Title

Investigating air-water interface adsorption phenomenon and the architectural organization of pleomorphic virus

Permalink

Author

Kang, Joon

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Investigating air-water interface adsorption phenomenon and the architectural organization of

pleomorphic virus

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Joon Kang

2024

ABSTRACT OF THE DISSERTATION

Investigating air-water interface adsorption phenomenon and the architectural organization of

pleomorphic virus

by

Joon Kang

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2024

Professor Hong Zhou, Chair

As the recent advances in cryogenic electron microscopy (cryoEM) have made it a tool of choice for determining biomolecular structures, it has been widely recognized that many complexes tend to have preferential orientation, aggregate, or disappear on cryoEM grids. However, the reasons for such misbehavior are not well understood, limiting the systematic approaches to resolve this problem. In this dissertation, we developed a theoretical formulation that explains these observations and rationalizes the use of surfactants in directly alleviating this phenomenon. We provided experimental evidence to demonstrate the effectiveness of various surfactants by visualizing the improved groEL particle distribution using cryogenic electron tomography (cryoET). We also conducted single particle analysis (SPA) on groEL to show that the surfactants do not adversely affect the protein structures at a near-atomic level. We further showed the effectiveness of surfactant on a more challenging membrane protein, CLC-1. In addition to testing the theoretical framework using a widely-test sample such as groEL, we also conducted a cryoET study of a more challenging zoonotic virus such as lymphocytic choriomeningitis virus (LCMV).

The structures of individually isolated LCMV proteins or their fragments have been reported, but the architectural organization and the full-length structure of these proteins remain unknown. To showcase the benefits of using cryoET and sub-tomogram averaging (STA), we conducted a structural study on LCMV, unveiling the architectural organization of viral proteins of LCMV virion and resolving the full-length *in situ* structure of the glycoprotein complex (GPC) with its transmembrane domain intact. CryoET revealed a great variation in the number of polymerases within each virion, adding new perspectives to the infection mechanism. Finally, our structure proposes potential key residues that can be considered for structure-guided vaccine design against LCMV and arenavirus.

The dissertation of Joon S. Kang is approved.

Emil Reisler

Feng Guo

Irene Chen

Oliver I. Fregoso

Hong Zhou, Committee Chair

University of California, Los Angeles

2024

This dissertation is dedicated to my wife Olivia & the furry friends Daehan, Toto, Nana, Morae

Table of Contents

Table of Figures

Acknowledgments

I would like to thank my dissertation advisor and mentor, Dr. Z. Hong Zhou, for his support and guidance over the past 5 years. I had many ups and downs throughout the years, but he was always there to support me and guide me in the right direction. We had some disagreements throughout the course of my Ph.D journey, but he embraced me with the utmost patience and understanding. I truly thank you for that. He was also very respectful and supportive of my career change and made sure that my dream of graduating before entering dental school became a reality. I really do not think I would have been able to achieve this if it were not for his leadership, wisdom, and altruistic nature. Thank you so much, Dr. Zhou, for your great mentorship and for helping me grow not only as a scientist but also as a human being.

I would like to thank my committee members, Dr. Emil Reisler, Dr. Feng Guo, Dr. Irene Chen, and Dr. Oliver Fregoso, for their guidance and support for my research and my career development. It was through their sincere care and wisdom that I was pushed to mature as a scientist. I would like to send a special Thank you to Dr. Feng Guo for being a wonderful mentor throughout the difficult times in my personal life. You gave me the special courage to move forward with my plan. Thank you so much, Dr. Guo.

I would like to thank my collaborator, Dr. Ming Luo, for his support and wisdom throughout our project on LCMV. His careful assessment of the progress and prompt communications allowed the project to flow smoothly and the manuscript to be submitted on time.

I would like to thank my lab members (current and old) for their warm support and wisdom. I would especially like to thank Dr. Muchen Pan, Dr. Yuntao Liu, Dr. Xian Xia, and Dr. Weiguang Wang for sharing computational knowledge with me and for being great friends. I would also like to thank Dr. Kang Zhou and Dr. Hui Wang for sharing their wisdom and being wonderful partners

in the main project. I would also like to thank Dr. David Strugatsky and Wong Hoi Hui for their help with my imaging sessions.

