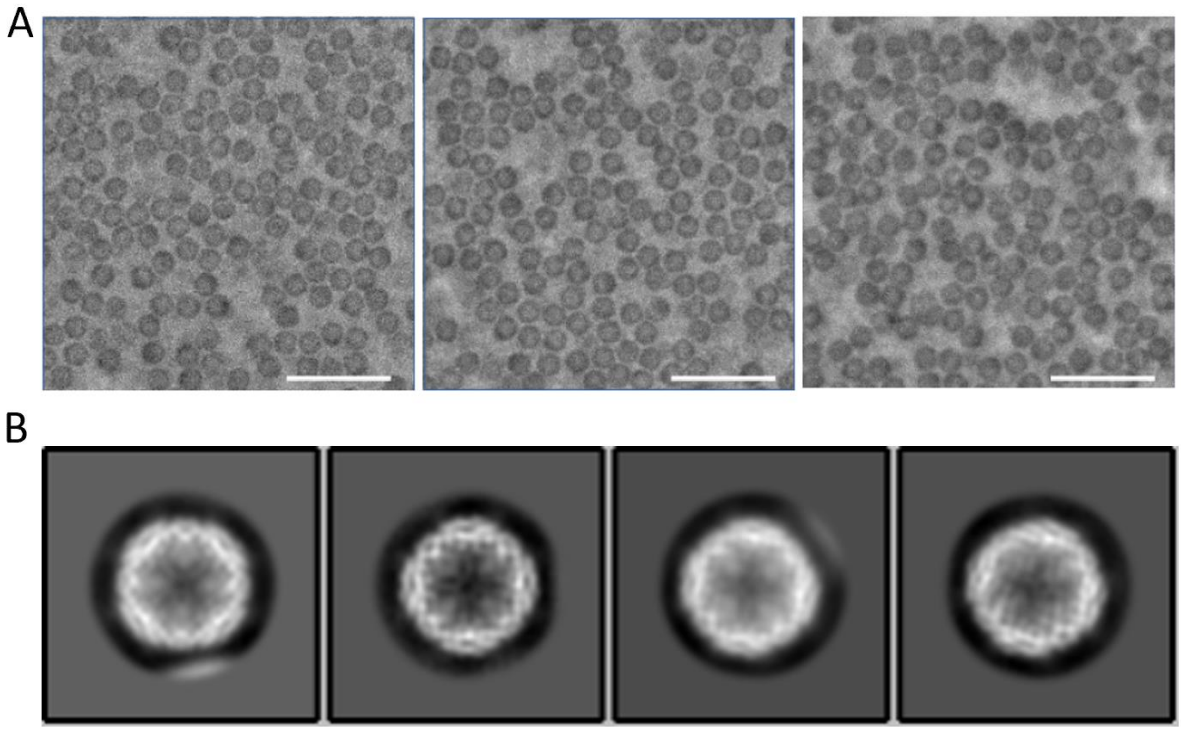


Figure 3.3. 2D and 3D classification results of the dataset II. (A) Representative micrographs with phase shift of 72, 99 and 144 degree, and the same defocus (0.5 μm). The scale bar represents 500 nm in length. (B) 2D class averages showing clear structural features.

The reconstruction of dataset II reached a resolution of 2.85 Å according to the "gold-standard" Fourier shell correlation (FSC) 0.143 criterion (Fig. 3.4 A), essentially the same as that of dataset I (Fig. 3.4 A). Both the random subset and around 90 degree phase shift subset of dataset I also reached similar resolution of that of the whole dataset I (Fig. 3.4A). The density maps of both datasets are also very consistent as shown by the FSC of the atomic model and the four maps (Fig. 3.4 B). The *RELION*-estimated accuracies of the center positions and euler angles of the different datasets in each iteration were plotted in Fig. 3.4C and 4D, respectively. Dataset II thus has similar quality with dataset I according to the FSC curves and the euler angle/position accuracy profile through the iterations. In the close-up view of the densities, the side-chains were clearly resolved and the atomic model visually fits well in the densities of both datasets (Fig. 3.4 E, F). The high quality of the density maps is also indicated by the validation statistics of the atomic models derived from the maps (Table 3.1).

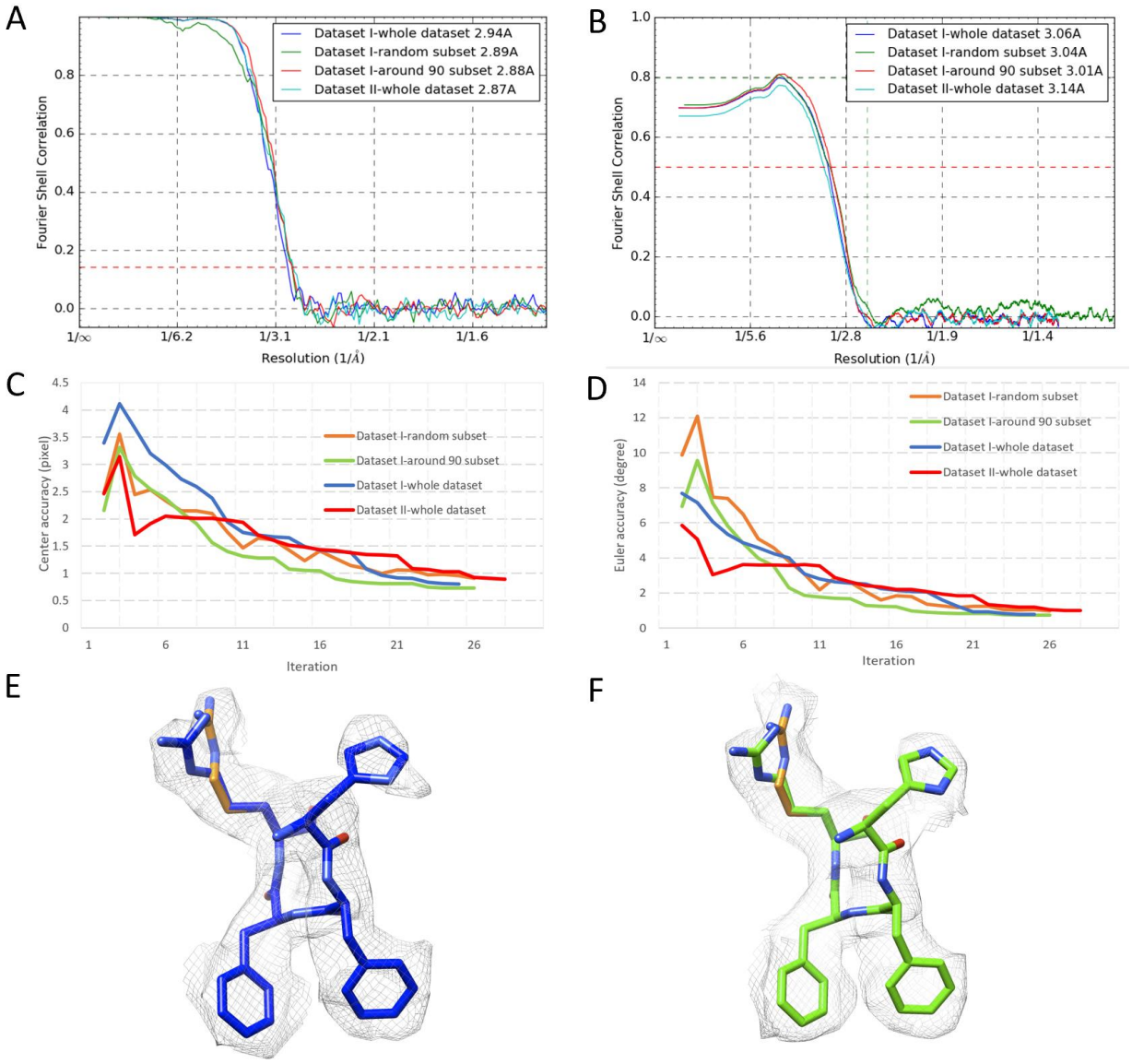


Figure 3.4. Comparison of the 3D structures reconstructed from dataset I and dataset II. (A) Noise-substitution corrected FSC curves of the dataset I and the dataset II. (B) The *phenix.mtriage* calculated model-map FSC curve of the three maps of dataset I and the map of dataset II. (C and D) The estimated accuracy of center (C) and Euler angles positions (D) in each iteration of 3D refinements reported by RELION/2.0. (E and F) Close-up view of the densities and the atomic model of an alpha-helical segment (a.a. 53–56) for dataset I (E) and dataset II (F).

3.4.2 Particles with phase shifts in the full range have similar quality

As the VPP was installed in the microscope only for about six months when datasets I and II were collected, we thus collected more datasets after an additional six to eight months to probe the variability of VPP and test if the results of datasets I and II could be reproduced. Figure 3.5A shows the phase shift trends of three of the six Volta spots included in dataset III. It is apparent that different Volta spots have different phase shift increasing rates, which is probably due to the intrinsic randomness of Volta potential generation process. The phase shift increasing rates are also much larger than the rates of datasets I and II and lack the clear biphasic profile seen in dataset II, which suggests that the age and/or the column environment can significantly influence the VPP and cause the variability from spot to spot and time tot time. The phase shift histogram (Fig. 3.5B) shows that dataset III has micrographs in the phase shift range of 0-250 degrees with the peak at 126-144 degrees. We could obtain a 3D reconstruction at 2.51 Å resolution using dataset III (Fig. 3.5C) after overcoming an unusual level of beam tilt issues (discussed below). To further investigate the relationship of phase shift and image quality, we divided the refined particles of dataset III into three subsets based on the *JSPR*-refined phase shift values: 0-120, 120-240, 240-360 degrees. The latter contains the least number of particles of 6,719. To ensure fair comparison, we randomly selected 6,719 particles from the first two ranges as the representative subsets of these phase shift ranges. The refinement results of the three subsets of particles (Fig. 3.5C) showed that all three subsets could be refined to similar resolutions, 2.59 Å for 0-120, 2.59 Å for 120-240, and 2.55 Å for 240-360 degree phase shifts. There is no systematic quality difference for either of these phase shift ranges.

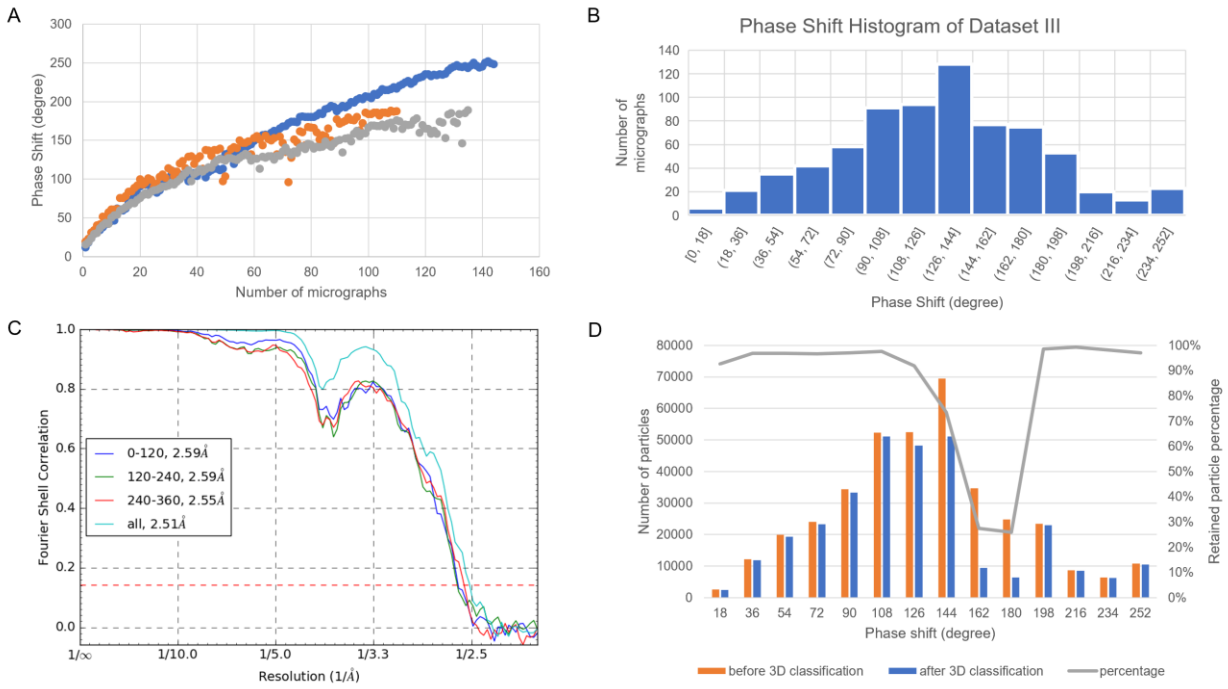


Figure 3.5. Image processing results of dataset III. (A) The time-courses of phase shifts of three different Volta phase plate spots showing different phase shift increment rates. (B) The phase shift histogram of dataset III determined by CTFFIND4. (C) The FSC curve of the half maps reconstructed with all particles and with particles in the phase shift range of 0–120, 120–240 and 240–360 degree. (D) The distribution of the number of particles before and after 3D classification using RELION and the percentage of retained particles after 3D classification as a function of phase shift.

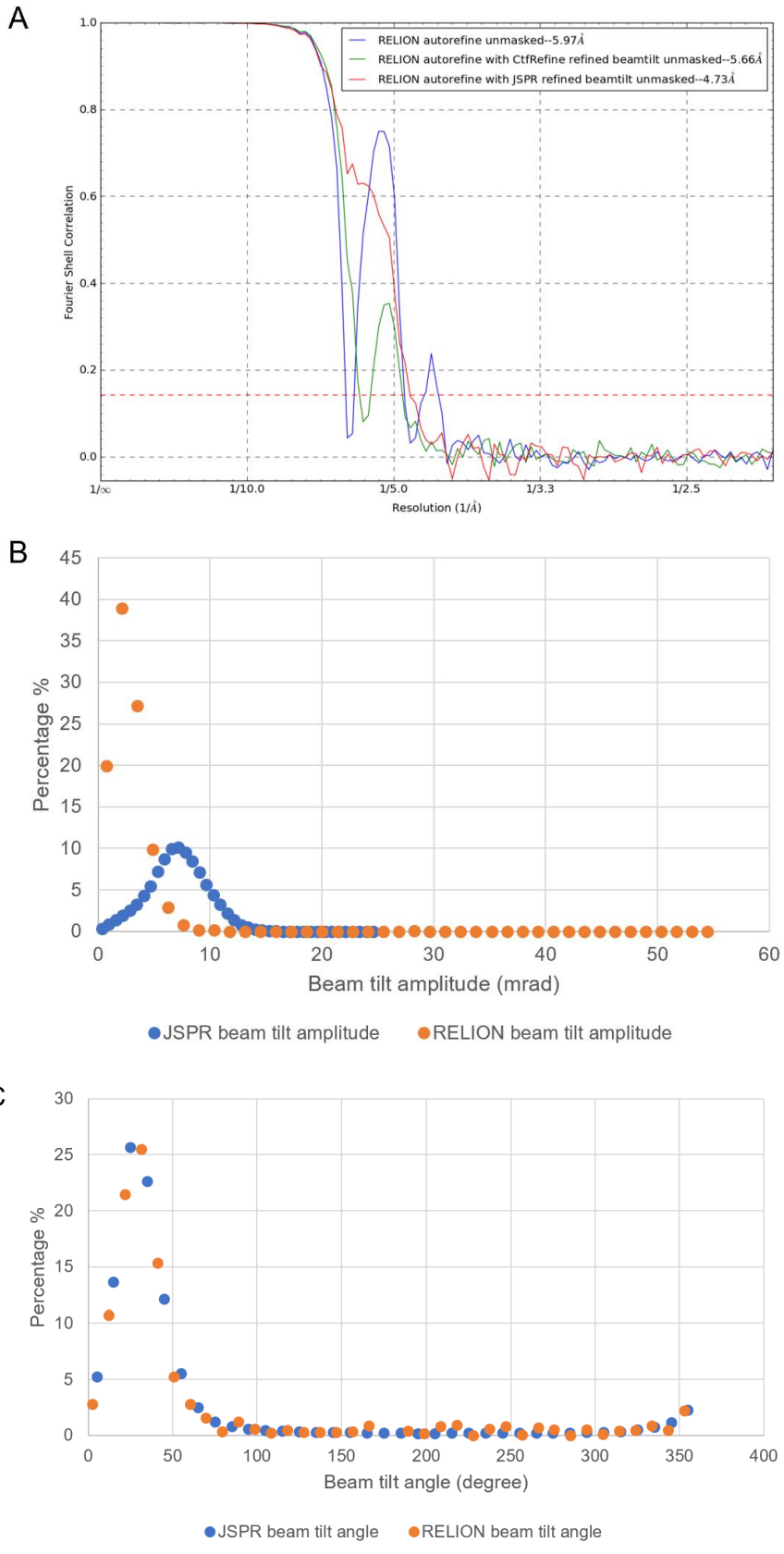
The fraction of retained particles after 3D classification was shown in Fig. 3.5D for the entire range of phase shifts. Except for the particles in the phase shift range of 160-180 degree, most particles in all other phase shift ranges were retained (>80%). Similar particle retention ratios were also observed for 2D classification of the particles (Fig. 3.9 C). We suspect that the drop in the 160-180 degree range was due to the low image phase contrast as the image condition for phase shifts in this range is essentially equivalent to VPP-free condition and the low resolution signals important for 2D alignment have been suppressed by the CTF.

3.4.3 Computational refinement and correction of beam tilt

Due to unknown reason(s), there was a large amount of beam tilt in dataset III. After *RELION/2.1* 3D autorefine, two strong oscillations were seen in the FSC curve (blue curve in Fig. 3.6A) of the unmasked maps, which indicates severe CTF artifacts. To identify the source of such CTF artifacts, we used *JSPR* (Guo and Jiang, 2013; Liu et al., 2016) to refine multiple parameters individually including defocus, astigmatism, beam tilt, magnification and distortion, and eventually found that beam tilt was the major cause. After iterative refinements using *JSPR*, the FSC oscillations were significantly reduced and the resolution was improved to ~2.5 Å (Fig. 3.5C). After the release of *RELION/3-beta*, we also used its CTFRefine function to refine beam tilt, astigmatism, phase shift (per micrograph) and defocus (per particle). Although the refinement could noticeably reduce the FSC oscillations (green curve in Fig. 3.6A), the remaining strong oscillation in the FSC curve suggests that the beam tilt effects were only partially corrected. Comparing the refined beam tilt parameters generated by *JSPR* and *RELION/3-beta*, the tilt angles are very consistent (Fig. 3.6 C) but the tilt magnitudes are different by about 3-fold (Fig. 3.6B). By replacing the *RELION* beam tilt parameters with that of *JSPR*, the FSC was improved and the remaining FSC oscillation could be eliminated (red curve in Fig. 3.6A).

Figure 3.6. Beamtilt refinement results of dataset III. (A) The FSC of unmasked maps reconstructed using the parameters of RELION/2.1 3D autorefine job (blue), RELION/3 CtfRefine job output star file (green), and the RELION/3 CtfRefine job output star file with beam tilt parameters replaced by JSPR refined beam tilt parameters (red). The dataset III was first refined with RELION/3 3D autorefine, which was limited to ~6 Å resolution based on the unmasked map FSC (blue). After running CtfRefine with one beam tilt group per micrograph in RELION/3, relion_reconstruct_mpi was used to reconstruct the half maps respectively (green). The rlnBeamTiltX and rlnBeamTiltY value of all particles were then replaced by the JSPR refined beam tilt parameters and half maps were reconstructed with relion_reconstruct_mpi (red). (B and C) Comparison of the beam tilt magnitudes (B) and angles (C) estimated by JSPR and RELION/3.

Figure 3.6 continued



3.4.4 Incomplete CTF model for VPP at low resolutions

Although dataset II has already shown that a single VPP position allows sufficient amount of images for near-atomic resolution 3D reconstructions, the number (164) of images in dataset II was still relatively small. Dataset IV was thus collected to test if a much larger number (461) of images using a single VPP position can still allow high-resolution 3D reconstructions. Fig. 3.7A shows the phase shift profile of dataset IV with two regions (orange triangle). The first region is very similar to dataset II which has a biphasic profile. The second region with the phase shift larger than 180 degree were assigned to a value in the range of 0-180 degree by *CTFFIND4*. By adding 180 degree to the phase shifts in this region, we are able to unwrap the phase shift to the range of 0 to 360 degree (blue dots). We noticed that some data points do not follow the phase increasing trend, due to failed CTF fitting based on visual inspection of the power spectra of the micrograph and the power spectra computed with fitted CTF parameters. The incorrect CTF fitting was likely due to the lack of sufficient number of visible Thon rings for images of small defocuses. Although about half of the images had phase shifts larger than 180 degrees (Fig. 3.7A) and the phase shift distribution has a peak at 194-212 degrees (Fig. 3.7B), this dataset still allows high-resolution reconstruction at around 2.93 Å resolution (Fig. 3.7D).

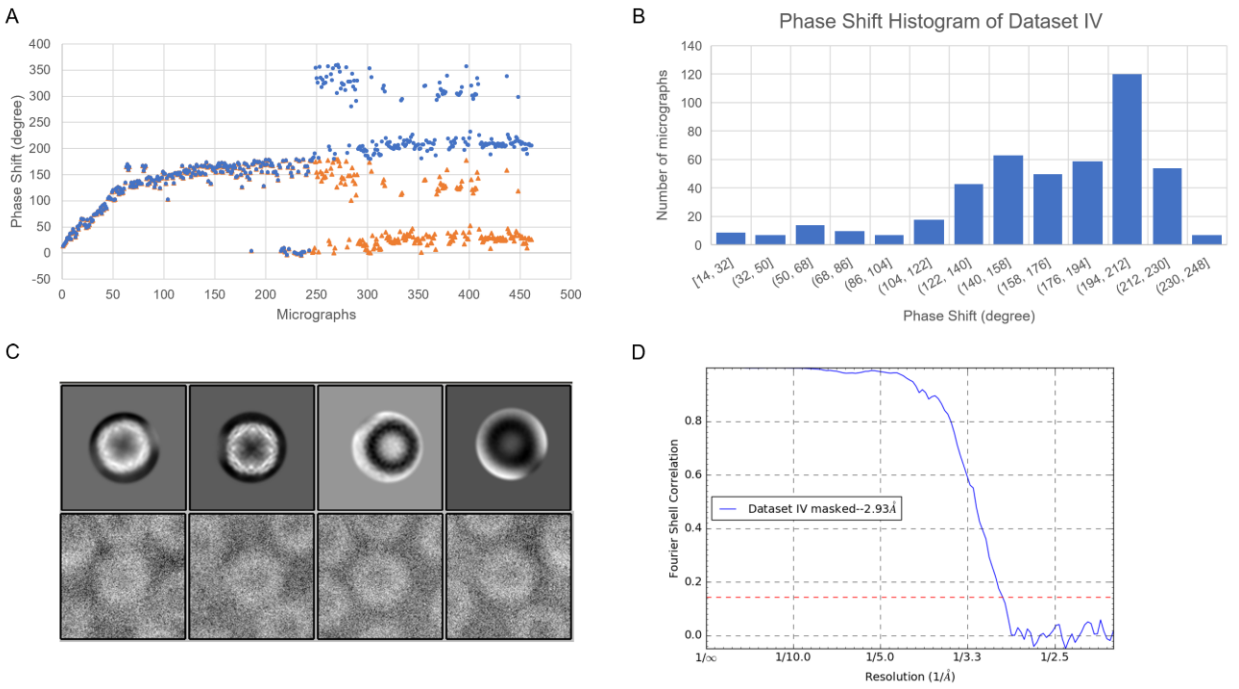


Figure 3.7. Image processing results of dataset IV. (A) The time-course of phase shifts determined by CTFFIND4 (orange triangles) and after manually adding 180 degree to the later micrographs (blue dots). (B) The phase shift histogram of dataset IV. (C) The white and black classes in the 2D classification results of dataset IV (top row) and representative particles in the black classes (bottom row). (D) The FSC curve for masked half maps.