I also would like to thank the Molecular Biology Institute (MBI), Biochemistry, Biophysics, and Structural Biology (BBSB) home area, and the Department of Microbiology, Immunology, and Molecular Genetics (MIMG) for providing me the opportunity to pursue my Ph.D journey at UCLA.

I would also like to thank my friends at New Life Vision Church for their support and prayers. I especially like to thank Bryan for being a true big bro to me. You really made me grow as a human being.

I would like to thank my wife, Olivia, for trusting me and supporting me from day one. We have had some tough times together, but we made it through every time, and here we are. I am so proud of what we have achieved together and I look forward to achieving bigger things with you for many years to come. I love you! And last but not least, my dearest Daehan, Toto, Nana, and Morae, Thank you for being there for me and teaching me what love is. I love you guys.

Biographical Sketch

Joon S. Kang

EDUCATION

University of California, Los Angeles
Candidate of Ph.D. in Molecular Biology (2019 – present)

California State Polytechnic University, Pomona.
Master of Science in Biological Sciences (2013 – 2016)

California State University, Fullerton
Master of Biotechnology (2011 – 2013)

University of California, Irvine
Bachelor of Science in Biological Sciences (2007-2011)

EXPERIENCE

Teaching Associate
University of California, Los Angeles, Department of Molecular Cell Developmental Biology
(2020, 2021)

Research Associate
California State Polytechnic University, Pomona, Department of Biological Sciences
(2016 – 2019)

Lecturer
California State Polytechnic University, Pomona, Department of Biological Sciences
(2017 – 2019)

Teaching Associate
California State Polytechnic University, Pomona, Department of Biological Sciences
(2014, 2022)

Clinical Care Extender
Hoag Hospital Newport Beach
(2015-2016)

PUBLICATIONS

Kang, JS, Zhou, K, Wang, H, Tang, S, Lyles, K, Luo, M, Zhou, ZH. Architectural organization and in situ fusion protein structure of lymphocytic choriomeningitis virus. Journal of Virology. <u>**Under Review**</u>

Kang, JS, Zhou, X, Liu, Y, Wang, K, Zhou, ZH. Theoretical framework and experimental solution for the air-water interface adsorption problem in cryoEM. Biophysics Reports (2023), 9(4): 215-229.

Pan, M, Alvarez, AL, **Kang, JS**, Wang, L, Fan, C, Zhou, ZH. Asymmetric Reconstruction of mammalian reovirus reveals interactions among RNA, transcriptional factor u2 and capsid proteins. Nature Communications (2021) 12, 4176.

Kang JS, Zhang AL, Faheem M, Zhang CJ, Ai N, Buynak JD, Welsh WJ, Oelschlaeger P. Virtual Screening and Experimental Testing of B1 Metallo-beta-lactamase Inhibitors. Journal of Chemical Information and Modeling (2018), 58, 1902-1914.

Xiang Y, Chang YN, Ge Y, **Kang JS**, Zhang YL, Liu XL, Oelschlaeger P, Yang KW. Azolylthioacetamides as a potent scaffold for the development of metallo-beta-lactamase inhibitors. Bioorganic & Medicinal Chemistry Letters (2017), 27, 5225-9.

Zhai L, Zhang YL, **Kang JS**, Oelschlaeger P, Xiao L, Nie SS, Yang KW. Trazolylthioacetamide: a valid scaffold for the development of New Delhi Metallo-beta-lactamase-1 (NDM-1) inhibitors. ACS Medicinal Chemistry Letters (2016), 7, 413-17.

Oelschlaeger P, Aitha M, Yang H, **Kang JS**, Zhang AL, Liu EM, Buynak JD, Crowder MW Meropenem and chromacef intermediates observed in IMP-25 metallo-beta-lactamase-catalyzed hydrolysis. Antimicrobial Agents and Chemotherapy (2015), 59, 4326-30.

Liu XL, Shi Y, **Kang JS**, Oelschlaeger P, Yang KW. Amino acid thioester derivatives: a highly promising scaffold for the development of metallo-beta- lactamase L1 inhibitors. ACS Medicinal Chemistry Letters (2015) 6, 660-664.

Yang SK, **Kang JS**, Oelschlaeger P, Yang KW. Azolylthioacetamide: a highly promising scaffold for the development of metallo-beta-lactamase inhibitors. ACS Medicinal Chemistry Letters (2015), 6, 455-560.