During the image processing of dataset IV, we were initially puzzled by the observation of black classes in the 2D classification result of *RELION/3.0* (Fig. 3.7C top row). However, the particles in the black classes all have white protein density (Fig. 3.7 C bottom row) but their phase shifts are larger than 180 degrees. Based on our understanding of the current CTF model, a contrast inversion is expected for the particles with phase shifts in the range of 180-360 degrees. On the contrary, we didn't observe such contrast inversion in the experimental data. It indicates that there is inconsistency between the current CTF model and the behaviour of images collected. Therefore, we began to realize that, when the Volta spot on the phase plate is not infinitely small, the phase shift at different spatial frequencies can be different as has been previously described (Danev et al., 2017) and the phase shifts at low resolutions should be smaller than 180 degrees as suggested by the lack of contrast inversion in the images. Such an understanding is corroborated by the observations of negative FSC at low resolutions of the maps reconstructed from the particles of phase shift larger than 180 degree and the particles with phase shifts less than 180 degree and the

transition from negative to positive FSC at ~90 Å resolution for both dataset III (Fig. 3.11 A) and dataset IV (Fig. 3.11 B).

3.5 Discussion

In this work, we studied if a Volta phase plate position is intrinsically limited to a small number of images (a few tens) and a narrow range of phase shifts (around 90 degrees). By acquiring four datasets with different strategies of changing VPP positions, varying number of images for a single position, and across more than six months' time period, we have shown that it is feasible to image a large number of images (>400) with full range of phase shifts (>180 degrees) using a single VPP position to obtain 3 Å and better resolution 3D reconstructions, and such results are reproducible despite noticeable variations among different VPP positions and different time periods.

Datasets I and II were collected with the same sample grid and imaging condition but with different VPP imaging strategies. Imaging on a single spot of VPP allowed us to collect enough data (dataset II) for high resolution structure determination. The apoferritin structure at 2.85 Å resolution obtained with dataset II is nearly identical to the structure reconstructed from the control (dataset I). This success with a single VPP position was initially surprising as it is at odds with the current understanding of VPP and the recommended data collection strategy involving VPP.
Datasets III and IV, which were collected more than six months later, further demonstrated that the "surprising" results of dataset II obtaining high-resolution 3D reconstruction from a large number of images with a wide range of phase shifts acquired using a single VPP position were indeed a reproducible outcome of VPP. Furthermore, the comparable resolutions achieved from equal number of particles in the three different phase shift ranges indicated that the increase of phase shift does not necessarily lead to the decrease of image quality.

Nevertheless, a large number of images using a single VPP position did result in particles of full range (>180 degrees) of phase shifts that require some modifications of the image processing tasks to address the shortcoming of current image processing methods. Current power spectrum-based CTF estimation methods cannot distinguish the phase shifts with 180 degree differences. However, the time-course of the phase shifts provides sufficient information to allow us to manually add 180 degrees to the fitted phase shift values of the later images and to obtain correct phase shifts. In the future, the CTF fitting methods should consider the time-course

information and automatically detect and correct the 180 degree ambiguity. Furthermore, by posing the phase shift determination task as a generalized 2D alignment task similar to the refinement of Euler angles, defocus, beam tilt, *etc*. and using projection matching as implemented in *JSPR* (Guo and Jiang, 2013), the power-spectra based estimates of the phase shifts can be further refined to improve the reconstructions. Currently, the CTF model considers uniform phase shifts across the entire resolution range. However, our data with phase shifts larger than 180 degrees have indicated that the phase shifts in the low resolution range should be smaller. An analytic form or computationally optimized approximation of the resolution-dependent phase shifts should be developed in future studies to help reduce the associated adverse effects from the inaccurate phase shifts at low resolutions and improve the resolution of the 3D reconstruction.

We hope that the results in this study may provide new insights into further improvement of both efficiency and robustness of VPP, and to help turn VPP into a plug-and-play device for high resolution cryo-EM.

3.6 Supplementary figures

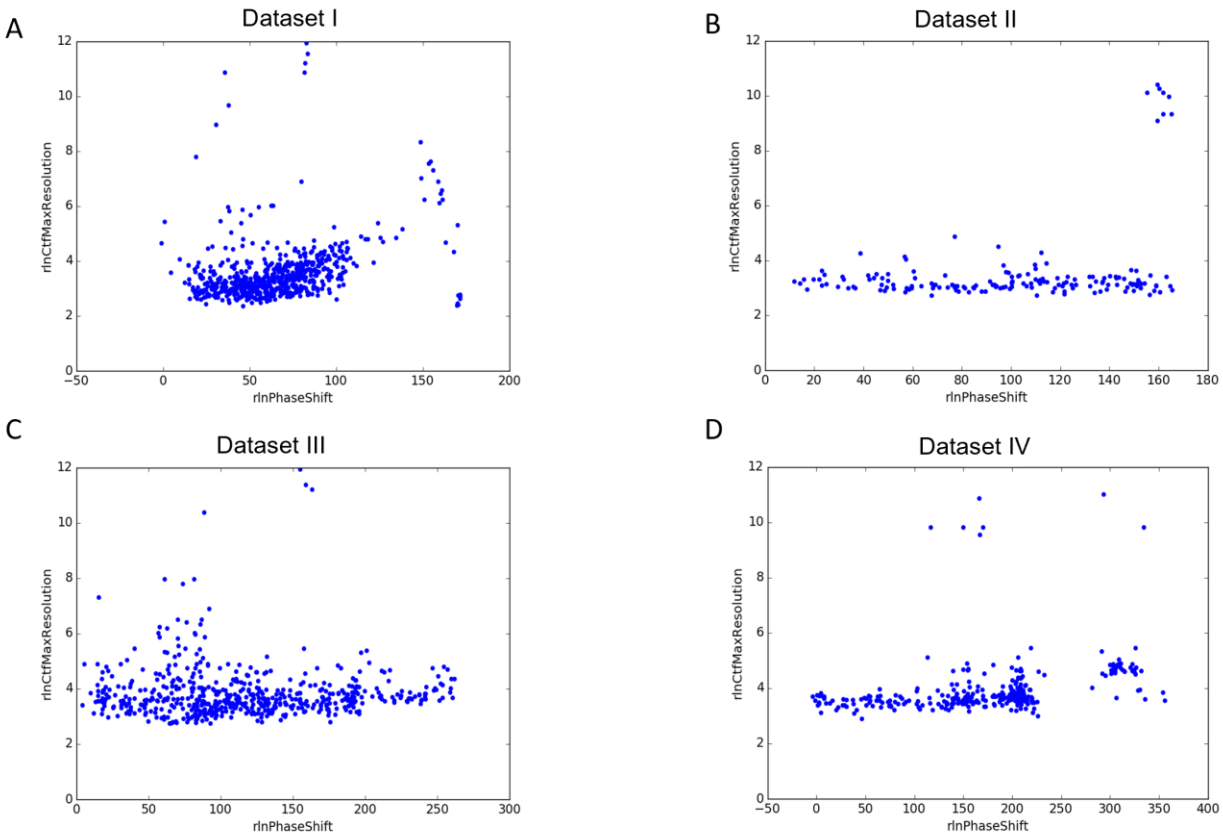


Figure 3.8. The CTFFIND4 estimated phase shift and maximum resolution of each micrograph of the four datasets.

Figure 3.9. The 2D and 3D classification retained particle percentage of the four datasets. The number of particles in different phase shift bins before 2D classification, after 2D classification, and after 3D classification are represented by the blue, orange, and grey bars respectively. The retained particle percentages after 2D classification and 3D classification are shown as the yellow and blue lines respectively. The 2D and 3D classification plot for dataset II used the particle numbers of the second round of processing.

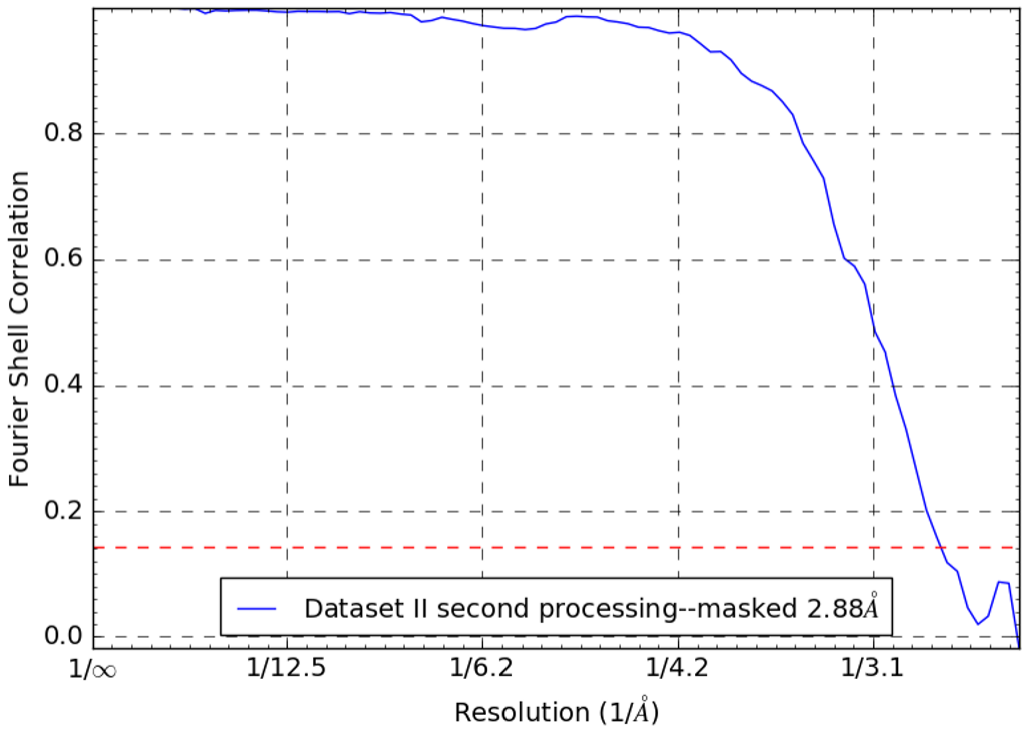


Figure 3.10. The FSC between the JSPR refined half maps of the second round reconstruction of dataset II.

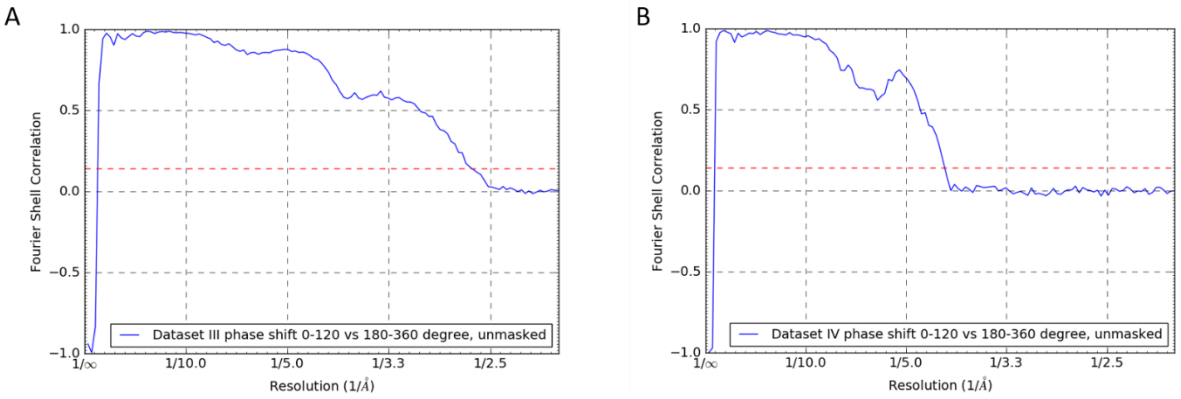


Figure 3.11. The FSC between the full unmasked maps reconstructed from 0-120 degree phase shift particles and 180-360 degree phase shift particles in dataset III (A) and dataset IV (B) showing the negative FSC at low resolutions.

CHAPTER 4. CRYO-EM STRUCTURE OF HETEROLOGOUS PROTEIN COMPLEX LOADED THERMOTOGA MARITIMA ENCAPSULIN CAPSID

This chapter is from a publication (Xiong et al., 2020) in collaboration with other scientists across different institutions. Their contributions are outlined in the journal website.

4.1 Abstract

Encapsulin is a class of nanocompartments that is unique in bacteria and archaea to confine enzymatic activities and sequester toxic reaction products. Here we present a 2.87 Å resolution cryo-EM structure of *Thermotoga maritima* encapsulin with heterologous protein complex loaded. It is the first successful case of expressing encapsulin and heterologous cargo protein in the insect cell system. Although we failed to reconstruct the cargo protein complex structure due to the signal interference of the capsid shell, we were able to observe some unique features of the cargo-loaded encapsulin shell, for example, an extra density at the fivefold pore that has not been reported before. These results would lead to a more complete understanding of the encapsulin cargo assembly process of *T. maritima*.

4.2 Introduction

Membrane-based organelles in eukaryotic cells are important cellular structures with various functions, including confining enzymes and substrates to enhance enzymatic activity and limit the propagation of potential damage caused by toxic products. In bacteria and archaea, compartmentation is realized by protein-based containers (Kerfeld et al., 2010; Martin, 2010). Encapsulin is a kind of protein compartment that exists in a wide variety of bacteria and archaea (Akita et al., 2007; Cornejo et al., 2013; Nicolas et al., 2017). Encapsulins are assembled by capsid proteins into an icosahedral shell with a diameter between 24 and 42 nm, which encapsulates native cargo proteins involved in oxidative stress and iron mineralization (Akita et al., 2007). The conserved C-terminal extension of native cargo proteins, which is called cargo-loading peptide (CLP), mediates the interaction between the interior surface of encapsulin and cargo protein (Sutter et al., 2008). There are also several examples of non-native cargo by fusing the CLP to heterologous proteins like teal fluorescent protein, superfolder green fluorescent protein (sfGFP),

and mNeonGreen in *Brevibacterium linens*, *T. maritima*, and *Myxococcus xanthus*, respectively (Cassidy-Amstutz et al., 2016; Hicks et al., 1998; Lau et al., 2018). However, the conformational changes and the packaging mechanisms of encapsulin and its cargos are poorly understood.

Multiple attempts have been made to solve the structure of encapsulin in *T. maritima* (Putri et al., 2017), *M. xanthus* (McHugh et al., 2014), and *Pyrococcus furiosus* (Akita et al., 2007) along with its cargo inside. However, none of these studies were able to obtain a high-resolution cargo structure. In a recent study of the cargo-loaded encapsulins from *Quasibacillus thermotolerans* (Giessen et al., 2019), X-ray crystallography was applied to solve the cargo protein structure before fitting it into the focused refined low-resolution cargo density of the encapsulin map. Considering their programmability and biocompatibility, encapsulin is a promising platform for the delivery of proteins or small ligands and nanoreactors (Avalos et al. 2013). Using *Escherichia coli* or yeast as a host, nonnative cargo proteins were packaged into encapsulins successfully in vivo (Cassidy-Amstutz et al., 2016; Lau et al., 2018). However, there are very few studies on the heterologous cargo-loaded encapsulins.

Here, using the insect cell expression system, we encapsulated a heterologous macromolecular protein complex named IDM complex in *T. maritima* encapsulin (hereafter, Encap/IDM complex), which contains Arabidopsis protein IDM1, IDM2, IDM3, HDP1, HDP2, and MBD7 with a total mass of 300 kDa which is the largest cargo so far (Sonotaki et al., 2017; Duan et al., 2017), and determined the encapsulin structure using cryo-EM to 2.87 Å resolution. Despite the interference of the inner cargo density, we have successfully solved the encapsulin shell structure with a large cargo inside.

4.3 Materials and Methods

4.3.1 Reconstitution and purification of IDM complex loaded *T. Maritima* encapsulin

The six IDM complex subunit genes were amplified from Arabidopsis cDNA. T. maritima encapsulin gene was amplified from pET14-GFP30-Encap (Addgene #86406). The IDM1 C-terminus is tagged with the amino acid sequence of CLP (LFTDKPITEIEEETSGGSENTGGDLGIRKL). All six IDM proteins and C-terminus His-tagged Encap were cloned into pFastBac1 individually. A truncated version of MBD7 was expressed from Ser232 to the end of the C-terminus. For IDM complex expression, a His-sfGFP tag and MBP

(maltose binding protein) tag were fused at the N-terminus of IDM1 and MBD7, respectively. While in Encap/IDM1 and Encap/IDM complex, a C-terminal fused CLP of IDM1 was used instead.

Recombinant baculoviruses were generated in Sf9 cells. All seven genes were co-expressed in insect cells with a density of 2 x 106 cells per mL and then collected after 60-72 h incubation at 27°C. Collected cell pellets were stored at −80 °C before protein purification.

The cell pellet was resuspended in buffer A [20 mM HEPES pH 7.5, 150 mM NaCl, 25mM Imidazole, 0.1 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP)] supplemented with Protease Inhibitor Cocktail and then lysed by sonication. The supernatant was collected after centrifugation at 18,000 × g for 30 min at 4 °C. The Ni-NTA resin was washed with 10 column volumes of buffer A and then incubated with the supernatant at 4°C for 1h. The complex was eluted by buffer B (buffer A with 250mM Imidazole) after buffer washing and then loaded into Superdex 200 10/300 GL column equilibrated with buffer A without imidazole. Fractions were collected and analyzed with SDS-PAGE and coomassie brilliant blue staining.

The reconstitution and purification of encapsulin loaded with IDM1 alone (hereafter, Encap/IDM1) followed the same procedure above with only the baculoviruses of encapsulin capsid and IDM1.

4.3.2 Cryo-EM sample grid preparation

A 4 µL amount of 1 mg/mL of either Encap/IDM or Encap/IDM1 complex sample was applied to glow discharged C-flatTM CF-2/1-4Cu-50 grids (Electron Microscopy Sciences, Hatfield PA, USA) and plunge frozen with a Cp3 plunger (Gatan, Inc., Pleasanton CA, USA), 90% humidity, ashless filter paper (Whatman® #1001-032) (Cytiva, Marlborough MA, USA), and 2 s of blotting.

4.3.3 Cryo-EM data acquisition

Images of Encap/IDM1 were taken on Talos F200C transmission electron microscope (Thermo Fisher Scientific, Waltham, MA, USA) recorded by the Ceta 16M camera (Thermo Fisher Scientific, MA, USA) with a pixel size of 2.04 Å.

The sample grid of Encap/IDM complex was imaged with a Titan Krios transmission electron microscope (Thermo Fisher Scientific, MA, USA). Images were recorded using a Gatan K3 Summit detector (Gatan, Inc., Pleasanton CA, USA) mounted on a Gatan Quantum energy filter (Gatan, Inc., Pleasanton CA, USA) using a 20 eV zero-loss slit in super-resolution counting mode. The movies were collected using Leginon software (Suloway et al., 2005) at a dose-rate of 30 electron/pixel/second. A nominal magnification of 81,000 was used for data collection resulting in 0.536 Å super-resolution pixel size. The total dose of each movie was 56.4 electrons/Å2 from 2.16 s exposure. Detailed imaging conditions for the dataset are listed in Table 4.1.

Table 4.1. Data collection and image processing statistics

Data Collection	
Voltage (kV)	300
Dose (e/Å^2)	56.4
Nominal Magnification	81,000
Super-resolution Pixel Size (Å)	0.658
Number of Movies	2,175
Number of frames/movie	40
Intended Defocus (μm)	0.5-2
Image processing	
Particles picked	1,535,389
Final #particles	112,241
Refined resolution (Å)	2.86

4.3.4 Image processing

The dataset of Encap/IDM complex consisted of 2175 movies. The raw movies were aligned and dose-weighted with Motioncor2/1.0.5 (Zheng et al., 2017). The motion-corrected micrographs were imported into cryoSPARC/v2 (Punjani et al., 2017) for the following processing steps. A total of 1,535,389 particles were extracted by template matching particle picking. After multiple rounds of heterogeneous refinement, a homogenous set of 112,241 particles was used in ab initio reconstruction with C1 symmetry and the final homogeneous refinement with icosahedral

symmetry in cryoSPARC/v2 to 3.39 Å resolution. Further icosahedral refinement with JSPR (Guo and Jiang, 2014; Liu et al., 2016) resulted in 2.87 Å resolution based on the "gold standard" Fourier Shell Correlation at FSC = 0.143 criterion. The same set of particles was also used for homogenous refinement with C1 symmetry to 4.29 Å resolution in cryoSPARC/v2. Detailed processing statistics for the dataset are listed in Table 4.1. The electron density map of the encapsulin shell was deposited in the Electron Microscopy Data Bank (EMDB; https://www.ebi.ac.uk/pbde/emdb) under accession number EMD-22617.

4.3.5 Model refinement

The PDB model 3DKT was fitted into the electron density maps in *Chimera* (Emsley et al., 2010; Yang et al., 2012) and refined with *Rosetta* (DiMaio et al., 2011; Wang et al., 2016). The refined monomer was visualized in *Coot* (Emsley et al., 2010; Emsley and Cowtan, 2004), refined with PHENIX (Afonine et al., 2018), and then subjected to symmetry refinement in a pentamer unit with Rosetta symmetry refinement. The refined atomic model has been deposited in the Protein Data Bank (PDB; https://www.rscb.org/) under accession number 7K5W.

4.4 Results

4.4.1 Self-assembly of heterologous macromolecular cargo loaded encapsulins in baculovirus expression system

To encapsulate the IDM complex in vivo, six subunits of IDM and encapsulin from T. maritima were constructed into individual operons and coexpressed in the insect cells. Based on cross-linking coupled mass spectrometry analysis (data not shown), the C-terminal end of IDM1 (the largest subunit of the IDM complex) was free from protein–protein interactions within the IDM complex. The 30 amino acid cargo-loading peptide was fused at the C-terminal of the IDM1 protein (Figure 4.1 a, b). Purification of the encapsulin particles was accomplished through Ni-NTA affinity of the surface-exposed His-tag at the C-terminal of the encapsulin protein, followed by gel filtration. The confirmation of the copurification of encapsulin and the IDM complex was done by SDS-PAGE (Figure 4.2 a, b). As a pilot test, we first only incorporated the IDM1 protein alone into the encapsulin and observed the additional density inside of the encapsulin shell as shown in Figure 4.2 c. Encouraged by the successful packaging of the IDM1 protein, we then tried

to incorporate the whole IDM complex, hoping that the whole IDM complex could be packaged inside of encapsulin. With IDM1 alone containing the cargo-loading peptide, the IDM complex should have been preassembled before being loaded into encapsulin. The cryo-EM image of purified particles of the whole IDM complex (Figure 4.2 d) indeed had larger density inside the shell compared to the IDM1 protein alone loaded encapsulin (Figure 4.2 c) as also shown in the 2D class averages of the Encap/IDM1 (Figure 4.2 e) and Encap/IDM complex (Figure 4.2 f). It indicates a successful assembly of the heterologous IDM complex inside of the encapsulin shell with the insect cell expression system. However, we did notice a significant number of contaminant bands in the SDS-PAGE gel. We reasoned that since only the encapsulin capsid had the His-tag, if we could see the bands of both the proteins, they must have been copurified as a complex.

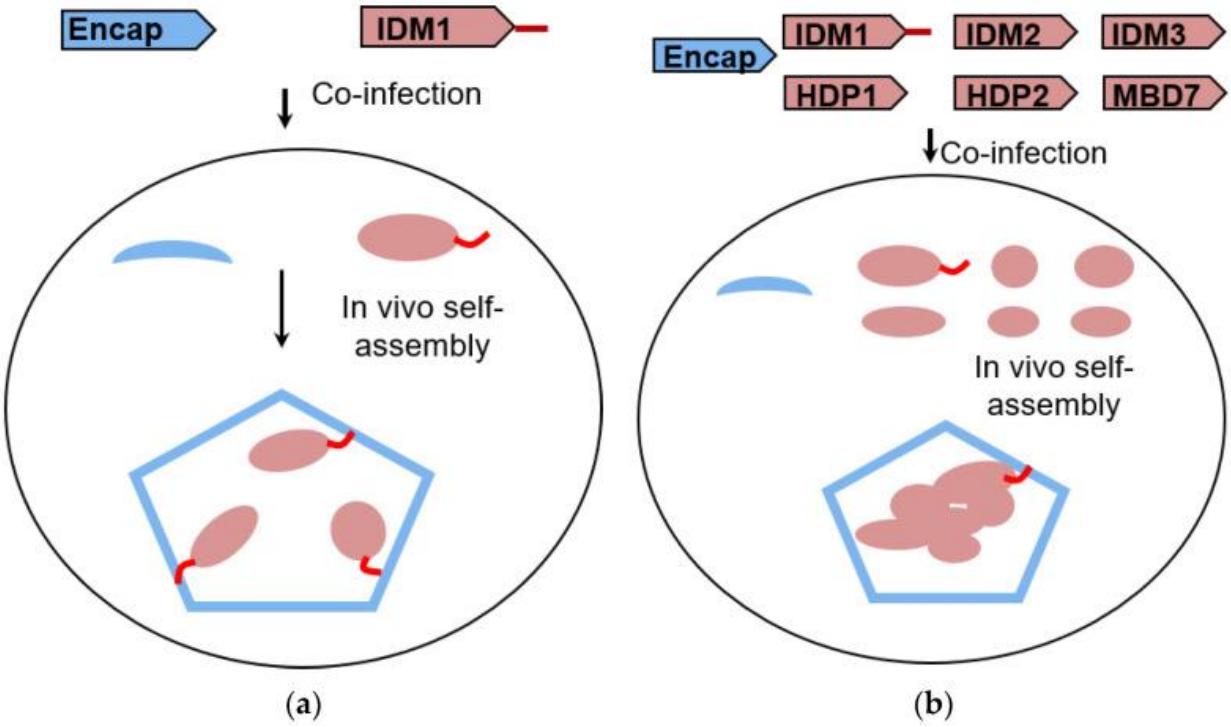


Figure 4.1. Schematic representation of heterologous cargo-loaded encapsulin expression in insect cells. Diagrams of in vivo encapsulin assembly resulted from the coexpression of Encap and cargo proteins in the baculovirus expression system with IDM1 as cargo alone (a) and the IDM holo complex as cargo (b).

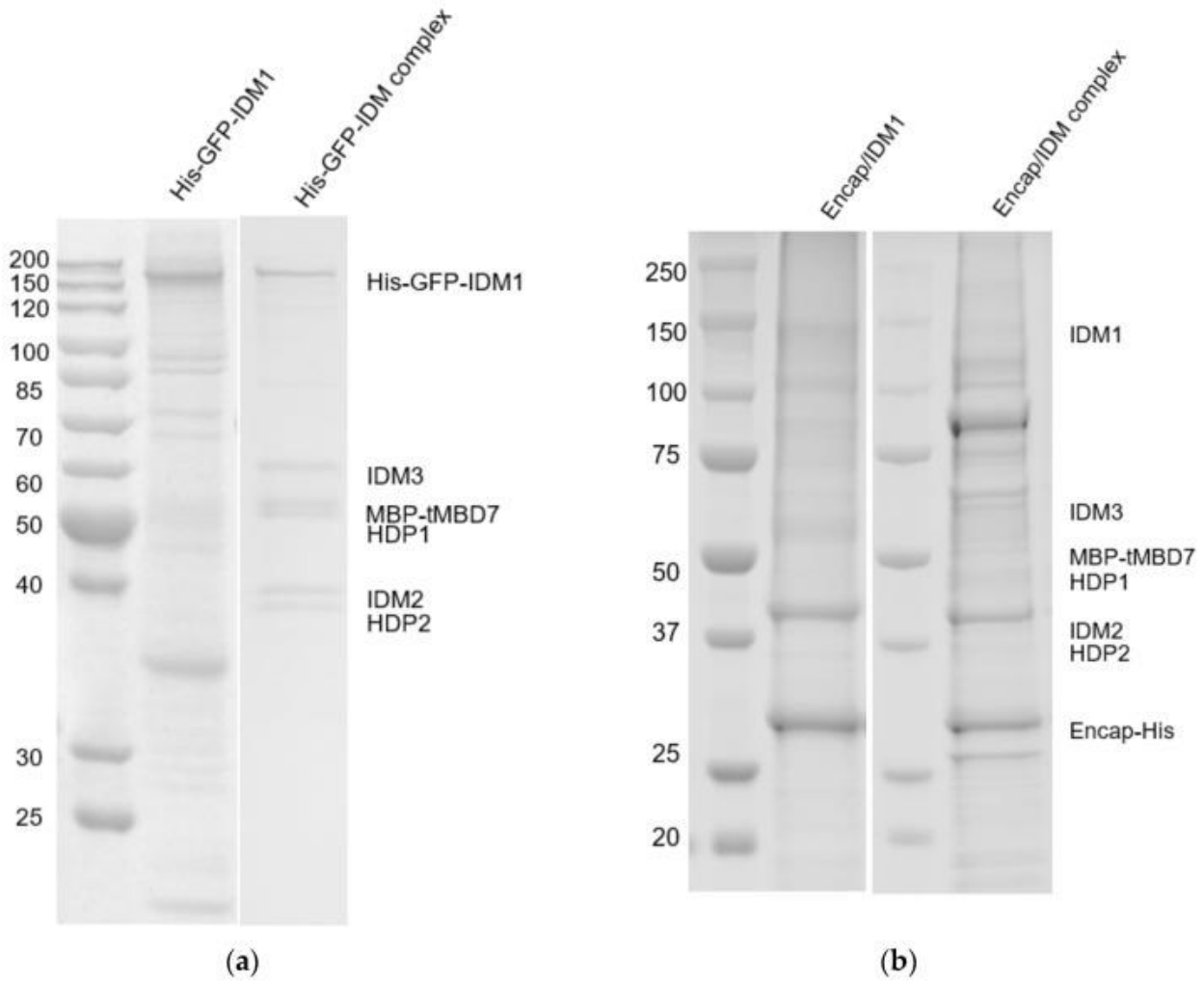
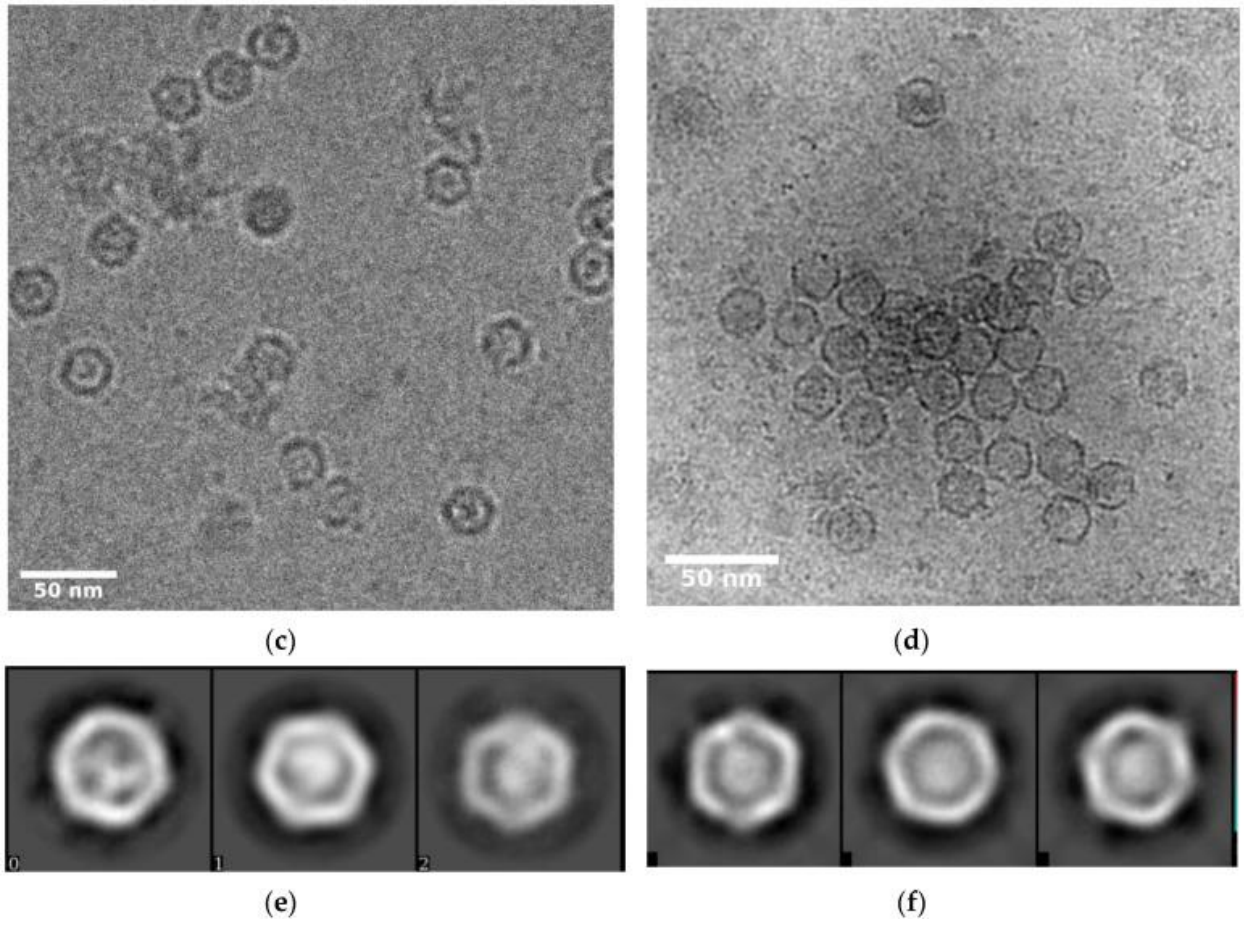


Figure 4.2. Production of heterologous cargo-loaded encapsulin. The SDS-PAGE result of His-GFP-IDM1 and His-GFP-IDM complex (a), Encap/IDM1 and Encap/IDM complex (b). His-GFP-IDM1: His-GFP fused at the N-terminal of IDM1. The IDM complex without encapsulin (a) was purified with the His-tag on the N terminal of IDM1 protein. In the Encap/IDM complex, there was no His-tag on IDM1 but was at the C-terminal of encapsulin capsid (Encap). The molecular weight of His-GFP tag, MBP tag, IDM1, IDM2, IDM3, HDP1, HDP2, and MBD7 are 32kDa, 40kDa, 131kDa, 39kDa, 52kDa, 46kDa, 33kDa, and 35kDa, respectively. Cryo-EM image of Encap/IDM1 (c) from Talos F200X G2 and Encap/IDM complex (d) from Titan Krios. Representative 2D class averages generated from the Encap/IDM1 (e) and the Encap/IDM complex dataset (f). The “matchto” processor in EMAN2 accessed through e2proc2d.py was used to filter the 2D class averages to the same level by matching their structure factor curves.

Figure 4.2 Continued



(c) (d)

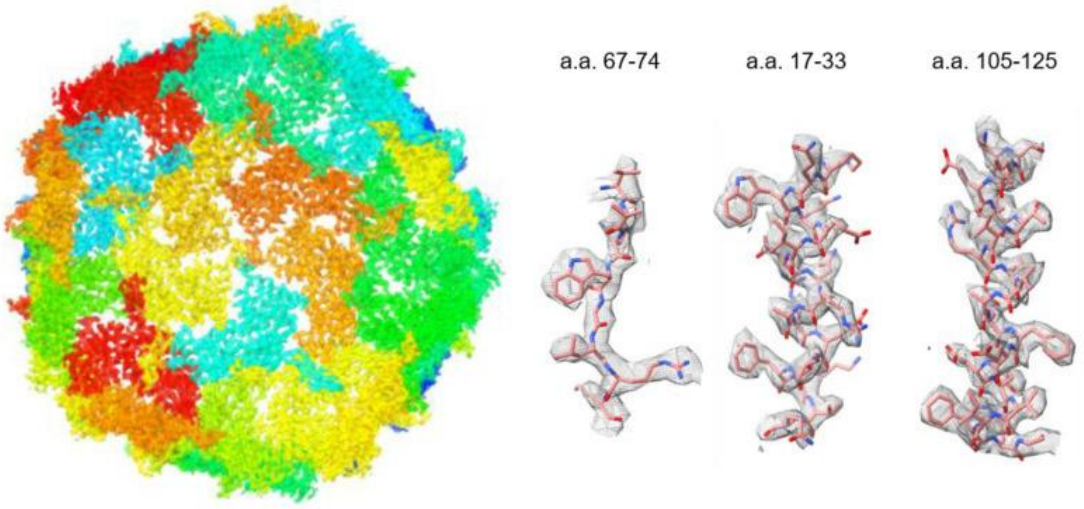
(e) (f)

4.4.2 Overall structure of IDM complex-loaded T. maritima encapsulin

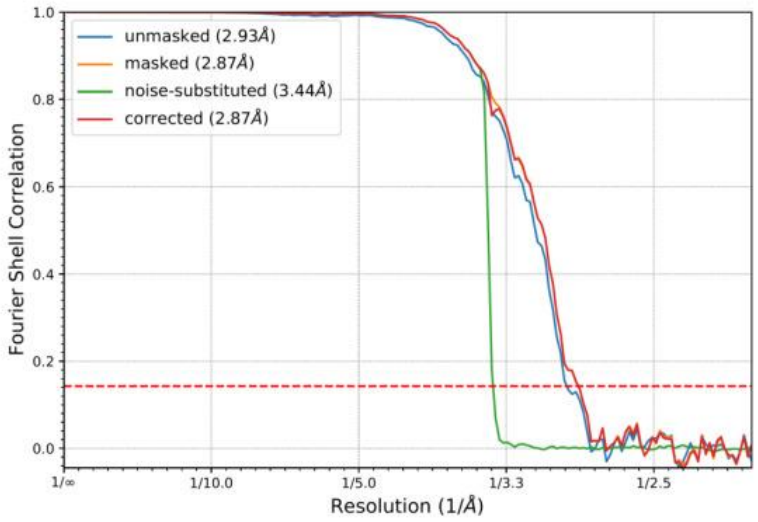
Using single-particle cryo-EM, we determined the structure of T. maritima encapsulin packaged with a heterologous IDM complex (Figure 4.4a). The overall resolution of the encapsulin icosahedral shell was solved at 2.87 Å based on the gold standard Fourier Shell Correlation (FSC) 0.143 threshold (Figure 4.3 c). To evaluate the cryo-EM density map, the crystallographic model of *T. maritima* encapsulin (PDB:3DKT) was docked into our density map. The crystal structure was further refined into the 2.87 Å resolution cryo-EM density map. The refined model fit the density map very well including the sidechain densities (Figure 4.3 b). After superimposing the refined model and original crystal structure of encapsulin (PDB:3DKT) (Sutter et al., 2008), the two structures were largely the same (RMSD = 0.681 Å) with only minor differences at the A-domain and the E-loop (Figure 4.4d). Both the A-domain loops and E-loops were slightly shifted without changing the overall shell structure of the Encap/IDM complex. It is consistent with the idea that A-domain loops and E-loops are the most flexible regions of encapsulin capsid protein. The electron density fitted monomer model demonstrating the map quality at E-loop and A-domain are presented in Figure 4.6. We did notice that there were missing densities at the Gly189 and Ala188 position, which was probably due to the flexibility of the A-loop. The density of those two residues were also missing the PDB-3DKT X-ray density map.

Figure 4.3. Overall structure of IDM complex-loaded T. maritima encapsulin. (a) The sharpened density map of Encap/IDM complex from the icosahedral reconstruction. (b) The closeup view of the side-chain densities of short segments of a.a.67-74, a.a.17-33, and a.a.105-125. (c) The FSC curve of the reconstructed cryo-EM half maps. (d) Structure alignment of the Encap (coral) and Encap/IDM monomer (light blue).

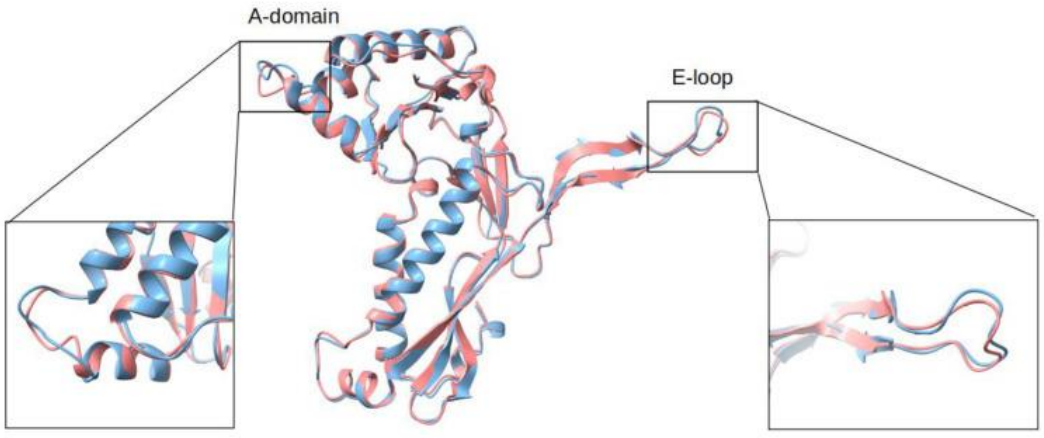
Figure 4.3 continued



(**a**) (**b**)



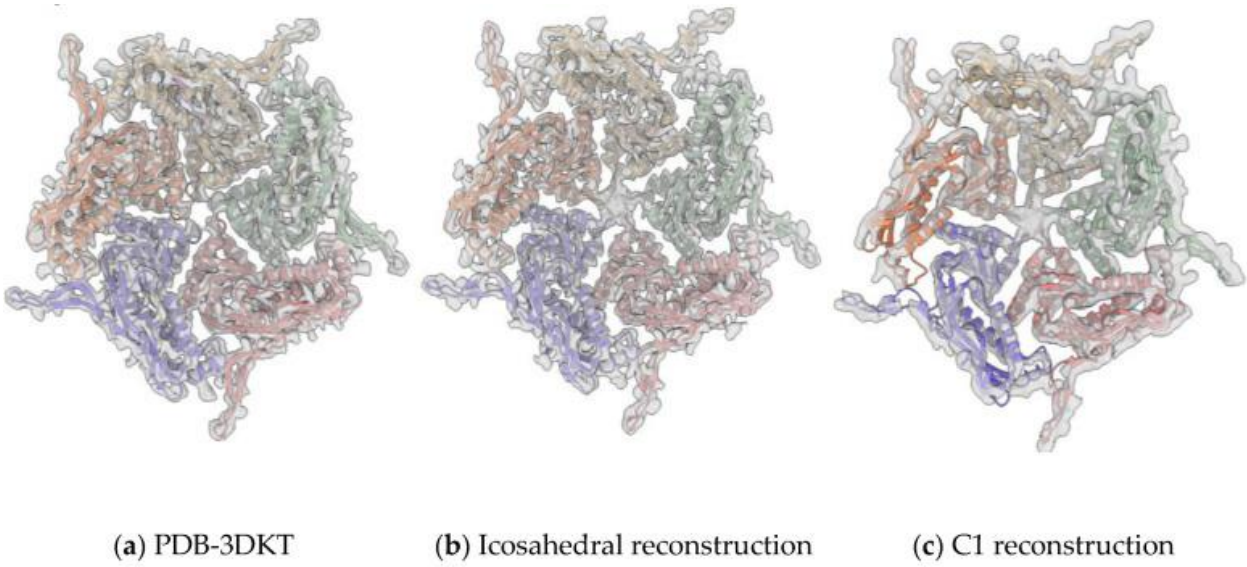
(**c**)



(**d**)

4.4.3 Pores of the Encap/IDM complex

The crystal structure of T. maritima encapsulin has multiple openings in its capsid shell, including the positively charged threefold pore and uncharged fivefold pore (Figure 4.4 a) (Sutter et al., 2008). The openings were thought to be the channel for the substrates and/or products of the cargo protein. In our IDM-loaded encapsulin structure, we have found extra density at the fivefold pore in both icosahedral shells (Figure 4.4 b). To test if the extra densities were an image-processing artifact of icosahedral reconstruction along the symmetric axes, we then performed de novo C1 reconstruction and found extra densities at all 12 fivefold pores in the C1 reconstruction (Figure 4.4 c). The extra density should thus be authentic structural features of the IDM-loaded encapsulin that might be caused by substrate binding or cell factors copurified from the insect cell. The C1 reconstructed map was refined to 4.29 Å resolution. The cryoSPARC-reported FSC curve is shown in Figure 4.8. Some part of the C1 reconstructed map is less resolved compared to the icosahedral map. We reasoned that when we performed the C1 symmetry refinement, the orientation alignment of the icosahedral shell of encapsulin was affected more by the cargo density. When we applied icosahedral symmetry, alignment was less vulnerable to the cargo density which did not have icosahedral symmetry and would have been averaged out in the icosahedral reconstruction. A zoom-in view of the extra density at the fivefold pore and conformational comparison of Encap and Encap/IDM complex is presented in Figure 4.5.



(**a**) PDB-3DKT (**b**) Icosahedral reconstruction (**c**) C1 reconstruction

Figure 4.4. Pores of the heterologous IDM-loaded encapsulin. The inner surface view along the fivefold axis of Encap/IDM complex. From left to right: the X-ray density map of PDB-3DKT (a), the crystal model of T. maritima fitted into the icosahedral reconstructed density map of our dataset (b), and the C1 symmetry reconstructed map (c).