Chapter 1: Introduction

1.1 Structural biology

The three-dimensional (3D) structural information of biological molecules, such as proteins, nucleic acids, and complex macromolecular assemblies, is the key to understanding their mechanistic function, underlying cellular processes, disease involvement, and drug discovery. Structural biology, with its arsenal of techniques such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), cryogenic electron microscopy (cryoEM), small angle X-ray scattering, and computer modeling, plays a crucial role in determining the 3D structure of biological macromolecules at atomic resolution. In recent years, protein structure prediction algorithms, such as Alphafold (1), have not only enhanced our understanding of biological molecules but have also opened up new avenues to tackle complex biological questions, demonstrating the significant role of cryoEM in advancing our knowledge in this field.

1.1.1 CryoEM

CryoEM has emerged as the leading technique for determining biomolecular structures, thanks to the groundbreaking work of Richard Henderson, Jacques Dubochet, and Joachim Frank. Their innovations in structure preservation via flash freezing, direct electron detectors, and computational image processing earned them the prestigious Nobel Prize in Chemistry in 2017 (2).

Figure 1.1 offers a comprehensive overview of the meticulous workflow involved in cryoEM sample preparation and imaging. The process begins with loading the sample onto the holey carbon side of the 2.5 mm diameter grid, which is glow-discharged beforehand. The loaded sample (2.5 µL) forms a meniscus on top of the carbon layer. The sample-loaded grid is then blotted with filter paper under controlled humidity, temperature, blotting force, and time, creating a thin sample-

embedded buffer layer on the grid. After removing the filter paper, the grid is plunged into liquid ethane, vitrifying the buffer. This step is critical as it minimizes crystal ice formation, enhancing the structural preservation of the sample and reducing strong diffraction of the electron beam. Finally, the grid is loaded into a transmission electron microscope (TEM), where the electrons emitted from the gun are manipulated by a series of lenses and pass through the sample on the grid. The transmitted electrons form the image, which is magnified and projected on the electron detector that directly transfers electron signals to a digitalized bitmap file.

Credit: Ke Ding
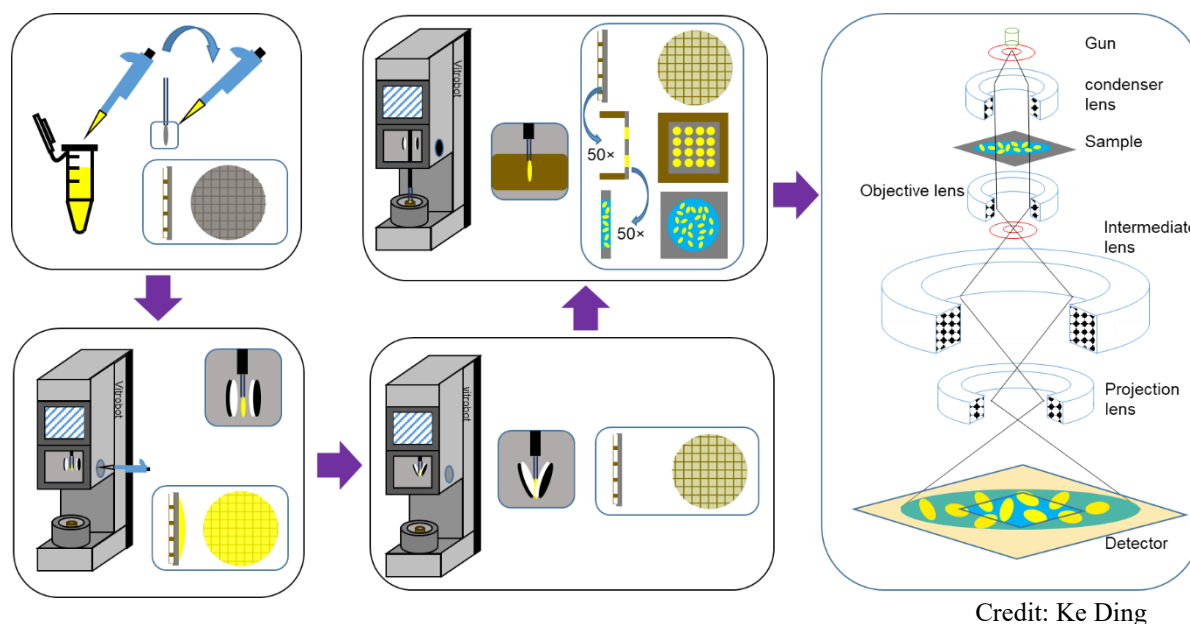
Figure 1.1: CryoEM workflow
CryoEM workflow of sample preparation and imging