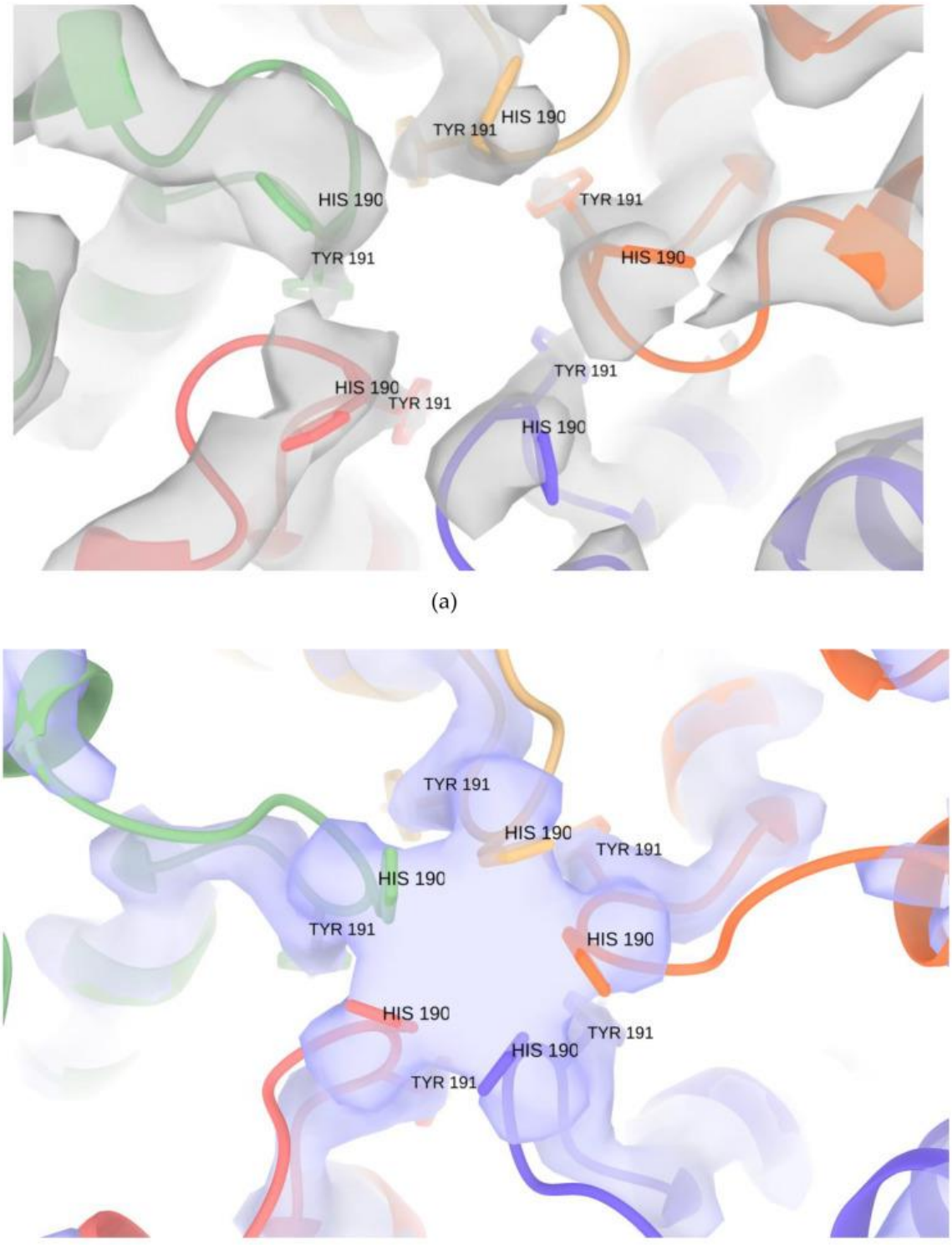


Figure 4.5. Conformation comparison of the crystal encapsulin structure (PDB:3DKT) fitted in the X-ray 2fo-fc map (grey) (a) and Encap/IDM complex (b) refined model fitted in the EM density map (purple) at the fivefold pore.

4.4.4 Structure of Encapsulated IDM Complex

We designed this experiment with the intention to protect the IDM complex from the air–water interface with the encapsulin shell and hoped to get a more uniform state of the IDM complex with the confined space inside of the shell. We have tried multiple image-processing software such as RELION and cryoSPARC and strategies including computational subtraction of the encapsulin shell signals to reconstruct the structure of the IDM complex inside of the Encap/IDM complex. However, we failed to obtain a good model for the IDM complex. It could be due to the heterogeneous compositions/conformations, low signal-to-noise ratio of the images, and interferences from the signals of the shell. In view of this failure, an attempt was made to recover the density of the CLP near the threefold axis with symmetry expansion and focused refinement in RELION. However, after multiple trials, the 3D classes of focused classification only contained the shell density. The density of CLP was expected to be hard to resolve since there was only one CLP per capsid, while when we did symmetry expansion, it would have been overwhelmed with the other expanded 59 sets of particles without the CLP. Therefore, even with a small number of particles having the loading peptide density, it could still be averaged out during the reconstruction. Although not being able to recover the CLP density at the threefold pore, we observed strong protein density inside of the encapsulin capsid in the focused refined map (Figure 4.7), which is another evidence of the existence of the inner protein complex. Future optimization of sample quality will be needed, for example, multiple CLPs fused to additional IDM subunits to enhance binding affinity and specificity, to allow successful structural determination of the encapsulated IDM complex.

4.5 Discussion

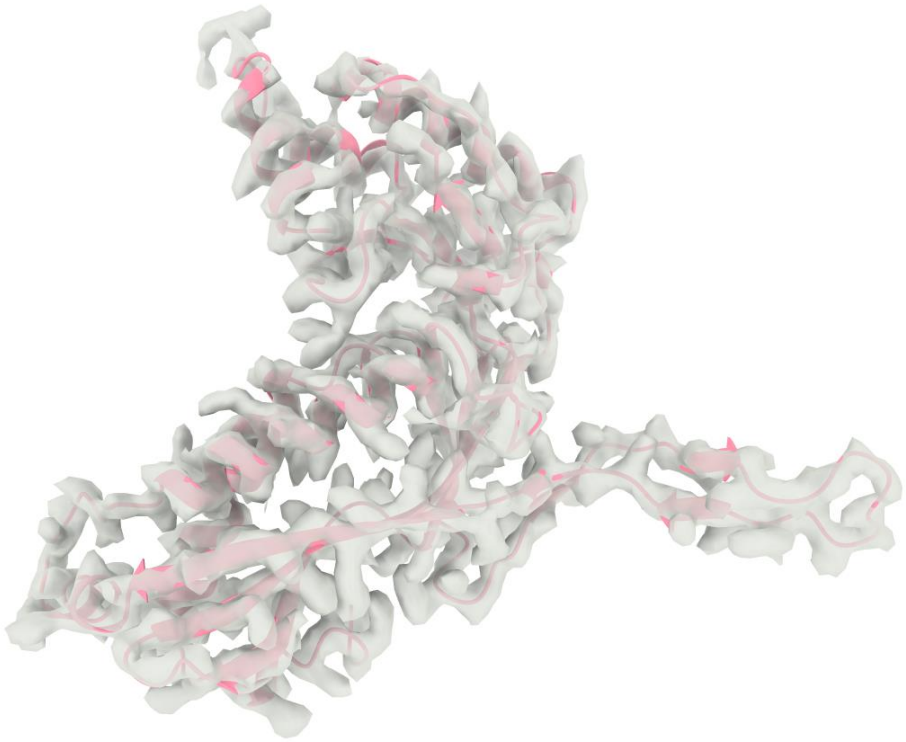
Our study is distinct from previous studies in three aspects. Firstly, we are presenting the first case of insect cell expression of encapsulin. As a reprogrammable platform, encapsulin has been expressed in several hosts such as E. coli, yeast, and mammalian cells but not insect cell systems. This project was designed to use encapsulin as a platform for solving protein complex structures that are particularly vulnerable to the air–water interface problem. The insect cell system has an advantage in expression of many eukaryotic proteins that require post-translational modifications and cannot be expressed in bacterial systems. This demonstration of successful

expression and formation of an intact encapsulin shell expands potential cargo targets that require insect cell expression systems.

Secondly, we have demonstrated the capability of coexpressing encapsulin and a heterologous cargo in the insect cell expression system. In a previous crystallography study, the *T. maritima* encapsulin was purified directly from its bacterial host and some of them might have contained native cargo (Sutter et al., 2008). In our study, encapsulin and cargo proteins were cloned separately in seven operons and self-assembled in vivo. The density inside of encapsulin in the TEM images indicated successful cargo loading and assembly.

Thirdly, the cargo that we chose to assemble inside of encapsulin was not a single protein but a multi-subunit complex which contributed to stronger cargo density inside of the encapsulin shell. Despite this, we were able to obtain a 2.87 Å encapsulin shell structure with a dataset of Encap/IDM complex. It is not surprising that we could not resolve the inner cargo structure and the loading peptide density, probably due to sample quality, the intrinsic flexibility of the IDM complex, the limited number of particles, and the interference of the shell signal. There is no doubt that the structure of the IDM complex inside and the loading peptide density on the shell would be more interesting findings. However, to characterize these further and hence to be able to shed light on the encapsulin assembly conformational changes, it will require significant improvement in both sample preparation and image-processing technique. Thus, if we can better handle these problems, it should be possible in the future to obtain an interpretable map of the IDM complex inside of encapsulin.

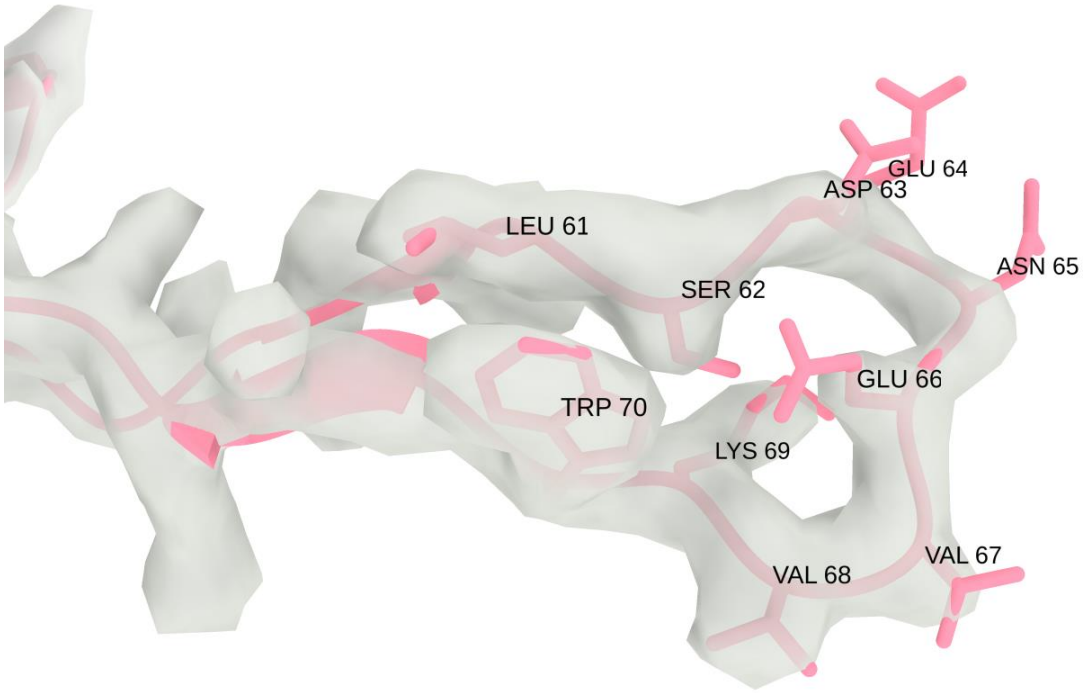
4.6 Supplementary figures



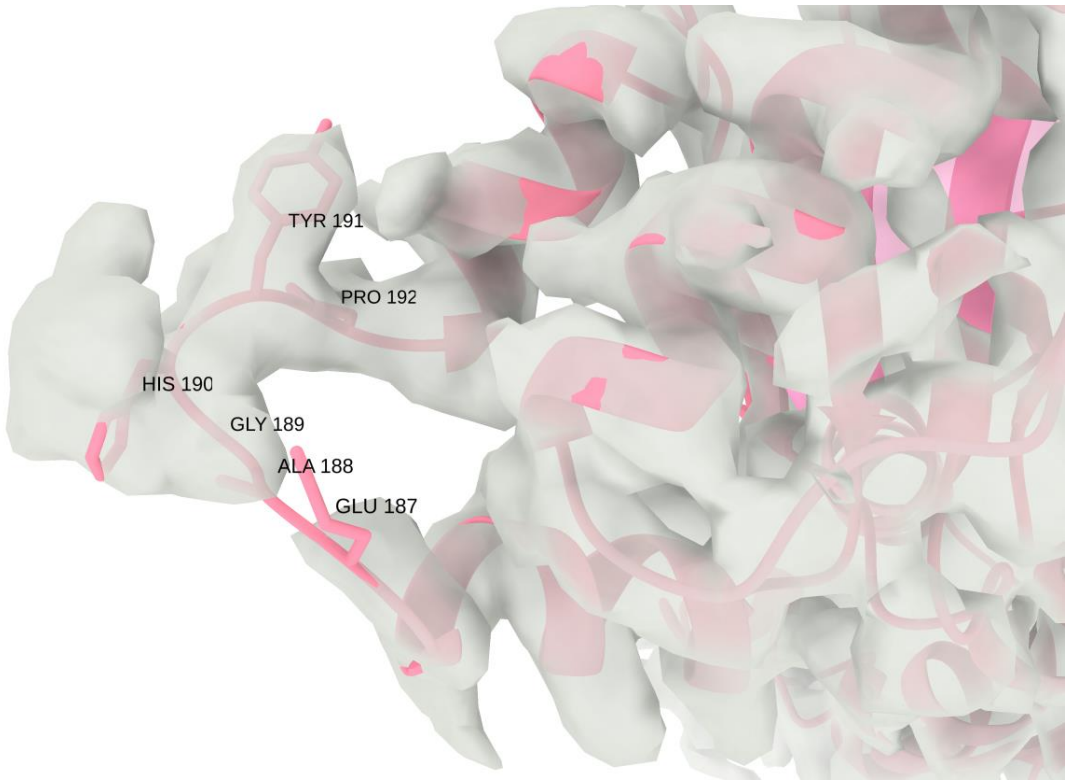
(a)

Figure 4.6. The rosetta refined monomer model (coral) superimposed with the density map (grey) (a). The map was sharpened by phenix.auto_refine. The zoom-in view of the E loop (b) and A-domain (c) with key residues shown in stick. The side chain densities of Trp70, Leu61 and Val68 are well resolved in our density map. Although the side chain densities of the other residues were missing or partially missing, the density map is able to provide enough information to model the backbone of the E loop.

Figure 4.6 continued



(b)



(c)

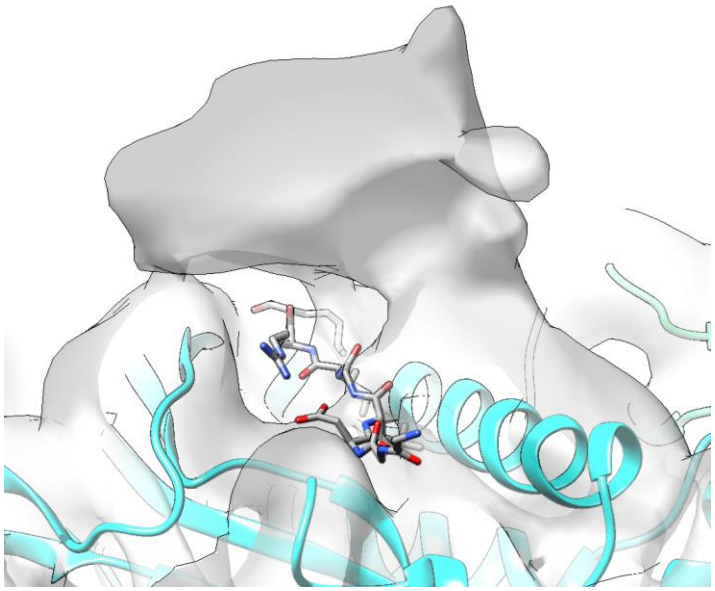


Figure 4.7. The extra density of the inner capsid in the focused refinement of the three-fold pore. Although there is no density for the cargo loading peptide (shown in stick), we can see the strong protein density connected to the inner surface of the encapsulin shell.

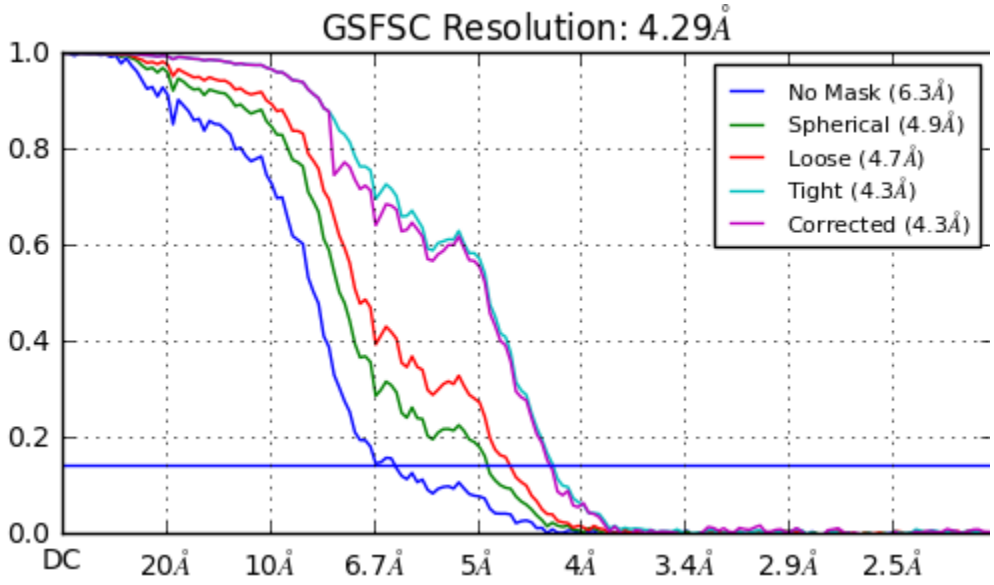


Figure 4.8. The FSC curve of C1 symmetry reconstructed map reported by cryoSPARC.

CHAPTER 5. HIGH RESOLUTION SINGLE PARTICLE CRYO-EM REFINEMENT USING JSPR

This chapter is from a publication (Sun et al., 2021) in collaboration with other scientists across different institutions. Their contributions are outlined in the journal website.

5.1 Abstract

JSPR is a single particle cryo-EM image processing and 3D reconstruction software developed in the Jiang laboratory at Purdue University. It began as a few refinement scripts for symmetric and asymmetric reconstructions of icosahedral viruses, but has grown into a comprehensive suite of tools for building ab initio reconstructions, high resolution refinements of viruses, protein complexes of arbitrary symmetries including helical tubes/filaments, and image file handling utilities. In this review, we will present examples achieved using JSPR and demonstrate recently implemented features of JSPR such as multi-aberration "alignments" and automatic optimization of masking for the assessment of map resolution using "true" FSC.

5.2 Introduction

Single particle cryo-EM has revolutionized structural biology in the past few years and become the method of choice for structural biology for a wide range of structures from large (10–100 MDa in mass and ∼200 nm in diameter), highly symmetric viruses (Dai and Hong Zhou, 2018; Fang et al., 2019) to small (sub-100 kDa), asymmetric protein or nucleotide structures (Zhang et al., 2019), and helical polymers (Fitzpatrick et al., 2017; Wang et al., 2019). Due to their size, symmetry, and higher signal to noise ratios (SNR) from the stronger electron scattering, icosahedral viruses were among the very few samples that could reach high resolution with cryo-EM (Jiang et al., 2008; Yu et al., 2008; Zhang et al., 2008) before the resolution revolution brought by the development of direct electron detectors. JSPR was initially developed for structure determination of viruses at Purdue University. This includes both symmetric reconstructions of the capsid shell with icosahedral symmetry (Chen et al., 2011; Guo et al., 2014; Jiang et al., 2008) and asymmetric reconstructions of the non-icosahedral structural components of the virus particles such as the phage tail and portal vertex (Guo et al., 2013; Jiang et al., 2006). It began as a few refinement scripts based on EMAN (Ludtke et al., 1999) and EMAN2 (Tang et al., 2007) libraries

but has grown into a comprehensive suite of tools for building *ab initio* reconstructions, high resolution refinements of viruses, protein complexes of arbitrary symmetries including helical tubes/filaments, and image file handling utilities. The general workflow from particle picking, CTF fitting, *ab initio* reconstruction using the random model approach, initial orientation determination using a consensus voting criterion, 3D reconstruction, and iterative refinements was previously detailed (Guo and Jiang, 2014). In this review, we will focus on high resolution refinement functions in JSPR, and useful image handling and parameter manipulation tools used to help translate data among various software packages, such as RELION (Scheres, 2012), and cryoSPARC (Punjani et al., 2017).

5.3 Generalized multi-aberration 2D alignment in addition to Euler angles and center positions

In single particle cryo-EM image processing, the 2D alignment step of the iterative refinement loop typically included only either a global search, or local refinement of particle Euler angles and center positions. The contrast transfer function (CTF) parameters of the particles, including defocus, astigmatism magnitude and angles, and phase shifts (if a phase plate was used), are pre-determined using the whole micrograph power spectra or smaller patches around the particles. This workflow is based on an implicit assumption that CTF fitting results were infinitely accurate. This assumption is obviously invalid or at least suboptimal as the power spectra-based CTF fitting methods have discarded a majority of the image information, such as phase. To more accurately determine CTF parameters and other aberration parameters, we generalized the Euler/center-only 2D alignment to also "align" these parameters as part of the iterative refinement process (Fig. 5.1A) (Guo and Jiang, 2014), which can take advantage of all the information of the particle images and the iteratively improved reference map quality.

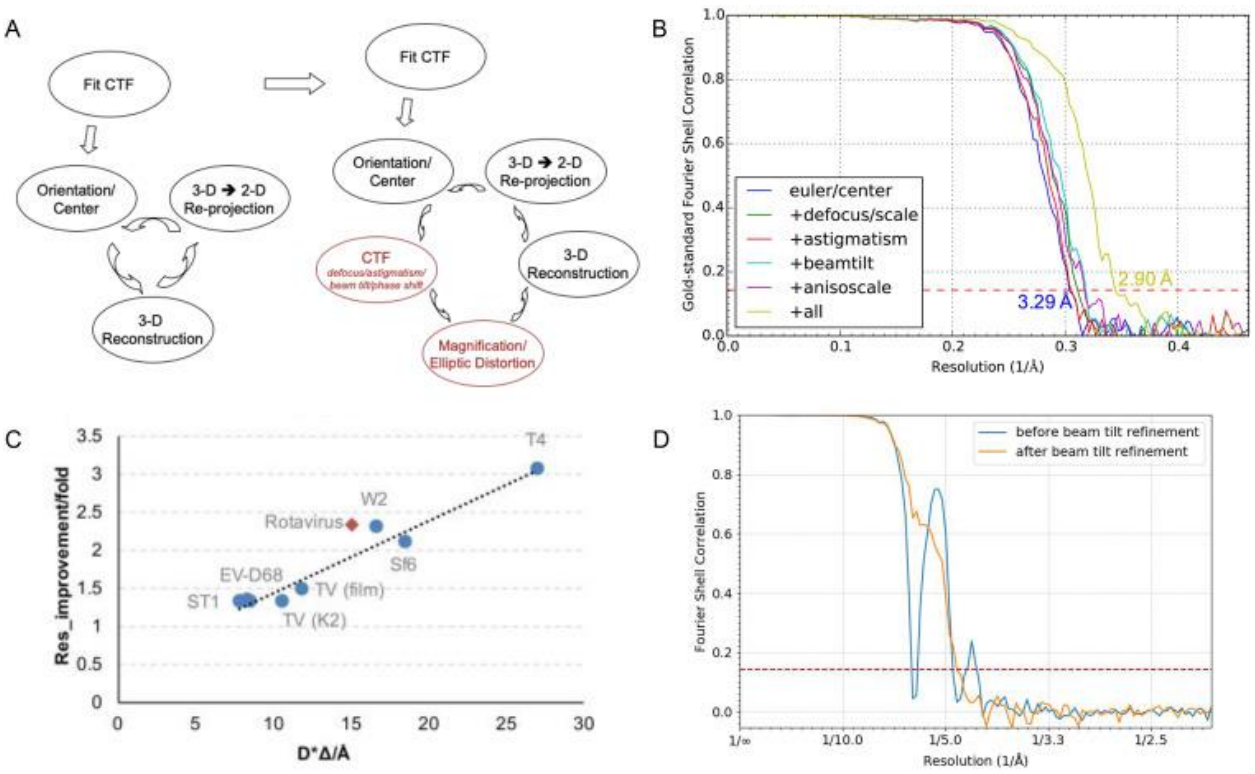


Figure 5.1. Generalized multi-aberration 2D alignment in JSPR. (A) The iterative refinement loop "aligns" multiple CTF parameters and geometric parameters in addition to the particle Euler angles and center positions. (B) Multi-aberration refinements improved the PCV2 structure from 3.3 Å to 2.9 Å resolution (Liu et al., 2016). The "anisoscale" parameter in the legend means elliptic distortion correction. (C) Correction of elliptic distortion could significantly improve the resolution of large viruses (Yu et al., 2016b). (D) The beam tilt induced coma aberrations can be corrected as indicated by the reduction of the large oscillations in the FSC curve (Li et al., 2019). (B, C, D) have been reproduced with the permission of the corresponding publishers.

We previously implemented different "aligners", where each aligner is refining one (for example, defocus or phase shift), or a few closely related parameters (such as astigmatism amplitude/angle or elliptic distortion amplitude/angle) (Guo and Jiang, 2014). The different aligners can be combined in user defined order during 2D alignment. More recently, a new aligner, *refineAll*, was developed to allow simultaneous refinement of all or a subset of the parameters, which improved alignment quality and also eliminated the need for the user to find the best order of sequential alignments. The parameters that *refineAll* supports include Euler angle, center, defocus, astigmatism, phase shift, beam tilt, spherical aberration, pixel size, and elliptic distortion. Users can easily turn on/off the refinement of any parameters and choose the refinement at the per particle or per micrograph level.

The refinement of aberration parameters has been shown not only to noticeably improve the reconstruction resolution, but sometimes essential to reach near-atomic resolution (4 Å and better) when aberrations are large. For example, the porcine circovirus 2 (PCV2) structure (Liu et al., 2016) was limited to 3.3 Å with only Euler/center refinement, but improved to 2.9 Å with aberration refinement/correction (Fig. 5.1B). While individual aberration correction could slightly improve the resolution, the holistic effect of multi-aberration corrections led to improvements in resolution (Fig. 5.1B) and side-chain densities.

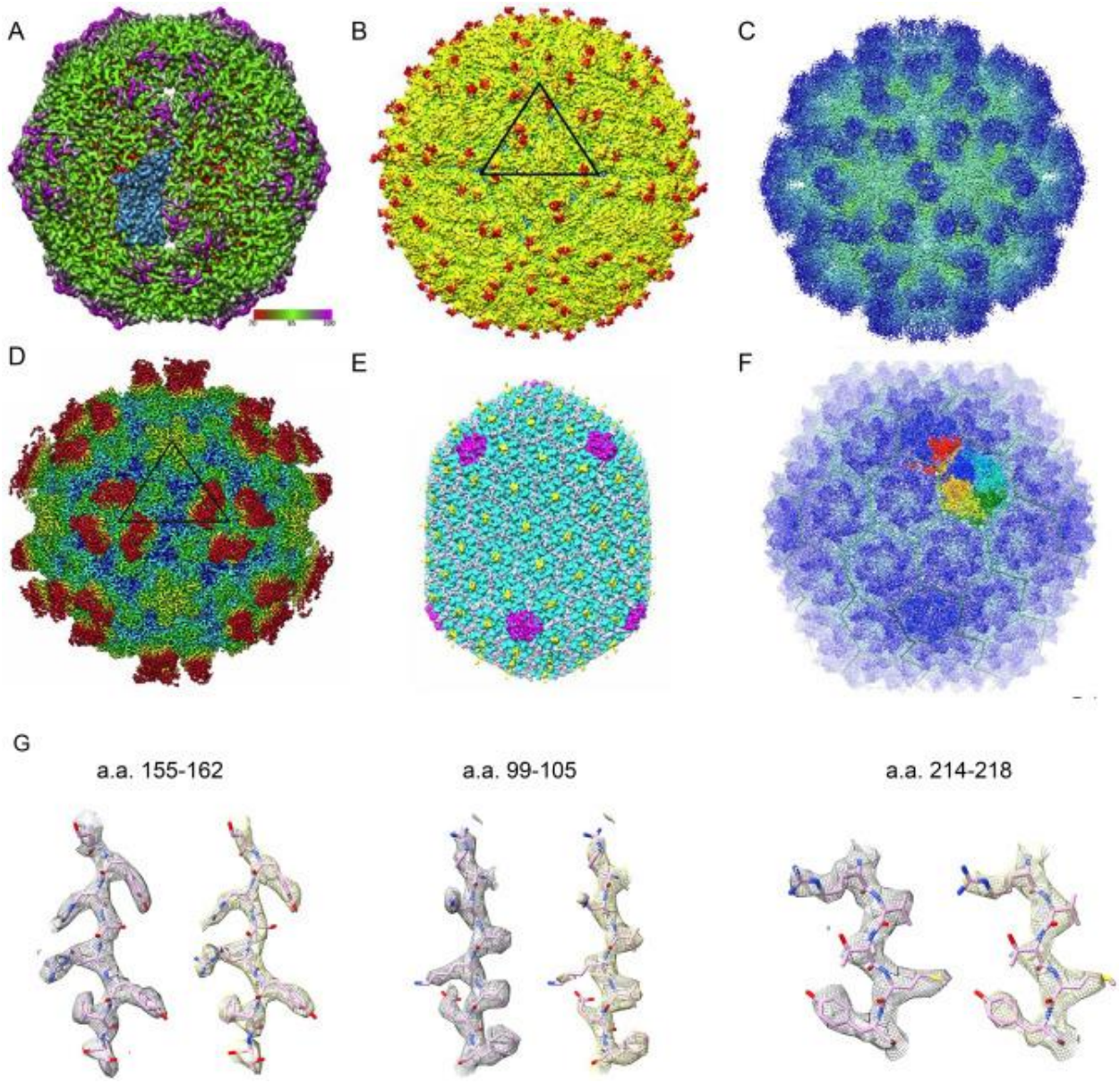


Figure 5.2. Gallery of high resolution virus structures reconstructed with JSPR. (A) 2.9 Å PCV2 structure (EMD-6555) using images recorded on photographic film (Liu et al., 2016). (B) 3.8 Å Zika virus structure (EMD-8116) (Sirohi et al., 2016). (C) 2.6 Å Tulane virus (EMD-8252) (Yu et al., 2016a). (D) 2.3 Å human rhinovirus B14 and antibody complex (EMD-8762) (Dong et al., 2017). (E) 3.3 Å bacteriophage T4 isometric head (EMD-8661) (Chen et al., 2017). (F) 2.9 Å bacteriophage Sf6 (EMD-8314) (Zhao et al., 2017). (G) The comparison of the side chain densities of PCV2 in maps with all the refinement parameters (left, grey mesh) and only the Euler/center parameters (right, yellow mesh) in three regions, a.a. 155–162, a.a. 99–105, and a.a. 214–218, superimposed with the PCV2 model (PDB:3JCI). The two density maps have been sharpened to the same level and displayed at the same contour level using *ChimeraX* (Goddard et al., 2018). (A–F) have been reproduced with the permission of the corresponding publishers.

A perfect imaging system would be free of aberrations, but in practice, all TEMs have residual aberrations either due to the fundamental limit of the magnetic lens optics, or misalignment of the lens and lens correctors (Hawkes, 2015). Elliptical distortion is a lens distortion caused by differing magnetic field strengths in the projection lenses in various directions. While TEM vendors generally correct the distortions for high magnifications (>40k) traditionally used for high resolution imaging, lower magnifications (<30k), often used with a K2 direct electron detector due to its small pixel size (5 μm), were found to have ∼2–3% elliptic distortion (Yu et al., 2016b). While low amounts of elliptical distortion are ignorable for studies of smaller protein targets, elliptical distortion correction for larger targets, such as viruses (Fig. 5.1C), is essential (Yu et al., 2016b). For example, the T4 head structure was limited to ∼7 Å resolution, but was improved to 3.3 Å resolution after elliptical distortion correction (Chen et al., 2017). The distortion parameters were determined as part of the 2D alignment, eliminating the need for pre-calibration which might not be feasible after data collection.

Beam tilt induced coma aberrations were first found to be significant in processing 2D crystal images (Henderson et al., 1986), but has now also been recognized as an important parameter in refinement for high resolution single particle cryo-EM (Glaeser et al., 2011). In our recent study on Volta phase plate imaging, we have collected a dataset with severe beam tilt that was accidentally introduced during data collection, likely because of an unknown software glitch (Li et al., 2019). Without beam tilt correction, the reconstruction was limited to ∼6 Å resolution and the Fourier shell correlation (FSC) curve had multiple large oscillations (Fig. 5.1D). We quickly identified the large beam tilt issue with the *refineAll* aligner in JSPR after we observed the reduction of oscillations in the FSC curve post beam tilt refinement/correction (Fig. 5.1D). The final resolution of 2.5 Å was achieved after additional multi-aberration refinement/correction (Li et al., 2019).

5.4 High resolution structures determined using JSPR

Most of the initial work with JSPR was done with virus reconstruction. As the scope of the projects began to expand beyond viruses, JSPR has been used to solve the structures of protein complexes and helical polymers. A few structures determined using JSPR will be briefly described here.

Viruses. Fig. 5.2 shows a gallery of virus structures solved with JSPR in recent years. The 2.9 Å PCV2 structure (Fig. 5.2 A) (Liu et al., 2016) has demonstrated the possibility of high resolution reconstruction with close-packed particles. It is also the first and only sub-3 Å structure (Fig. 5.1 B) using photographic film. The 3.8 Å Zika virus structure (Fig. 5.2 B) (Sirohi et al., 2016) provides a foundation for analysis of the antigenicity and pathogenesis of Zika virus. The 2.6 Å Tulane virus structure (Fig. 5.2 C) (Yu et al., 2016a) demonstrated the feasibility of achieving high resolution reconstructions using antibody-based affinity grids to enrich low concentration particles on the EM grid. The 2.3 Å structure of human rhinovirus B14 with C5 antibody Fab (Fig. 5.2 D) (Dong et al., 2017) depicts the highest resolution of a virus-antibody complex. The 3.3 Å T4 phage head structure (Chen et al., 2017) (Fig. 5.2 E) highlighted the severe negative impact of elliptic distortion on large structures and the dramatic improvement of the resolution from ~7 Å to 3.3 Å by elliptic distortion correction using JSPR. The 2.9 Å bacteriophage Sf6 structure (Fig. 5.2 F) (Zhao et al., 2017) was the first tailed phage structure determined to sub-3 Å resolution.

Protein complexes. It is sometimes misunderstood that JSPR only supports icosahedral viruses, probably due to our previous work mainly featuring viruses. However, JSPR supports all point-group symmetries including all platonic solids (I, O, and T symmetries), dihedral symmetries (D), cyclic symmetries (C), and no symmetry. Fig. 5.3 A shows the 2.7 Å structure of the T20S proteasome submitted to the 2016 map challenge (emcd108). The two JSPR maps submitted to the map challenge (this T20S structure and emcd132 for GroEL) were assessed among the best density maps (Heymann et al., 2018; Marabini et al., 2018; Pintilie and Chiu, 2018) and the assessors included comments such as, "JSPR is always among the best" (Marabini et al., 2018). In a recent study using the Volta phase plate, a 2.5 Å apoferritin structure (Li et al., 2019) (Fig. 5.3 B) was obtained with JSPR after refining and correcting large beam tilts accidentally introduced during data collection (Fig. 5.1 D). We have also refined the asymmetric ribosome structure, EMPIAR-10107 (Desai et al., 2017), to ~3 Å with JSPR (Fig. 5.3C). These examples demonstrated the capability of JSPR for reconstructing non-virus structures.

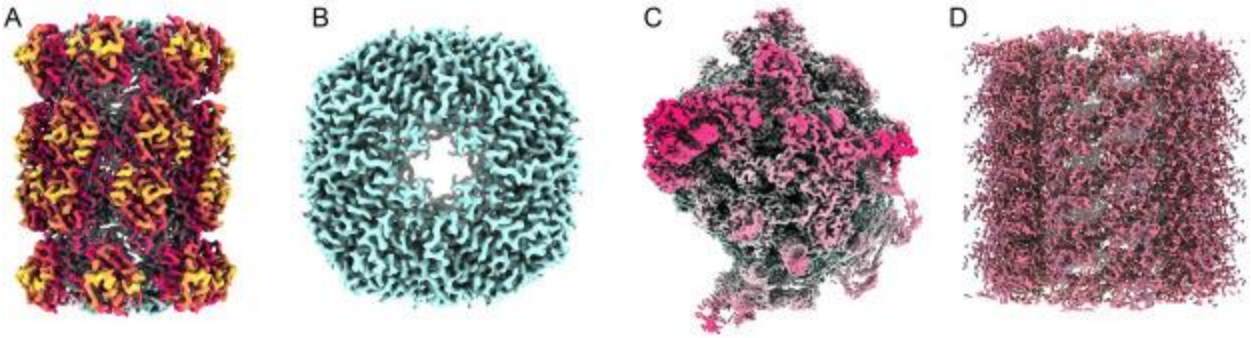


Figure 5.3. Gallery of high resolution non-virus structures reconstructed with JSPR. (A) 2.7 Å T20S proteasome structure submitted to the 2016 Cryo-EM Map Challenge (emcd108). (B) 2.5 Å apoferritin structure from Volta phase plate data (Li et al., 2019). (C) 3 Å ribosome structure solved using the EMPIAR-10107 dataset (Desai et al., 2017). (D) 3 Å VipA/VipB helical structure reconstructed from the EMPIAR-10019 dataset (Kudryashev et al., 2015).

Helical polymers. Support for helical structures using the Iterative Helical Real Space Reconstruction (IHRSR) strategy (Egelman, 2007) was recently added to JSPR. The "gold standard" split of a dataset is done at the micrograph or filament/tube level instead of the particle (i.e. helical segment) level to avoid inflated FSCs due to the same region of pixels in neighboring but overlapping segments being separated into different subsets. Fig. 5.3D shows a ~3 Å structure of the VipA/VipB protomer Type VI secretion system contractile sheath (EMPIAR-10019) (Kudryashev et al., 2015) with both *de novo* indexing and refinement by JSPR. The new helical indexing and classification method will be published separately.

5.5 Utilities in JSPR

JSPR is currently a CPU-based image processing software that runs on Linux-based computing environments, an ad hoc collection of workstations, CPU cycle-scavenging HTCondor (Jiang et al., 2008), and dedicated Linux clusters managed by a job queue system, such as PBS or Slurm. It has been programmed in a way that whenever an unexpected stop occurs, it is able to automatically resume at the last stopping point. This is realized by dividing large tasks, for example, the 2D alignment, into batches and tracking the completion state of each batch.

In addition to the iterative refinement function, JSPR also includes a rich set of utilities that facilitate analysis of the refinement results and interchanges of results with other image processing software. Here we only briefly describe three such utilities while a more comprehensive list of

utilities is available in the software that is freely downloadable from the corresponding author's website (https://jiang.bio.purdue.edu/jspr).

images2lst.py and *images2star.py***.** Both programs were initially designed to convert files from different software packages, for example, cisTEM's database (*∗.db*) file or cryoSPARC's (*∗.cs*) file, into *lst* file used by JSPR or *star* file used by RELION, and to allow interchanging between JSPR and RELION. Later, they were enhanced with many additional functions. For example, it allows users to easily check/remove duplicate particles, add parameters to a *lst/star* file, copy parameters from another *lst/star* file, perform subset selection based on the desired range of any parameters, perform symmetry expansion or the reverse, remove excessive number of particles in the dominant views, subtract projections with real space linear scaling of pixel levels or Fourier space per resolution shell amplitude scaling, or virtually shift reconstruction center position for localized reconstruction of arbitrary sizes without the need to re-extract/clip the particles into another dataset, *etc*.

*plotDist.py***.** This script allows the user to plot the histogram of arbitrary parameters or the correlation of any pair of parameters in a *lst* or *star* file and the plot can be interactively examined or saved into a PDF file. These plots are useful in evaluating data (e.g., distribution of defocus values) or the refinement behavior (e.g., distribution of Euler angle changes or how Euler angle changes are correlated to defocus value changes, *etc.*).

trueFSC.py. Although FSC has been widely used for the estimation of the cryo-EM map resolution after its first introduction (Harauz and van Heel, 1986), how to properly mask out the background (i.e. solvent flattening) is often a dilemma as the FSC tends to report different resolution values with different levels of masking. If the mask is too tight, the resolution will be overestimated because of the mask correlation at high resolution. If the mask is too loose, the resolution will be underestimated, which means the map quality is better than the reported resolution. An optimal mask should contain all protein pixels, exclude background/noise pixels as much as possible, and use a soft transition layer (i.e. tapering) of appropriate slope between the protein and background regions. The masking is typically based on a thresholding method for which the quality of masking is critically dependent on the threshold value. Current software still requires the user to manually input the threshold for mask generation, which can be tedious and often leads to suboptimal masking. To address this problem, we developed a python script, *trueFSC.py,* which can automatically find the best threshold and mask slope for reliable

resolution estimation. The user only needs to specify the two half maps. A plot of the FSC curves in PDF format and two masks for the half maps will be generated. There are two different methods for automatic threshold determination that have been implemented. The preferred approach is to specify the mass of the protein complex, which is used to calculate a threshold to make the total volume of all pixels with larger values equal to the expected volume assuming an average protein density (1.35 g/cm3) (Fischer et al., 2004). The second automatic threshold determination method is based on Otsu's thresholding method that automatically finds a cutoff value so that the variances within the two groups are minimum (Otsu, 1979). Starting from a sharp mask based on the threshold, *trueFSC.py* will then extend it with a soft slope (raised cosine, 1+cos(r)2 in which r is the total width of the slope) at an optimal slope width that maximizes the resolution without inflating the FSC of phase randomized maps (Chen et al., 2013). Fig. 5.4A shows an example of a *trueFSC.py* generated FSC and the corresponding soft mask. As a comparison, a large spherical mask will underestimate the resolution (Fig. 5.4B) while an overly-tight mask will inflate the resolution, as indicated by the inflated FSC of phase randomized maps using the same tight mask (Fig. 5.4 C). We expect this feature will provide a more convenient and accurate FSC calculation, plotting, and resolution reporting tool that can benefit the cryo-EM community.

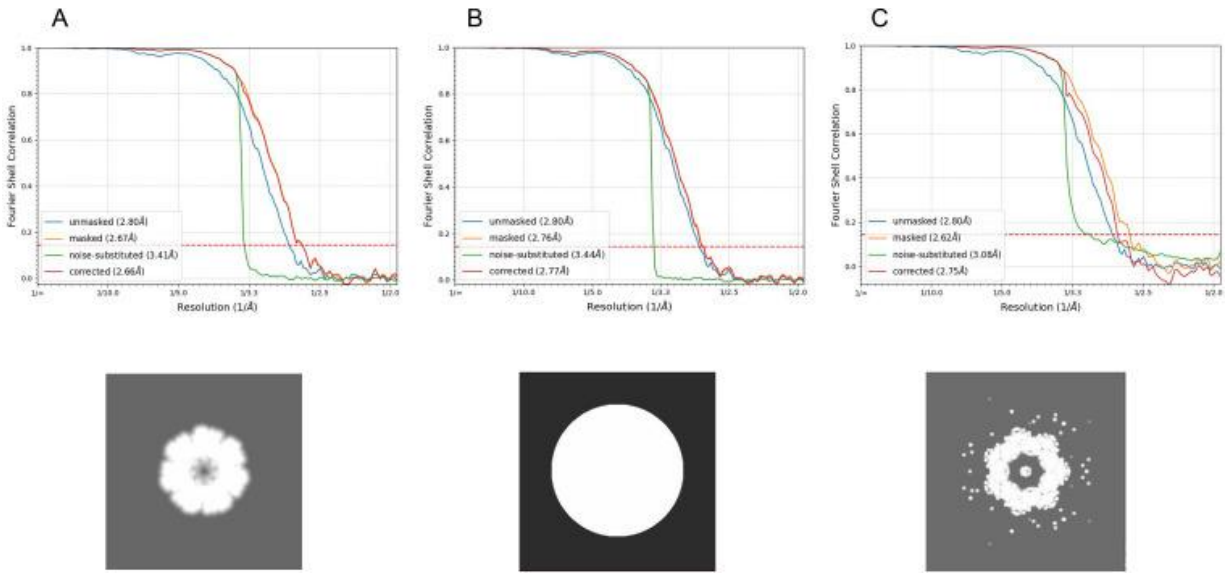


Figure 5.4. Examples of different levels of masking and their effect on FSC curves. (A) Reliable resolution estimate using the optimal mask automatically determined by *trueFSC.py*. (B) Large spherical mask underestimates the resolution. (C) Overly tight mask inflates the resolution if the FSC = 0.143 criterion is applied directly to the FSC of masked maps (2.62 Å). The large FSC values beyond the phase randomization cutoff resolution (~3.4 Å) in the noise-substituted map FSC curve (green) provide clear signs that the mask is too tight.

5.6 Conclusion

The highlight of JSPR is the refinement and correction of multiple aberrations, including CTF parameters and geometric scaling/distortions, which have slowly been recognized in the field. We are happy to see other software packages, for example, cryoSPARC 2.12 and RELION 3.1 (Zivanov et al., 2019), now also include similar functions in their recent updates. Currently, high resolution refinement in JSPR is limited by the homogeneity of the sample. Future developments will need to address the heterogenous states, especially the continuously varying conformations due to the intrinsic dynamics of the structure. While a definite solution is still to be found, an attractive concept is to picture the distribution of particles as a lower-dimension manifold embedded in the hyperspace with dimensions equal to the number of pixels in a particle image (Frank and Ourmazd, 2016). While the manifold for a homogenous set of particles will have only three dimensions for the particles views, the manifold for heterogeneous particles will have additional dimensions to represent the conformations with the number of extra-dimensions ranging from small for simple one or few hinge motions to large for continuously varying conformations throughout the structure. Algorithms based on either classic machine learning methods (Frank and Ourmazd, 2016) or the more recent deep neural networks approach (Zhong et al., 2019) can be explored to learn the manifold.

CHAPTER 6. HELICAL INDEXING IN REAL SPACE WITHOUT THE NEED FOR FOURIER LAYER LINES

This chapter is from a publication (in press) in collaboration with other scientists. Their contributions would be outlined in the journal website.

6.1 Abstract

Biological structures with helical symmetries of distinct twist, rise, and axial symmetry are abundant and span a wide range of organisms and functions. Performing de novo helical indexing remains challenging because of the steep learning curve involved in Fourier space layer lines. The unknown amount of out-of-plane tilt and the existence of multiple conformations of the helices further complicate indexing. In this work, we introduce a real-space indexing method that leverages the prior knowledge of the tilt and in-plane angles of the helical filaments/tubes, robust ab initio 3D reconstruction capabilities in single particle cryo-EM to obtain asymmetric reconstructions, and automatic indexing of helical parameters directly from the asymmetric density maps. We validated this approach using data from multiple helical structures of distinct helical symmetries, diameters, flexibility, data qualities, and heterogeneous states. The fully automated tool we introduce for real space indexing, HI3D, uses the 2D lattice in the autocorrelation of the cylindrical projection of a 3D density map to identify the helical symmetry. HI3D can often successfully determine the helical parameters of a suboptimal 3D density map with intermediate evidence that can also help assess the map quality. Furthermore, this tool is usable independently as a Web application that can be accessed free of installation. With these methods, de novo helical indexing will be significantly more accessible to researchers investigating structures of helical filaments/tubes using cryo-EM.