The next step is computational image processing. The most widely used approach to obtaining high-resolution reconstruction is single particle analysis (SPA), as shown in Figure 1.2 (3). It is worth mentioning here that biomolecules are vulnerable to beam-induced damage. To overcome this, the average dosage of electrons exposed to each biomolecule is minimized by fractionating the exposure time. As such, multiple fractionated frames of bitmaps are collected as a movie per imaged region. The frames are aligned and merged into a single micrograph, with beam-induced

2

movements between frames corrected, using software such as Motioncorr (4). The next step is called the contrast transfer function (CTF) estimation. CTF is represented as contrast (intensity) versus spatial frequency, revealing how much contrast of each spatial frequency information is transferred to the final image. In the ideal situation, every spatial frequency wave has full contrast, retaining the full information in a given micrograph. However, in reality, some phase-shifted waves can combine to weaken or silence each other, leading to a loss of information i.e., the zeros of the contrast transfer function (CTF) of each micrograph (3). By adjusting the focus of the objective lens, one can shift the CTF, allowing the spatial frequency (features) that were missing in the previous CTF to now be visible. Therefore, images are collected at various defocus settings to restore the missing gap during the subsequent data processing. In the CTF estimation step, the CTF that matches that of each micrograph is computationally estimated by determining the defocus of each micrograph using software such as CTFFIND 4 (5). As a result, data from micrographs with different defocus can be merged to fill the CTF zeros, restoring the information.

Next, each biomolecule of interest is picked and extracted from every micrograph, resulting in tens of thousands of extracted "particles." These extracted particles are subjected to two-dimensional (2D) classification, which utilizes center and in-plane rotation to align the particles and obtain class averages based on structural similarity. The particles in the desired 2D average groups are used to obtain an *ab initio* three-dimensional (3D) model, which is used as a low-resolution reference for 3D classification, which helps to identify distinct conformations, macromolecular heterogeneities, or subsets of particles with different structural integrity (3). Next, iterative orientation refinement is performed that is monitored by the Fourier shell correlation (FSC) curve, which provides information on the level of the signal-to-noise ratio (SNR) as a function of the spatial frequency. The FSC curve is constructed by computing the correlation coefficient

between the Fourier transforms of the two volumes that are derived from the dataset split in halves (6). The orientation refinement continues until the two volumes converge, generating a refined 3D model. Finally, the existing model can be fitted into the refined 3D map for structural comparison, or the protein sequence can be fitted into the map to build a *de novo* 3D model of the protein. With enough particles averaged, the final refined reconstruction can reveal high-resolution features such as the protein backbone and side chain densities (Figure 1.2).
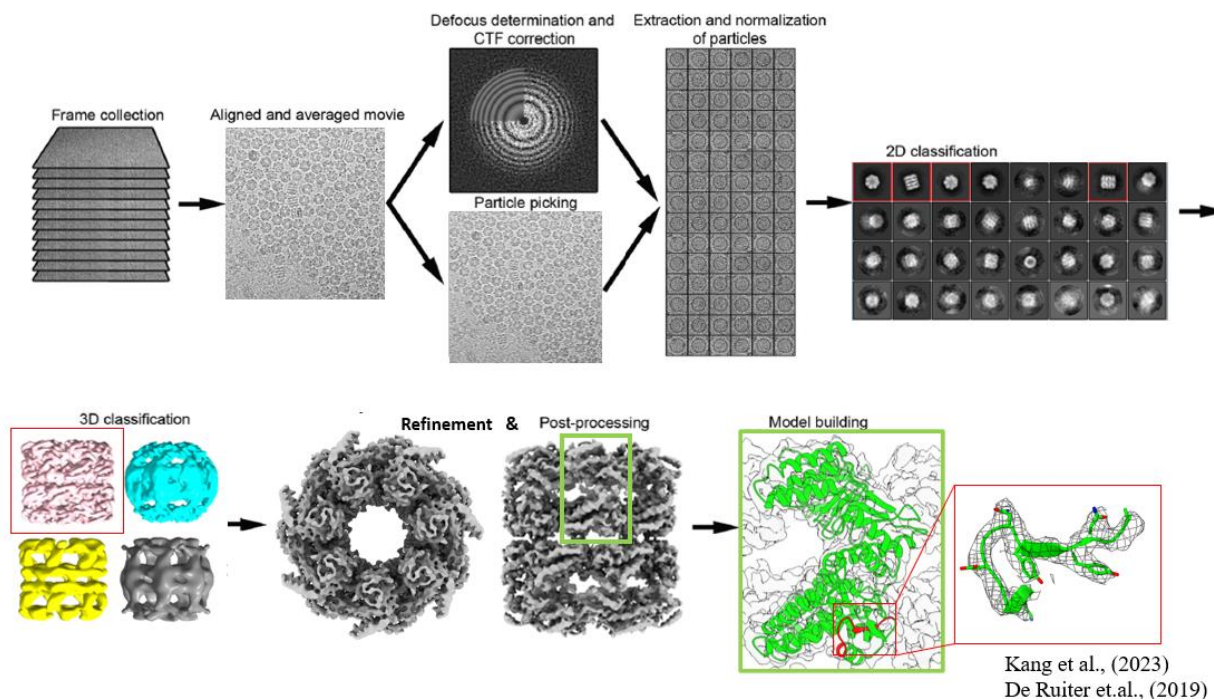
Figure 1.2: CryoET workflow
CryoET workflow of data processing and modeling