6.2 Introduction

Nature has adopted helical arrangements for many important biological structures, from viruses to cellular components. Notable examples include double-stranded DNA, actin, microtubule, bacterial flagella and pili, filamentous viruses such as TMV and Ebola, and phage tails. In recent years, increasing evidence demonstrates that many cellular proteins undergo conformational changes to form helical filaments either in a regulated process, such as human

acetyl-CoA carboxylase (Lynch et al., 2020), or in pathological conditions, such as distinct Tau protein filaments in Alzheimer's diseases and other neurodegenerative diseases (Berriman et al., 2003; Kidd, 1963; Terry, 1963). Helical structures were among the initial targets of 3-D electron microscopy in 1968 (De Rosier and Klug, 1968) and many helical structures have been solved to near-atomic resolutions since 2003 using the classic Fourier-Bessel method (Yonekura et al., 2003), and Iterative Helical Real Space Refinement (IHRSR) method (Egelman, 2000). Currently, RELION (He and Scheres, 2017) is the most frequently used program for helical reconstruction for its superior performance over other software.

A helix is a 1D crystallographic arrangement of asymmetrical units that can be described in rise and twist. Rise is the distance along the helical axis between the two neighboring asymmetric units. The twist is defined as the rotation angle around the helical axis between the neighboring asymmetric units. The helical symmetry parameters are traditionally determined by indexing the layer line patterns in the 2D power spectra of helical particle images, using the same principle for indexing as was used for the "Photo 51", the X-ray diffraction of dsDNA (Franklin and Gosling, 1953) that led to the discovery of the double-helix structure. In this method, the correct orders of Fourier-Bessel functions are manually assigned to layer lines, which is not only mathematically difficult for many users to master but also intrinsically demanding on the quality of the data. A critical step of Fourier layer line based indexing is identifying and extracting a long, straight, and well-ordered helical filament/tube in the micrograph to produce clearly defined layer lines in Fourier space. However, helical structures with a high degree of curvature and heterogeneous conformations produce poorly defined layer lines, making Fourier Bessell indexing impossible. Only when clear helical layer lines are seen in Fourier space, the helical twist and rise can be derived, and the final structure can be refined with the helical symmetry estimation. However, helical indexing could still be ambiguous and error-prone even when strong layer lines are available, as the unknown amount of out-of-plane tilt further complicates the layer line pattern (Egelman, 2014). For these reasons, incorrect helical indexing has been found in several reports (Egelman, 2014; Egelman and Wang, 2021). The IHRSR method was developed (Egelman, 2000) to overcome the challenges of the classic Fourier-Bessel reconstruction method with flexible, curved helical structures by treating a helical filament/tube as a series of overlapping segments (i.e. single particles) of which the orientation/center parameters could be individually refined. Although this method could also automatically refine the helical parameters around an initial value, it could

only converge to the correct helical parameters if the initial rise/twist values were close to the correct values. In reality, it is hard to find an accurate initial value, especially when dealing with a new system without having any prior knowledge of it.

Herein, we introduce a new method to determine helical symmetry in real-space. Using existing single-particle reconstruction (SPR) image processing tools, we have devised a strategy to obtain 3D reconstructions of helices without knowledge of helical parameters to estimate the helical symmetry in real space using HI3D (Helical Indexing using the cylindrical projection of a 3D map). This method is based on 2D lattice indexing using the auto-correlation function (ACF) on the cylindrical projection of a 3D map. We have demonstrated that this approach can yield accurate helical symmetry parameters for multiple experimental datasets. The samples of these datasets vary in function, size, symmetry, flexibility, heterogeneity, and data quality. The easy access of the HI3D method as a Web app will also help one better understand helical symmetry, how helical symmetry and its underlying 2D crystal lattice are related. We believe this method is easier to understand and use than the Fourier layer line method, and it is openly accessible online, significantly reducing the difficulty in determining de novo helical structures.

6.3 Materials and Methods

6.3.1 Test datasets

Among the four test datasets, three datasets were downloaded from EMPIAR, and the HIV tube dataset was kindly provided by Dr. Peijun Zhang (Table 6.1). For the TMV dataset, segments were extracted from the micrographs with the provided coordinate file. For the MAVS CARD dataset, segments were picked with cryoSPARC's reference-free filament tracer (Punjani et al., 2017). The reference-free 2D classification was performed for all extracted segments. All good quality 2D classes were selected and kept. For the VipA/VipB dataset, segments were first picked with the template-free cryoSPARC filament tracer. Good looking 2D classes were selected for template-guided filament tracer. Then, 2D classification was performed in cryoSPARC and good quality 2D classes were kept. For the HIV tube dataset, all filaments were manually picked and extracted with RELION/3.1 (Scheres, 2012; Zivanov et al., 2018). As above, cryoSPARC was used for reference-free 2D classification. For all datasets, the selected, good segments were subject

to ab initio asymmetric reconstruction usingRELION/3.1. The resulting maps were uploaded to HI3D for automated helical indexing.

6.3.2 Ab initio asymmetric reconstruction with constrained Euler angles

After extracting helical segments and using 2D classification to remove the poor-quality segments, the Euler angle (ϕ, ψ, θ) search-constrained ab initio reconstruction was performed with RELION/3.1 using Class3D with a bare cylinder as the initial model. The tilt angle search range was constrained to 90±15° given the prior knowledge that the helical axis should be nearly parallel to the image plane. The in-plane rotation angle psi is also constrained (±5°) around the angle derived from filament selection. The reconstructed maps were generated without imposing any point or helical symmetry. These reconstructed maps were then uploaded to HI3D to determine the helical parameters.

6.3.3 Implementation and availability of HI3D Web app

We have used streamlit (https://github.com/streamlit/streamlit), which is open-source software for the development of Python-based Web apps, to implement the HI3D Web app. The hosted HI3D app and the source code will be freely accessible from the authors' Web site (http://jiang.bio.purdue.edu/HI3D). The trackpy (https://github.com/soft-matter/trackpy) library was used for automated detection of peaks (i.e. "diffraction spots") in the ACF image of the cylindrical projection.

6.4 Results

6.4.1 Real space helical symmetry estimation with HI3D web app

HI3D was developed to overcome the above mentioned issues with Fourier layer line based helical indexing. It only requires an input 3D map to automatically output the helical rise and twist parameters. The internal configuration of HI3D is described in Fig. 6.1. Since it was designed for 3D maps reconstructed without symmetry that is arbitrarily positioned/oriented (Fig. 6.1A), HI3D will first find the center of the density map and shift it to the center of the box and vertically align it along the Z-axis (Fig. 6.1B). Then, the cylindrical projection of the 3D map is obtained by

resampling the map in cylindrical coordinates and summing along the radial direction (Fig. 6.1C). This cylindrical projection is equivalent to unwrapping a helical structure into its corresponding 2D crystal. If the input asymmetric map has clear helical structure features, the cylindrical projection appears as a well-ordered 2D crystal (Fig. 6.1C). Conversely, the cylindrical projection of an asymmetric structure would not produce a 2D lattice organization. Thus, the apparent crystalline order of the cylindrical projection is a useful diagnostic of the presence and quality of the helical structure features in the asymmetric reconstruction.

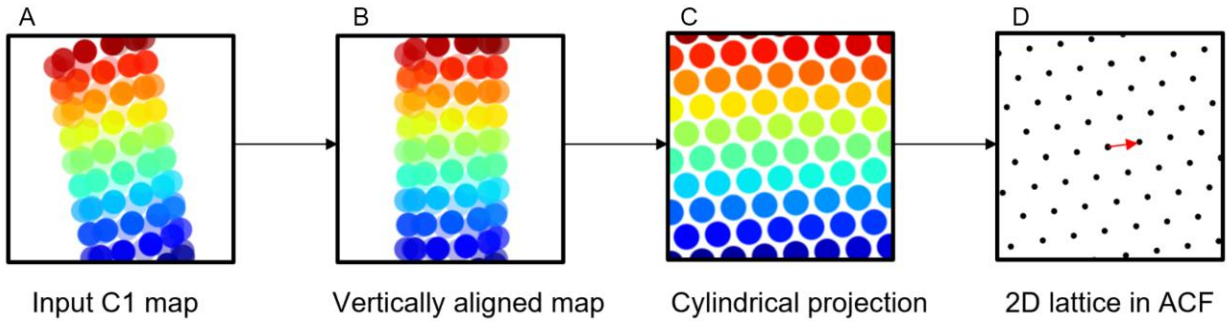


Figure 6.1. HI3D workflow. The input asymmetric density map that is arbitrarily positioned and orientated (A) is automatically centered and vertically aligned (B). The aligned 3D map is resampled in the cylindrical coordinate to generate the cylindrical projection (C) to mathematically convert a helical structure in the original 3D map into a "2D crystal" image. The auto-correlation function of the cylindrical projection would generate a 2D lattice (D) that visually resembles the diffraction spots of a crystal. The unit cell vector (red arrow in D) with the shortest distance to the equator would correspond to the helical twist (x-coordinate) and rise (y-coordinate).

The task of identifying helical parameters in the original asymmetric reconstruction is now mathematically converted to an easier, more intuitive task of indexing the 2D crystal of the cylindrical projection. To facilitate automatic indexing of the 2D crystal, HI3D then computes the auto-correlation function (ACF) of the cylindrical projection (Fig. 6.1D), which results in a 2D lattice of much sharper "diffraction spots" with the quality (i.e. sharpness) of the spots being strongly correlated to the quality of the helical organization. In the ACF, the x-dimension corresponds to the azimuthal angle around the helical axis while the y-dimension is defined as the shift along the helical axis. Users can use the mouse to interactively index the helical parameters by hovering the mouse over the "diffraction spot" closest to the equator of the ACF image to display the corresponding helical twist (x-coordinate) and rise (y-coordinate).

HI3D has also implemented an automated 2D lattice indexing function to not only eliminate the need for interactive indexing, but also improve the accuracy of helical twist and rise estimates by least-square fitting of all "spots". The automated 2D lattice indexing was achieved by two different methods. The first is to treat the detected "diffraction spots" on a generic 2D lattice without considering its origin from the helical lattice and to find the two-unit cell vectors that best fit the "spot" positions. The unit cell vector with the smallest distance to the equator will be chosen to report the helical twist (x-coordinate) and rise (y-coordinate). The second method considers the 2D lattice as a thin slab of lattice unwrapped from the helical lattice and identifies the rise and twist in sequential order. First, it sorts the Y-coordinates of all "spots", calculates the distance between neighbor values, and takes the most frequent, non-zero spacing as the helical rise. Second, it sorts the X-coordinates of the "spots" in neighboring rows (i.e. with Y-coordinates separated by helical rise), calculates the spacings between neighbor values, and takes the most frequent, non-zero value as the helical twist. To ensure the resulting helical twist/rise parameters are reliable, HI3D uses both methods to generate the rise and twist and requires that the difference between the results of the two methods is less than a threshold, for example, 1 Å in rise and 1° in twist. When the density map has clear helical structure features, the two methods reliably generate consistent helical twist and rise parameters. However, the twist and rise values from these two methods tend to diverge when the quality of the helical structure becomes worse. HI3D can alert the user about inconsistency among helical parameter estimates. In both methods, the C-symmetry of the helix can be calculated as the number of equally spaced peaks on the equator or equivalently as Csym=360°/twist if Csym is an integer. To further improve the accuracy of the helical twist and rise parameters, a local optimization of both parameters is performed to sub-pixel accuracy by maximizing the sum of ACF values at the "diffraction spots" positions expected by the helical twist/rise.

HI3D has been set up as a Web app (Fig. 6.2) which is convenient to access using an internet browser, without the need for installation. In addition to UI for data intake (Fig. 6.2A) and displaying the output helical twist/rise parameters (Fig. 6.2B), its Web UI also displays the intermediate data, such as the X/Y/Z section views, cylindrical projection, ACF with detected "spots" (Fig. 6.2C), and the vector representing the twist/rise parameters (Fig. 6.2B) for diagnostic purposes. These intermediate data also clearly relate the helical lattice to the corresponding 2D lattice and how the two lattices could be interconverted through a wrapping (2D → helical lattice)

or unwrapping (helical → 2D lattice) process. Thus, HI3D can also serve as an educational tool to illustrate the basic concepts of helical structures and the formation of helical structures via wrapping a thin slab of 2D crystals.

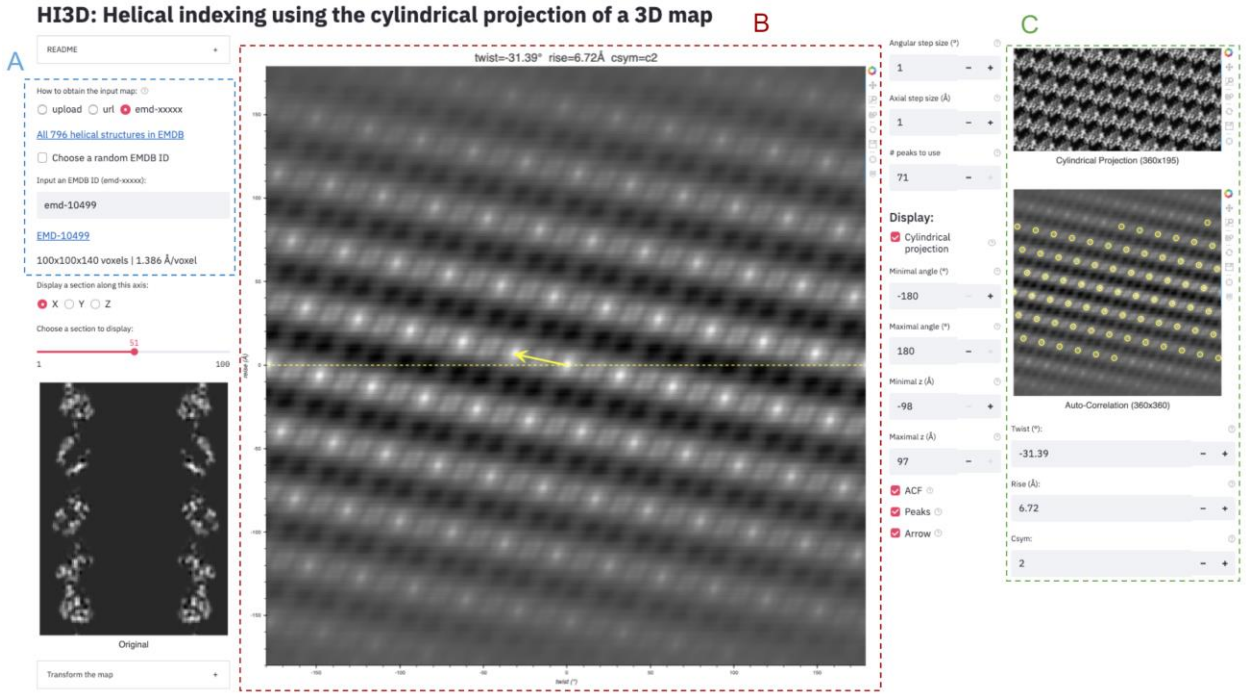


Figure 6.2. HI3D Web app user interface. The user interface of HI3D consists of three major parts. (A) The left part shows the input panel and the X/Y/Z section of the input map. Users can input a map in three ways. The first one is to upload a map from the local directory. The second is to provide the URL, for example, of a cryoSPARC output map. The third is dedicated to the helical structures in EMDB by either entering an EMDB ID or randomly choosing a helical structure in EMDB. (B) The central panel shows the output twist, rise, and C-sym values in text and as a vector centered in the ACF image. The right panel (C) shows the cylindrical projection (top), ACF of the projection image (middle), and input fields (bottom) for the user to overwrite the rise and twist if the automated detection fails.

6.4.2 Case studies of helical structures deposited in EMDB

To facilitate the validation of HI3D against a wide range of helical structures, HI3D includes an input mode to retrieve EMDB-deposited helical structures (Fig. 6.2A). A user can either input a specific EMDB ID or HI3D can randomly select one from all currently available helical structures in EMDB. HI3D keeps an up-to-date list of all helical structures in EMDB by automatically querying EMDB for the complete list of all helical structures. HI3D performs well

in tests with the EMDB helical structures as shown in the three examples of distinct cases, EMD-23871 (Tau paired helical filament extracted from PrP-CAA Patient brain tissue (Hallinan et al., 2021)), EMD-6179 (F-actin (Galkin et al., 2015)), and EMD-30129 (Helical stem of the cleaved double-headed nucleocapsids of Sendai virus (Zhang et al., 2021)) (Fig. 6.3), with the HI3D-reported rise and twist values almost identical to the published values despite the drastically different qualities of the lattice in the ACF images of these structures. Therefore, HI3D is a robust reporter of helical parameters for EMDB maps. Since the helical rise and twist parameters are not consistently reported across different EMDB entries or even for the same EMDB entry across the EMDB-mirroring sites, HI3D can be useful as a validation tool to complement EMDB. The convenience provided by HI3D to quickly examine many helical structures in EMDB and obtain their helical parameters also makes it a useful educational tool for illustrating helical structure and symmetry.

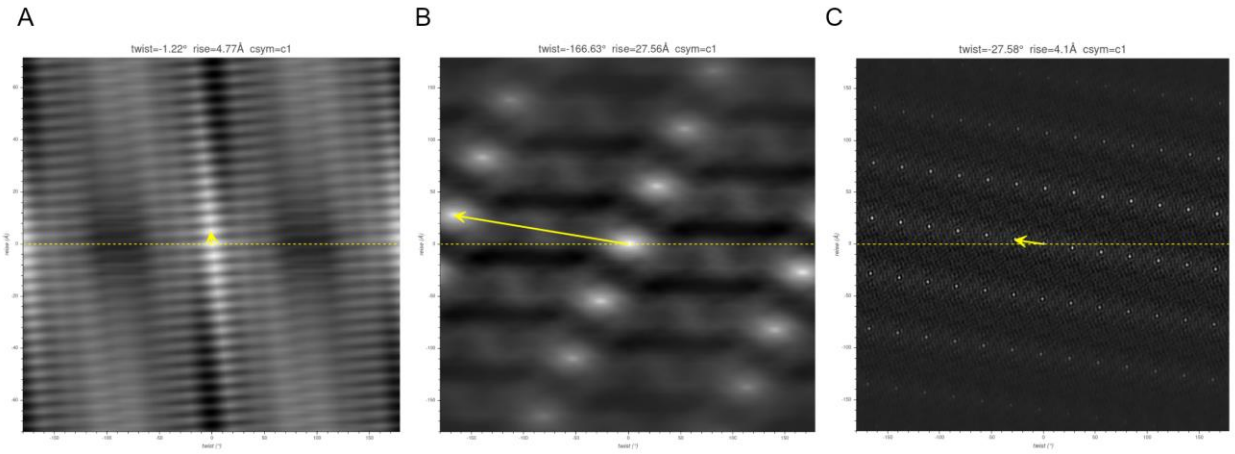


Figure 6.3. HI3D results for three helical structures in EMDB. (A) EMD-23871 (Tau paired helical filament extracted from PrP-CAA Patient brain tissue); (B) EMD-6179 (F-actin); (C) EMD-30129 (Helical stem of the cleaved double-headed nucleocapsids of Sendai virus). The published rise and twist for these three datasets are (4.77 Å, -1.2°), (27.6 Å, 166.7°), and (4.09 Å, -27.58°), respectively.

6.4.3 Case studies of ab initio asymmetric reconstructions

The ultimate goal of HI3D is to facilitate de novo determination of helical structures without the need for prior knowledge of helical symmetry that are typically obtained from Fourier layer line-based indexing, a complicated, error-prone approach. To gauge its feasibility, we tested multiple experimental datasets spanning a wide range of helical symmetries, diameters, flexibility,

data qualities, and heterogeneous states. In these tests, we did not impose any point or helical symmetry (i.e. without any symmetry specified), but took advantage of single particle reconstruction methods and constrained the orientation search for 2 of the 3 Euler angles to a small range around expected values (see Methods). Since the resulting 3D density maps were not expected to have an ideal helical structure, these tests would gauge the robustness of HI3D in its reporting of helical symmetry parameters.

TMV (Figure 6.4)

We chose the tobacco mosaic virus (TMV) as our first test case for it is the first structure reconstructed with 3DEM (De Rosier and Klug, 1968), and has been well studied as a model system throughout the years. It was also used as a robust test sample for a new algorithm in subtomogram averaging (Sanchez et al., 2020). The TMV capsid is primarily composed of a 17 kDa coat protein (CP) that is organized into a helix. In Figure 6.4, the ab initio reconstructed asymmetric map generated with Euler angle constraints already displays detailed structure features. After uploading this map to HI3D, the cylindrical projection and ACF are automatically generated (Fig. 6.4). Although the cylindrical projection visually displays an obvious 2D crystalline pattern, it is also obvious that the "crystalline" pattern has long range disorder, which explains the much weaker spots in the outer region of the ACF and also indicates that the asymmetric reconstruction was still far from ideal. Despite the suboptimal quality of this asymmetric reconstruction, HI3D reports twist and rise values close to the published values (Table 6.1). Of note, although we have downsampled the segment images eight times resulting in a pixel size (2.552 Å), which is much larger than the rise (1.408 Å), the error of HI3D output rise value is only 0.16 Å or is ~6% of the pixel size. The HI3D determined twist has an opposite sign due to the wrong handedness of the asymmetric map. Single particle reconstructions are intrinsically equivalent for both handedness and the correct handedness can only be determined with additional information, such as tilt experiments or the atomic structure details.

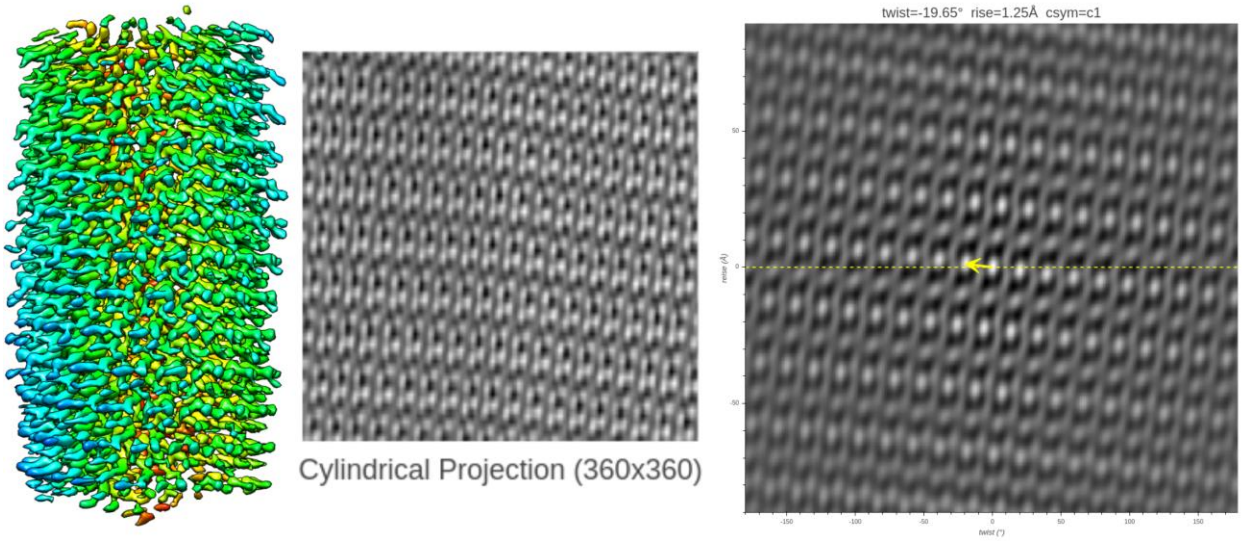


Figure 6.4. HI3D results for ab initio asymmetric reconstructions of TMV dataset (EMPIAR-1022). From left to right: surface view of the asymmetric reconstruction, cylindrical projection, HI3D output helical twist, rise and C-sym in text and as a unit cell vector in the ACF image.

MAVS-CARD (Figure 6.5)

We selected the MAVS CARD EMPIAR dataset 10031 (Xu et al., 2015) as our second case for it has been historically assigned wrong helical twist, rise, and C-symmetry (Egelman, 2014). With our method, the helical parameters can be easily determined without any ambiguity as shown by the 2D lattice of "spots" in the ACF image (Fig. 6.5). After ab initio asymmetric reconstruction, the dominant class of the three output classes has correct structural features, while the other two classes were obvious junk classes (surface view of all three classes are shown in Fig. 6.8). While the cylindrical projection clearly displays a 2D crystalline pattern, differences in the "unit cells" were also apparent (Fig. 6.5). Despite the suboptimal asymmetric reconstruction, HI3D could output rise (5.04 Å) and twist (-101.22°) values that are nearly identical to the published values (rise=5.088Å, twist=-101.436°) (Table 6.1).

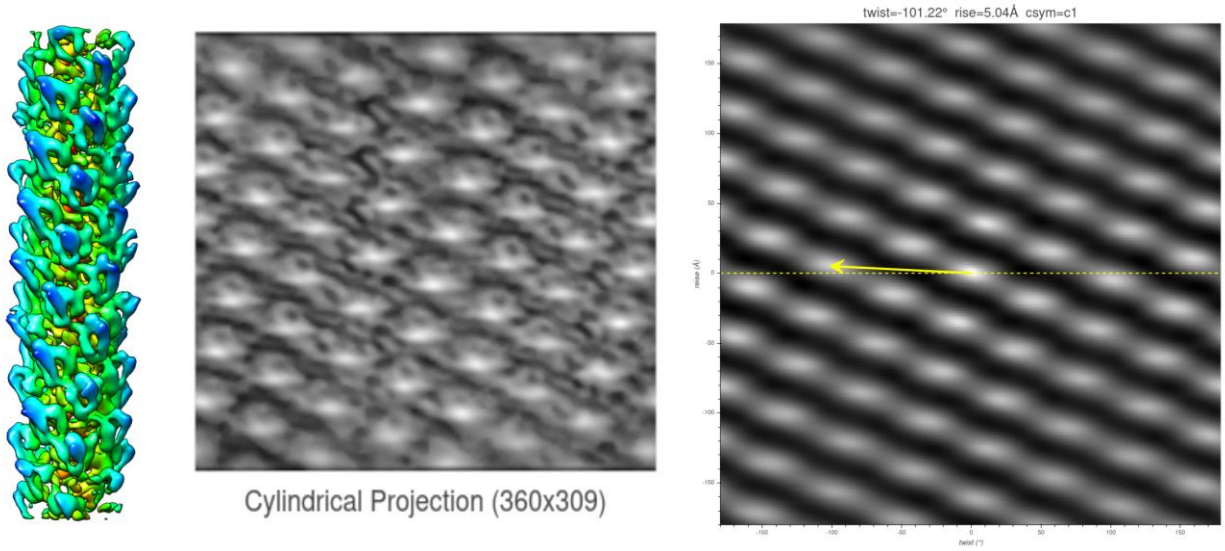


Figure 6.5. HI3D results for ab initio asymmetric reconstructions of MAVS CARD dataset (EMPIAR-10031). From left to right: surface view of the asymmetric reconstruction, cylindrical projection, HI3D output helical twist, rise and C-sym in text and as a unit cell vector in the ACF image.

Table 6.1 Comparison of HI3D reported helical parameter with the published values.

Dataset	EMPIAR	EMDB	Symmetry	# of particles	Pixel Size (Å)	Published Rise (Å)	Published Twist (°)	HI3D Rise (Å)	HI3D Twist (°)
TMV	10305	10129	C1	21,744	2.552	1.406	22.038	1.25	-19.69
MAVS CARD	10031	6428	C1	118,301	2.1	5.088	-101.436	5.04	-101.22
VipA/VipB	10019	2699	C6	8,899	2	21.8	29.4	21.83	-29.31
HIV tubes	N/A	10239	C1	2,296	2.44	6.95	-31.13	6.96	31.14
		10246	C1	1,814	2.44	6.44	-28.68	6.39	28.62

VipA/VipB (Figure 6.6)

Our third test case, VipA/VipB, is a type VI secretion system analogous to helical tails of myophages. VipA/VipB is a relatively rigid helical tube with an outer diameter of about 300 Å. We used the Vibrio Cholerae VipA/VipB dataset (EMPIAR 10019) (Kudryashev et al., 2015). HI3D analysis of our ab initio asymmetric reconstruction estimated the rise and twist at 21.83 Å and -29.31°, respectively, which is nearly identical to the previously published helical rise (21.8 Å) and twist (29.4°) (Kudryashev et al., 2015). HI3D also correctly detected the intrinsic C6 symmetry around the helical axis although only asymmetric symmetry was used for the reconstruction.

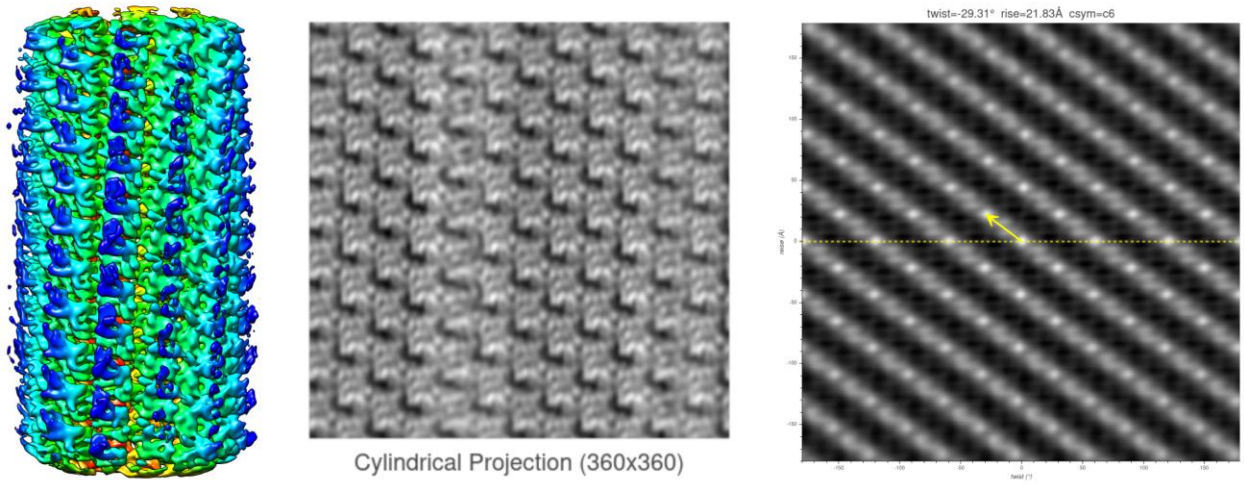


Figure 6.6. HI3D results for ab initio asymmetric reconstructions of VipA/VipB dataset (EMPIAR-10029). From left to right: surface view of the asymmetric reconstruction, cylindrical projection, HI3D output helical twist, rise and C-sym in text and as a unit cell vector in the ACF image.

HIV Tubes (Figure 6.7)

The data for our fourth test case, HIV tubes, provided by Dr. Peijun Zhang, was a more challenging dataset. HIV capsid protein is known to organize into a variety of structures, such as cones and tubes. Tube assemblies can also vary in diameter and helical symmetry. Dr. Peijun Zhang's recent work revealed high resolution HIV tube structures with seven helical symmetries (Ni et al., 2020). The dataset that we obtained only contains 5,529 particles from 39 tubes. After ab initio asymmetric reconstruction into three classes, two classes showed well resolved hexons

(Fig. 6.7), indicating that both structures were correct. The largest class contains 41.5% particles with HI3D-reported helical rise of 6.96 Å and twist 31.14° (Fig. 6.7). The second most populated class was composed of 32.8% particles with HI3D-reported helical rise of 6.39 Å and twist -28.62° (Fig. 6.7). The least populated class only had 25.7% particles and did not yield a clear structure. It is interesting to note that the cylindrical projection showed interleaved, well resolved, and poorly resolved angular/vertical stripes of hexons (Fig. 6.7), suggesting that these asymmetric reconstructions were far from ideal. Nevertheless, the 2D lattice in the ACF was sufficiently clear to estimate the helical symmetry very close to the reported values (Table 6.1).

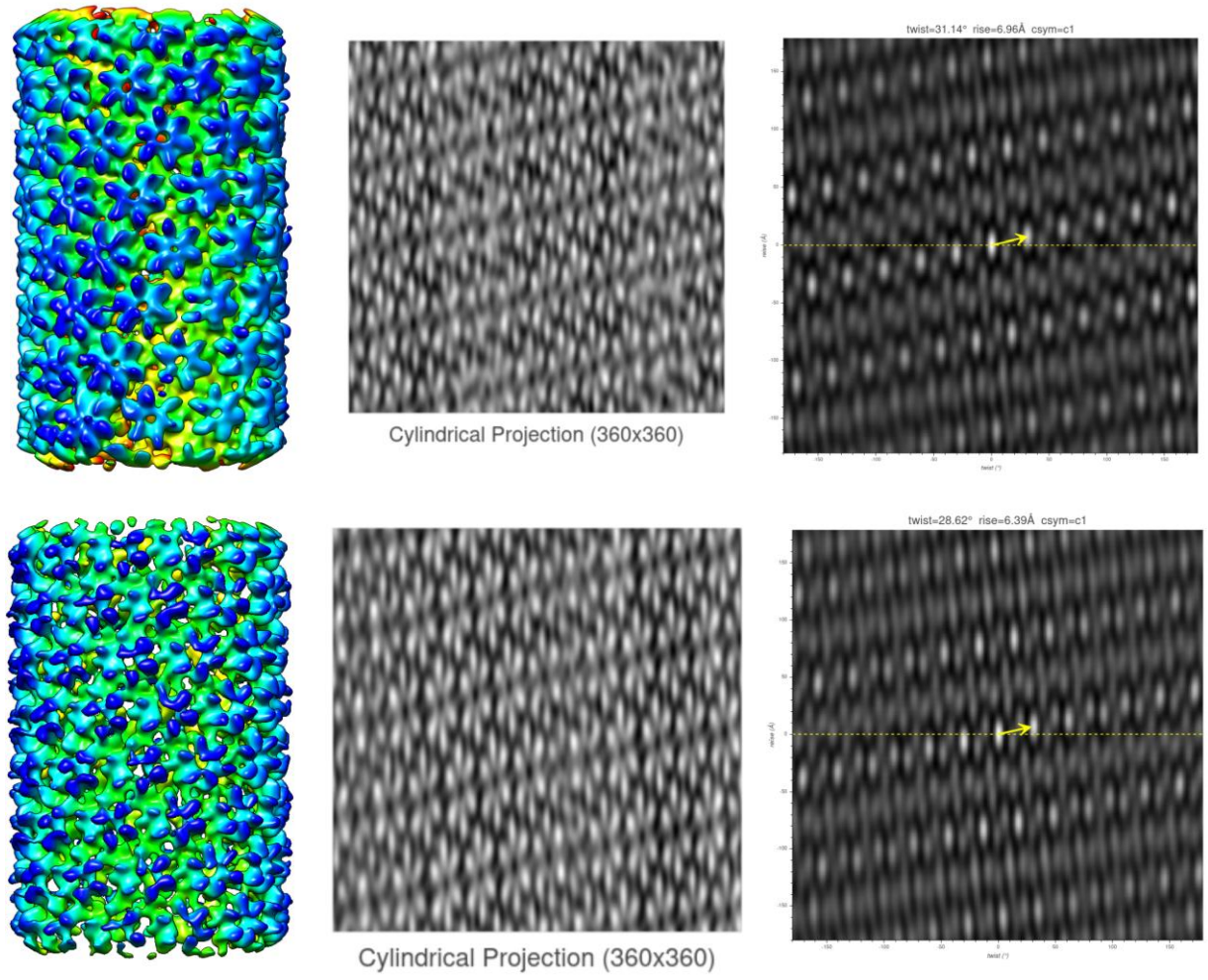


Figure 6.7. HI3D results for ab initio asymmetric reconstructions of HIV capsid protein (provided by Dr. Peijun Zhang) that resulted in two distinct asymmetric structures. From left to right: surface view of the asymmetric reconstruction, cylindrical projection, HI3D output helical twist, rise and C-sym in text and as a unit cell vector in the ACF image. The upper panel shows the best class with helical rise and twist of (6.96, 31.14), while the bottom panel is the 2nd class which yields helical rise and twist of (6.39, -28.62).

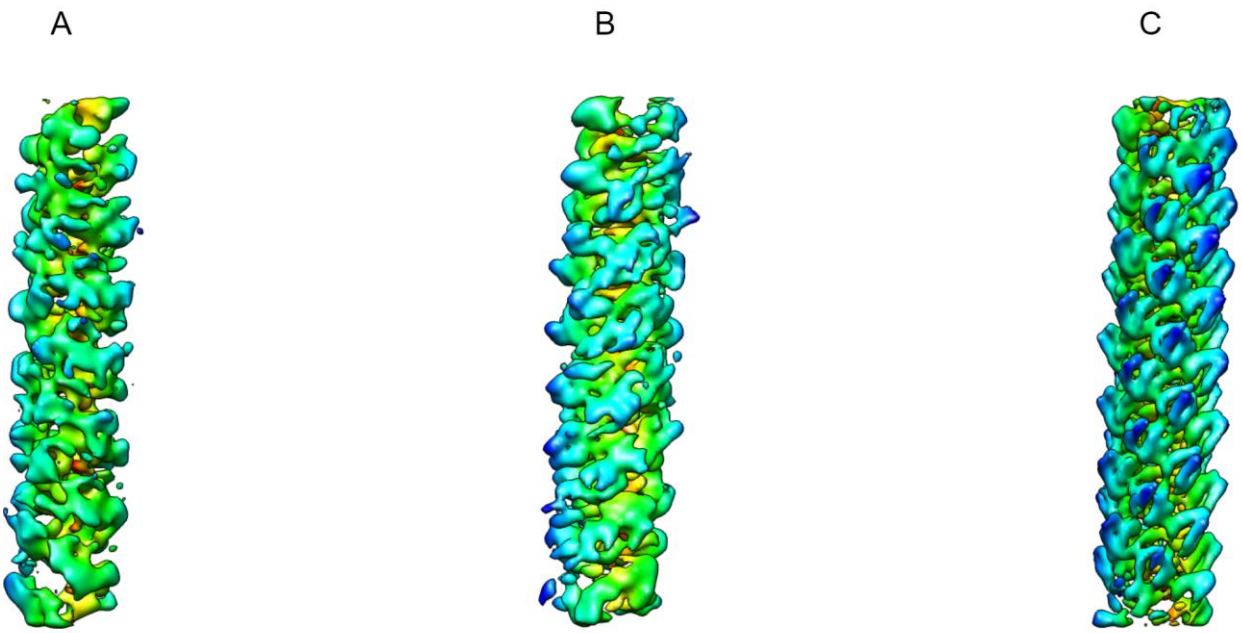


Figure 6.8. 3D surface view of all three classes of ab initio reconstruction of MAVS CARD dataset. All maps are slighted tilted. A and B are the two junk classes. The density map in C is more intact comparing to the other two. C is the only map that returned the correct helical parameters.

6.5 Discussion

The work we present here introduces a unique method, HI3D, to analyze helical symmetry in real space using asymmetric reconstructions. It offers an alternative approach to overcome the challenges associated with Fourier layer line indexing and/or confirm previously assigned helical symmetry estimations obtained from other methods. To demonstrate its robustness, we have tested it with four experimental datasets varying in helical diameter, symmetry, flexibility, and heterogeneity. From suboptimal asymmetric reconstructions, it has successfully estimated helical parameters that are accurate enough for the downstream helical refinement of all test cases. As a Web app, it is installation-free and convenient to use. It can be easily integrated into any existing image processing workflow. Its user-friendly interface improves access to new users. There is no stringent requirement on the input map resolution. Furthermore, the “crystalline order” of the cylindrical projections and sharpness of the “diffraction spots” in the ACF images could provide the users an intuitive assessment of the quality of the input maps. It can also be used as a validation tool for maps deposited to EMDB for easy access to the helical symmetry information.

Compared with the helical symmetry search methods in the existing software, like RELION helical_toobox and cryoSPARC Helical Symmetry Search utility, HI3D demonstrated more robust performance, especially for asymmetric reconstructions with suboptimal helical structure features. In these software, the symmetry search is cross-correlation based, which relies on user input to define a relatively small search range for both helical twist and rise parameters. For unknown systems, it can be hard to guess an accurate search range. Therefore, they are not global symmetry search methods and perform poorly when provided with full range. In contrast, HI3D is designed to directly estimate the helical parameters globally by default. For bad quality maps, for example, with varying crystalline order in different stripes of the cylindrical projection (Fig. 6.7), users can select to use only the clearly defined regions in the cylindrical projection to improve the 2D lattice in the ACF. From this aspect, HI3D has the advantage of lowering the quality threshold needed for helical indexing. Coupled with HI3D indexing, some suboptimal asymmetric maps, with partial disordered regions, can still be used to decipher the helical parameters as shown in our four examples.

Despite all these advantages, like all other methods, there is also a shortcoming of HI3D. Different ab initio asymmetric reconstructions from the same dataset can yield different apparent helical symmetries, and none may be correct. Multiple ab initio reconstruction jobs are recommended to find a consistent solution. Based on our experience, for high quality datasets, whatever reconstruction software is used, there was a high probability for at least one class to converge to the correct structure. However, for low-quality or challenging datasets, it could be that none of the reconstructed maps is correct.

We acknowledge that the ab initio asymmetric reconstructions would not always output the correct structure. For each test datasets, we have performed multiple ab initio Class3D reconstruction jobs and within each job, three models were generated. We only output the helical parameters that we have consistently obtained in the independent runs. Users are advised to perform multiple ab initio asymmetric reconstruction and use the 2D lattice generated by HI3D to examine the quality of the map. If a clear 2D lattice with sharp peaks is observed, the input map would be good enough for helical indexing with HI3D. Otherwise the output parameters wouldn’t be trustable.

Overall, helical samples remain one of the most challenging targets of cryoEM studies because of the complexity/ambiguity involved in helical symmetry estimation and heterogeneity.

We hope this work can provide more resources for education, make structural research involving helical structures more accessible, and provide a new direction for helical indexing.

6.6 Acknowledgement

This work was supported in part by grants from NIGMS U24GM116789, NIAID R01AI111095, NINDS U01NS110437, NIA RF1AG071177 to W.J. and the NIH T32 GM132024 Molecular Biophysics Training Program at Purdue (https://molbiophys.science.purdue.edu) to B.G. We thank Dr. Peijun Zhang for providing the HIV tube dataset as a test for HI3D. We thank the Purdue Rosen Center for Advanced Computing (RCAC) (https://www.rcac.purdue.edu) for providing the computing resources for the development and application of HI3D.

PUBLICATIONS

1. Chen Sun, Xueyong Xu, Frank S. Vago, Pengwei Huang, Ming Tang, Xi Jiang, Wen Jiang. “2.6 Å Structure of Tulane Virus With Minor Mutations Leading to Receptor Change.” (Manuscript)
2. Chen Sun, Brenda Gonzalez, Wen Jiang. “Ab Initio Helical Indexing in Real Space Without the Need for Fourier Layer Lines.” (Submitted)
3. Xia, Ming, Pengwei Huang, Chen Sun, Ling Han, Frank S. Vago, Kunpeng Li, Weiming Zhong et al. "Bioengineered Norovirus S60 Nanoparticles as a Multifunctional Vaccine Platform." ACS nano 12, no. 11 (2018): 10665-10682.
4. Kunpeng Li*, Chen Sun*, Thomas Klose, Frank S. Vago, Jose Irimia-Dominguez, Ruben Vidal, Wen Jiang. “Sub-3 Å Apoferritin Structure Determined With Full Range of Phase Shifts Using A Single Position Of Volta Phase Plate.” Journal of Structural Biology (2019). (*co-first author)
5. Xiong, Xiansong*, Chen Sun*, Frank S. Vago, Thomas Klose, Jiankang Zhu, and Wen Jiang. "Cryo-EM Structure of Heterologous Protein Complex Loaded Thermotoga Maritima Encapsulin Capsid." Biomolecules 10, no. 9 (2020): 1342. (*co-first author)
6. Irimia-Dominguez, Jose, Chen Sun, Kunpeng Li, Barry B. Muhoberac, Grace I. Hallinan, Holly J. Garringer, Bernardino Ghetti, Wen Jiang, and Ruben Vidal. "Cryo-EM structures and functional characterization of homo-and heteropolymers of human ferritin variants." Scientific reports 10, no. 1 (2020): 1-10.
7. Chen Sun, Brenda Gonzalez, Frank S. Vago, and Wen Jiang. "High resolution single particle Cryo-EM refinement using JSPR." Progress in Biophysics and Molecular Biology 160 (2021): 37-42.
8. Runrun Wu, Jeremy W. Bakelar, Karl Lundquist, Zijian Zhang, Katie M. Kuo, David Ryoo, Yui Tik Pang, Chen Sun, Tommi White, Thomas Klose, Wen Jiang, James C.Gumbart and Nicholas Noinaj. “Plasticity within the barrel domain of BamA mediates a hybrid-barrel mechanism by BAM” (In press)

REFERENCES

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367.

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., Adams, P.D., 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol 74, 531–544.

Akbar M. Dastjerdi, David R. Snodgrass, Janice C. Bridger, Characterisation of the bovine enteric calici-like virus, Newbury agent 1, FEMS Microbiology Letters, Volume 192, Issue 1, November 2000, Pages 125–131, https://doi.org/10.1111/j.1574-6968.2000.tb09370.x

Akita, F., Chong, K.T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., Nakagawa, A., 2007. The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. J. Mol. Biol. 368, 1469–1483.

Avalos, J.L., Fink, G.R., Stephanopoulos, G., 2013. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. Nat. Biotechnol. 31, 335–341.

Baraibar, M.A., Muhoberac, B.B., Garringer, H.J., Hurley, T.D., Vidal, R., 2009. Unraveling of the E-helices and Disruption of 4-Fold Pores Are Associated with Iron Mishandling in a Mutant Ferritin Causing Neurodegeneration. J. Biol. Chem. 285, 1950–1956.

Barbé, L., Le Moullac-Vaidye, B., Echasserieau, K. et al. Histo-blood group antigen-binding specificities of human rotaviruses are associated with gastroenteritis but not with in vitro infection. Sci Rep 8, 12961 (2018). https://doi.org/10.1038/s41598-018-31005-4

Berriman, J., Serpell, L.C., Oberg, K.A., Fink, A.L., Goedert, M., Crowther, R.A., 2003. Tau filaments from human brain and from in vitro assembly of recombinant protein show cross-beta structure. Proc. Natl. Acad. Sci. U. S. A. 100, 9034–9038.

Cassidy-Amstutz, C., Oltrogge, L., Going, C.C., Lee, A., Teng, P., Quintanilla, D., East-Seletsky, A., Williams, E.R., Savage, D.F., 2016. Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. Biochemistry 55, 3461–3468.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H.W., Henderson, R., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135, 24–35.

Chen, Z., Sun, L., Zhang, Z., Fokine, A., Padilla-Sanchez, V., Hanein, D., Jiang, W., Rossmann, M.G., Rao, V.B., 2017. Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. Proc. Natl. Acad. Sci. U. S. A. 114, E8184–E8193.

Conley, M.J., McElwee, M., Azmi, L. et al. Calicivirus VP2 forms a portal-like assembly following receptor engagement. Nature 565, 377–381 (2019). https://doi.org/10.1038/s41586-018-0852-1

Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S., & Landau, N. R. (1997). Change in coreceptor use correlates with disease progression in HIV-1–infected individuals. The Journal of experimental medicine, 185(4), 621-628.

Cornejo, E., Abreu, N., Komeili, A., 2014. Compartmentalization and organelle formation in bacteria. Current Opinion in Cell Biology. https://doi.org/10.1016/j.ceb.2013.12.007

Dai, X., Hong Zhou, Z., 2018. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science 360, eaao7298.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., Baumeister, W., 2014. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. Proc. Natl. Acad. Sci. U. S. A. 111, 15635–15640.

Danev, R., Nagayama, K., 2001. Transmission electron microscopy with Zernike phase plate. Ultramicroscopy 88, 243–252.

Danev, R., Tegunov, D., Baumeister, W., 2017. Using the Volta phase plate with defocus for cryo-EM single particle analysis. Elife 6. https://doi.org/10.7554/eLife.23006

Day, J.M., Ballard, L.L., Duke, M.V., Schefflerm B.E. and Zsak, L. (2010). Metagenomic analysis of the turkey gut RNA virus community. Virology Journal 7: 313.

De Rosier, D.J., Klug, A., 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature. https://doi.org/10.1038/217130a0

Desai, N., Brown, A., Amunts, A., Ramakrishnan, V., 2017. The structure of the yeast mitochondrial ribosome. Science 355, 528–531.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., André, I., 2011. Modeling symmetric macromolecular structures in Rosetta3. PLoS One 6, e20450.

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

Duan, C.-G., Wang, X., Xie, S., Pan, L., Miki, D., Tang, K., Hsu, C.-C., Lei, M., Zhong, Y., Hou, Y.-J., Wang, Z., Zhang, Z., Mangrauthia, S.K., Xu, H., Zhang, H., Dilkes, B., Andy Tao, W., Zhu, J.-K., 2017. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. Cell Research. https://doi.org/10.1038/cr.2016.147

Dubochet, J., Adrian, M., Chang, J., Homo, J., Lepault, J., McDowall, A., & Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Quarterly Reviews of Biophysics, 21(2), 129-228. doi:10.1017/S0033583500004297

Egelman, E.H., 2000. A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy 85, 225–234.

Egelman, E.H., 2007. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. J. Struct. Biol. 157, 83–94.

Egelman, E.H., 2014. Ambiguities in helical reconstruction. https://doi.org/10.7554/eLife.04969

Egelman, E.H., Wang, F., 2021. Cryo-EM is a powerful tool, but helical applications can have pitfalls. Soft Matter 17, 3291–3293.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501.

Fan, X., Zhao, L., Liu, C., Zhang, J.-C., Fan, K., Yan, X., Peng, H.-L., Lei, J., Wang, H.-W., 2017. Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM. Structure 25, 1623–1630.e3.

Fang, Q., Zhu, D., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., Chen, Z., Sun, Y., Gross, M.L., Van Etten, J.L., Zhang, X., Rossmann, M.G., 2019. Near-atomic structure of a giant virus. Nat. Commun. 10, 388.

Farkas T, Lun CWP, Fey B. Relationship between genotypes and serotypes of genogroup 1 recoviruses: a model for human norovirus antigenic diversity. J Gen Virol. 2014;95(Pt 7):1469-1478. doi:10.1099/vir.0.064675-0

Farkas, T., Sestak, K., Wei, C., Jiang, X., 2008. Characterization of a rhesus monkey calicivirus representing a new genus of Caliciviridae. J. Virol. 82, 5408–5416.

Fischer, H., Polikarpov, I., Craievich, A.F., 2004. Average protein density is a molecular-weight-dependent function. Protein Sci. 13, 2825–2828.

Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., Scheres, S.H.W., 2017. Cryo-EM structures of tau filaments from Alzheimer's disease. Nature 547, 185.

Frank, J. O. A. C. H. I. M., and B. R. I. A. N. Shimkin. "A new image processing software system for structural analysis and contrast enhancement." Proc. gth Int. Conf. Electron Microsc. 1978.

Frank, J., Ourmazd, A., 2016. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. Methods 100, 61–67.

Franklin, R.E., Gosling, R.G., 1953. Molecular configuration in sodium thymonucleate. Nature 171, 740–741.

Galkin, V.E., Orlova, A., Vos, M.R., Schröder, G.F., Egelman, E.H., 2015. Near-atomic resolution for one state of F-actin. Structure 23, 173–182.

Giessen, T.W., Orlando, B.J., Verdegaal, A.A., Chambers, M.G., Gardener, J., Bell, D.C., Birrane, G., Liao, M., Silver, P.A., 2019. Large protein organelles form a new iron sequestration system with high storage capacity. eLife. https://doi.org/10.7554/elife.46070

Glaeser, R.M., Typke, D., Tiemeijer, P.C., Pulokas, J., Cheng, A., 2011. Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. J. Struct. Biol. 174, 1–10.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E., 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25.

Grant, T., Rohou, A., Grigorieff, N., 2018. cisTEM: User-friendly software for single-particle image processing. https://doi.org/10.1101/257618

Guo, F., Jiang, W., 2013. Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses, in: Methods in Molecular Biology. pp. 401–443.

Hallinan, G.I., Hoq, M.R., Ghosh, M., Vago, F.S., Fernandez, A., Garringer, H.J., Vidal, R., Jiang, W., Ghetti, B., 2021. Structure of Tau filaments in Prion protein amyloidoses. Acta Neuropathol. 142, 227–241.

Handley SA, Thackray LB, Zhao G, et al. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. Cell. 2012;151(2):253-266. doi:10.1016/j.cell.2012.09.024

Harauz, G., van Heel, M., 1986. Exact filters for general geometry three dimensional reconstruction. sbio.uct.ac.za 73, 146–156.

Hawkes, P.W., 2015. The correction of electron lens aberrations. Ultramicroscopy 156, A1–64.

He, S., Scheres, S.H.W., 2017. Helical reconstruction in RELION. J. Struct. Biol. 198, 163–176.

Heggelund, J.E., Varrot, A., Imberty, A., Krengel, U., 2017. Histo-blood group antigens as mediators of infections. Curr. Opin. Struct. Biol. 44, 190–200.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. Journal of molecular biology, 213(4), 899-929.

Henderson, R., Baldwin, J.M., Downing, K.H., Lepault, J., Zemlin, F., 1986. Structure of purple membrane from halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. Ultramicroscopy 19, 147–178.

Hettler, S., Kano, E., Dries, M., Gerthsen, D., Pfaffmann, L., Bruns, M., Beleggia, M., Malac, M., 2018. Charging of carbon thin films in scanning and phase-plate transmission electron microscopy. Ultramicroscopy 184, 252–266.

Heymann, J.B., Marabini, R., Kazemi, M., Sorzano, C.O.S., Holmdahl, M., Mendez, J.H., Stagg, S.M., Jonic, S., Palovcak, E., Armache, J.-P., Zhao, J., Cheng, Y., Pintilie, G., Chiu, W., Patwardhan, A., Carazo, J.-M., 2018. The First Single Particle Analysis Map Challenge: A Summary of the Assessments. J. Struct. Biol.

Hicks, P.M., Rinker, K.D., Baker, J.R., Kelly, R.M., 1998. Homomultimeric protease in the hyperthermophilic bacterium Thermotoga maritima has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria. FEBS Lett. 440, 393–398.

Hsu EC, Sarangi F, Iorio C, et al. A single amino acid change in the hemagglutinin protein of measles virus determines its ability to bind CD46 and reveals another receptor on marmoset B cells. J Virol. 1998;72(4):2905-2916. doi:10.1128/JVI.72.4.2905-2916.1998

Huang P, Farkas T, Zhong W, et al. Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. J Virol. 2005;79(11):6714-6722. doi:10.1128/JVI.79.11.6714-6722.2005

Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J., Chiu, W., 2008. Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. Nature 451, 1130–1134.

Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J Virol. 1972;10(5):1075-1081. doi:10.1128/JVI.10.5.1075-1081.1972

Kerfeld, C.A., Heinhorst, S., Cannon, G.C., 2010. Bacterial Microcompartments. Annual Review of Microbiology.

Kidd, M., 1963. Paired Helical Filaments in Electron Microscopy of Alzheimer's Disease. Nature.

Kudryashev, M., Wang, R.Y.-R., Brackmann, M., Scherer, S., Maier, T., Baker, D., DiMaio, F., Stahlberg, H., Egelman, E.H., Basler, M., 2015. Structure of the type VI secretion system contractile sheath. Cell 160, 952–962.

Lau, Y.H., Giessen, T.W., Altenburg, W.J., Silver, P.A., 2018. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. Nature Communications. https://doi.org/10.1038/s41467-018-03768-x

Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. (2017) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. Jan 4;46(D1): D708-D717. PMID: 29040670. PMCID: PMC5753373.

L'Homme, Y., Sansregret, R., Plante-Fortier, E., Lamontagne, A.M., Ouardani, M., Lacroix, G. and Simard, C. (2009). Genomic characterization of swine caliciviruses representing a new genus of Caliciviridae. Virus Genes 39: 66-75.

Li, K., Sun, C., Klose, T., Irimia-Dominguez, J., Vago, F.S., Vidal, R., Jiang, W., 2019. Sub-3 Å apoferritin structure determined with full range of phase shifts using a single position of volta phase plate. J. Struct. Biol. 206, 225–232.

Liang, Y.-L., Khoshouei, M., Radjainia, M., Zhang, Y., Glukhova, A., Tarrasch, J., Thal, D.M., Furness, S.G.B., Christopoulos, G., Coudrat, T., Danev, R., Baumeister, W., Miller, L.J., Christopoulos, A., Kobilka, B.K., Wootten, D., Skiniotis, G., Sexton, P.M., 2017. Phase-plate cryo-EM structure of a class B GPCR–G-protein complex. Nature 546, 118–123.

Lindesmith, L.C., Brewer-Jensen, P.D., Mallory, M.L., Debbink, K., Swann, E.W., Vinjé, J., Baric, R.S., 2018. Antigenic Characterization of a Novel Recombinant GII.P16-GII.4 Sydney Norovirus Strain With Minor Sequence Variation Leading to Antibody Escape. J. Infect. Dis. 217, 1145–1152.

Liu, Z., Guo, F., Wang, F., Li, T.-C., Jiang, W., 2016. 2.9 Å Resolution Cryo-EM 3D Reconstruction of Close-Packed Virus Particles. Structure 24, 319–328.

Lochridge, V.P., Hardy, M.E., 2007. A single-amino-acid substitution in the P2 domain of VP1 of murine norovirus is sufficient for escape from antibody neutralization. J. Virol. 81, 12316–12322.

Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97.

Lynch, E.M., Kollman, J.M., Webb, B.A., 2020. Filament formation by metabolic enzymes-A new twist on regulation. Curr. Opin. Cell Biol. 66, 28–33.

Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268. PMID: 30976793; PMCID: PMC6602479.

Madeley CR, Cosgrove BP. 1976. Caliciviruses in man. Lancet i:199–200. (Letter.) doi:10.1016/S0140-6736(76)91309-X.

Mahar, J.E., Donker, N.C., Bok, K., Talbo, G.H., Green, K.Y., Kirkwood, C.D., 2014. Identification and characterization of antibody-binding epitopes on the norovirus GII.3 capsid. J. Virol. 88, 1942–1952.

Mallagaray, A., Creutznacher, R., Dülfer, J. et al. A post-translational modification of human Norovirus capsid protein attenuates glycan binding. Nat Commun 10, 1320 (2019).

Marabini, R., Kazemi, M., Sorzano, C.O.S., Carazo, J.M., 2018. Map Challenge: Analysis using a Pair Comparison Method based on Fourier Shell Correlation. J. Struct. Biol. https://doi.org/10.1016/j.jsb.2018.09.009

Martin, W., 2010. Evolutionary origins of metabolic compartmentalization in eukaryotes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365, 847–855.

McHugh, C.A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A.A., Heymann, J.B., Winkler, D.C., Lam, A.S., Wall, J.S., Steven, A.C., Hoiczyk, E., 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. EMBO J. 33, 1896–1911.

Mikalsen, A.B., Nilsen, P., Frøystad-Saugen, M., Lindmo, K., Eliassen, T.M., Rode, M. and Evensen, O. (2014). Characterization of a novel calicivirus causing systemic infection in atlantic salmon (Salmo salar L.): proposal for a new genus of Caliciviridae. PLoS One. 9(9): e107132

Mor, S.K., Phelps, N.B.D., Ng, T.F.F., Subramaniam, K., Primus, A., Armien, A.G., McCann, R., Puzach, C., Waltzek, T.B. and Goyal, S.M. (2017). Genomic characterization of a novel calicivirus, FHMCV-2012, from baitfish in the USA. Arch. Virol. 2017 162(12):3619-3627.

Naeve, C. W., V. S. Hinshaw, and R. G. Webster. "Mutations in the hemagglutinin receptor-binding site can change the biological properties of an influenza virus." Journal of virology 51.2 (1984): 567-569.

Ni, T., Gerard, S., Zhao, G., Dent, K., Ning, J., Zhou, J., Shi, J., Anderson-Daniels, J., Li, W., Jang, S., Engelman, A.N., Aiken, C., Zhang, P., 2020. Intrinsic curvature of the HIV-1 CA hexamer underlies capsid topology and interaction with cyclophilin A. Nat. Struct. Mol. Biol. 27, 855–862.

Nichols, R.J., Cassidy-Amstutz, C., Chaijarasphong, T., Savage, D.F., 2017. Encapsulins: molecular biology of the shell. Crit. Rev. Biochem. Mol. Biol. 52, 583–594.

Oka T, Wang Q, Katayama K, Saif LJ. Comprehensive review of human sapoviruses. Clin Microbiol Rev. 2015;28(1):32-53. doi:10.1128/CMR.00011-14

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.

Parra F, Prieto M. Purification and characterization of a calicivirus as the causative agent of a lethal hemorrhagic disease in rabbits. J Virol 1990; 64:4013–4015.

Parra F, Prieto M. Purification and characterization of a calicivirus as the causative agent of a lethal hemorrhagic disease in rabbits. J Virol 1990; 64:4013–4015

Pedersen NC, Elliott JB, Glasgow A, Poland A, Keel K. An isolated epizootic of hemorrhagic-like fever in cats caused by a novel and highly virulent strain of feline calicivirus. Vet Microbiol. 2000 May 11;73(4):281-300. doi: 10.1016/s0378-1135(00)00183-8. PMID: 10781727; PMCID: PMC7117377.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Pintilie, G., Chiu, W., 2018. Assessment of Structural Features in Cryo-EM Density Maps using SSE and Side Chain Z-Scores. J. Struct. Biol. https://doi.org/10.1016/j.jsb.2018.08.015

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296.

Putri, R.M., Allende-Ballestero, C., Luque, D., Klem, R., Rousou, K.-A., Liu, A., Traulsen, C.H.-H., Rurup, W.F., Koay, M.S.T., Castón, J.R., Cornelissen, J.J.L.M., 2017. Structural Characterization of Native and Modified Encapsulins as Nanoplatforms for in Vitro Catalysis and Cellular Uptake. ACS Nano 11, 12796–12804.

Qiu Q, Dewey-Mattia D, Subramhanya S, et al. Food recalls associated with foodborne disease outbreaks, United States, 2006-2016. Epidemiol Infect. 2021;149:e190. Published 2021 Jul 19. doi:10.1017/S0950268821001722

Ravn, V., Dabelsteen, E., 2000. Tissue distribution of histo-blood group antigens. APMIS 108, 1–28.

Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". Nucl. Acids Res. 42(W1), W320-W324 - doi: 10.1093/nar/gku316

Rohou, A., Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221.

Sanchez, R.M., Zhang, Y., Chen, W., Dietrich, L., Kudryashev, M., 2020. Subnanometer-resolution structure determination in situ by hybrid subtomogram averaging - single particle cryo-EM. Nat. Commun. 11, 3709.

Scheres, S.H.W., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.

Schultheiß, K., Pérez-Willard, F., Barton, B., Gerthsen, D., Schröder, R.R., 2006. Fabrication of a Boersch phase plate for phase contrast imaging in a transmission electron microscope. Rev. Sci. Instrum. 77, 033701.

Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T.C., Rossmann, M.G., Kuhn, R.J., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science 352, 467–470.

Smits SL, Rahman M, Schapendonk CM, van Leeuwen M, Faruque AS, Haagmans BL, Endtz HP, Osterhaus AD. Calicivirus from novel Recovirus genogroup in human diarrhea, Bangladesh. Emerg Infect Dis. 2012 Jul;18(7):1192-5. doi: 10.3201/eid1807.120344. PMID: 22709854; PMCID: PMC3376821.

soft-matter, n.d. GitHub - soft-matter/trackpy: Python particle tracking toolkit [WWW Document]. URL https://github.com/soft-matter/trackpy (accessed 8.26.21).

Sonotaki, S., Takami, T., Noguchi, K., Odaka, M., Yohda, M., Murakami, Y., 2017. Successful PEGylation of hollow encapsulin nanoparticles from Rhodococcus erythropolis N771 without affecting their disassembly and reassembly properties. Biomater Sci 5, 1082–1089.

streamlit, n.d. GitHub - streamlit/streamlit: Streamlit — The fastest way to build data apps in Python. URL https://github.com/streamlit/streamlit (accessed 8.26.21).

Structural Characterization of Native and Modified Encapsulins as Nanoplatforms for in Vitro Catalysis and Cellular Uptake, n.d. https://doi.org/10.1021/acsnano.7b07669.s001

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151, 41–60.

Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M.J., Stetter, K.O., Weber-Ban, E., Ban, N., 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat. Struct. Mol. Biol. 15, 939–947.

Szepanski S, Gross HJ, Brossmer R, Klenk HD, Herrler G. A single point mutation of the influenza C virus glycoprotein (HEF) changes the viral receptor-binding activity. Virology. 1992;188(1):85-92. doi:10.1016/0042-6822(92)90737-a

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., Ludtke, S.J., 2007. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46.

Tange, O., 2018. GNU Parallel 2018. https://doi.org/10.5281/zenodo.1146014

Terry, R.D., 1963. THE FINE STRUCTURE OF NEUROFIBRILLARY TANGLES IN ALZHEIMER'S DISEASE. Journal of Neuropathology and Experimental Neurology. https://doi.org/10.1097/00005072-196310000-00005

Traum, J. Vesicular exanthema of swine. J. Amer. vet. med. Ass. 88, 316--327 (1936).

Wang, F., Gu, Y., O'Brien, J.P., Yi, S.M., Yalcin, S.E., Srikanth, V., Shen, C., Vu, D., Ing, N.L., Hochbaum, A.I., Egelman, E.H., Malvankar, N.S., 2019. Structure of Microbial Nanowires Reveals Stacked Hemes that Transport Electrons over Micrometers. Cell 177, 361–369.e10.

Wang, Ray Yu-Ruei, Yifan Song, Benjamin A. Barad, Yifan Cheng, James S. Fraser, and Frank DiMaio. 2016. "Automated Structure Refinement of Macromolecular Assemblies from Cryo-EM Maps Using Rosetta." eLife 5 (September). https://doi.org/10.7554/eLife.17219.

Wang, Z., Li, C., Ellenburg, M., Soistman, E., Ruble, J., Wright, B., Ho, J.X., Carter, D.C., 2006. Structure of human ferritin L chain. Acta Crystallogr. D Biol. Crystallogr. 62, 800–806.

Wei C, Meller J, Jiang X. Substrate specificity of Tulane virus protease. Virology. 2013;436(1):24-32. doi:10.1016/j.virol.2012.10.010

Wolf, S., Reetz, J. and Otto, P. Genetic characterization of a novel calicivirus from chicken. Arch. Virol. 2011; 156: 1143-1150.

Xu, H., He, X., Zheng, H., Huang, L.J., Hou, F., Yu, Z., de la Cruz, M.J., Borkowski, B., Zhang, X., Chen, Z.J., Jiang, Q.-X., 2015. Correction: Structural basis for the prion-like MAVS filaments in antiviral innate immunity. eLife. https://doi.org/10.7554/elife.07546

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A., Ferrin, T.E., 2012. UCSF Chimera, MODELLER, and IMP: An integrated modeling system. Journal of Structural Biology. https://doi.org/10.1016/j.jsb.2011.09.006

Yonekura, K., Maki-Yonekura, S., Namba, K., 2003. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature 424, 643–650.

Yu, G., Li, K., Huang, P., Jiang, X., Jiang, W., 2016a. Antibody-Based Affinity Cryoelectron Microscopy at 2.6-Å Resolution. Structure 24, 1984–1990.

Yu, G., Li, K., Jiang, W., 2016b. Antibody-based affinity cryo-EM grid. Methods 100, 16–24.

Yu, G., Li, K., Liu, Y., Chen, Z., Wang, Z., Yan, R., Klose, T., Tang, L., Jiang, W., 2016. An algorithm for estimation and correction of anisotropic magnification distortion of cryo-EM images without need of pre-calibration. J. Struct. Biol. 195, 207–215.

Zernike, F., 1942. Phase contrast, a new method for the microscopic observation of transparent objects. Physica 9, 686–698.

Zhang, D., Huang, P., Zou, L., Lowary, T.L., Tan, M., Jiang, X., 2014. Tulane Virus Recognizes the A Type 3 and B Histo-Blood Group Antigens. J. Virol. 89, 1419–1427.

Zhang, K., 2016. Gctf: Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12.

Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.-C., Schmid, M.F., Das, R., Chiu, W., 2019. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. Nat. Commun. 10, 5511.

Zhang, N., Shan, H., Liu, M., Li, T., Luo, R., Yang, L., Qi, L., Chu, X., Su, X., Wang, R., Liu, Y., Sun, W., Shen, Q.-T., 2021. Structure and assembly of double-headed Sendai virus nucleocapsids. Commun Biol 4, 494.

Zhao, H., Li, K., Lynn, A.Y., Aron, K.E., Yu, G., Jiang, W., Tang, L., 2017. Structure of a headful DNA-packaging bacterial virus at 2.9 Å resolution by electron cryo-microscopy. Proc. Natl. Acad. Sci. U. S. A. 114, 3601–3606.

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

Zhong, E.D., Bepler, T., Davis, J.H., Berger, B., 2019. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. arXiv [q-bio.QM].

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., Scheres, S.H., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7. https://doi.org/10.7554/eLife.42166

Zivanov, J., Nakane, T., Scheres, S.H.W., 2019. Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Datasets in RELION-3.1.