1.1.2 CryoET

The innovations of SPA using cryoEM have shed light on the structural details of biological molecules and multicomponent complexes. However, SPA largely depends on molecules that have been highly purified, engineered, or fractionated. Also, the reconstituted multicomponent complexes may lack physiologically relevant cofactors (7). Cryogenic electron tomography (cryoET) is a breakthrough technology for 3D visualization and structural investigation of cellular

"biomolecules sociology"—inter-protein interactions and spatial arrangement—in their native state (8). This approach is based on acquiring a series of images taken at different tilt angles of the sample with respect to the electron beam axis.

CryoET can be applied to a wide range of biological samples, from isolated macromolecules to multicellular organisms. Figure 1.3 offers the general workflow of cryoET. The typical workflow of cryoET is similar to that of cryoEM, including the sample preparation and preservation via vitrification. The grid is loaded onto a cryo-electron microscope to acquire images of the sample from -60° to 60° with 3° increments by tilting the specimen stage. At each tilt angle, fractionated frames are collected, forming a movie per tilt angle. Similar to cryoEM, the frames are drift-corrected and averaged with Motioncorr (4). The output is a single motion-corrected image per tilt angle. Each tilt series, containing ~40 to 41 motion-corrected images at different tilt angles, is then reconstructed into 3D tomograms by weighted back projections using the IMOD software package (9). For this reason, during the sample preparation, the sample is typically mixed with nano-sized high-contrast fiducial markers (5-20 nm gold beads), which are utilized in aligning the tilt series for tomogram reconstruction. The final tomograms can be deconvoluted and corrected for the missing-wedge artifacts with IsoNet (10) using defocus values estimated by CTFFIND4 (5) to increase the signal-to-noise ratio and reduce resolution anisotropy (Figure 1.3). CryoEM and cryoET can be combined to visualize protein structural integrity at the individual level and the protein distribution at the global scale.

1.1.2.1 Subtomogram averaging approach of cryoET

In cryoET, the imaging area is exposed multiple times to generate the tilt series. As such, the total electron dosage is ~100 e$^-$/Å^2 per imaging area, limiting the number of electrons that can be exposed per tilt angle. Single particle cryoEM, on the other hand, tolerates higher electron dose

per imaging area as the imaging area needs to be exposed once. As a result, individual cryoET images are much noisier than SPA images (11). To overcome this, subtomogram averaging (STA) can be incorporated to average features present in multiple copies, i.e., isotropic structures, within the tomograms, namely glycoproteins studded on the virion envelope (Figure 1.3). For example, 3D coordinates of glycoproteins can be picked as individual particles, similar to SPA, and extracted as subtomograms and aligned. Aligned subtomograms are grouped into classes, selected, and averaged to reinforce common signals, generating 3D reconstruction with improved signal-to-noise ratio and higher resolution than present in the original tomograms and the individual subtomograms. By combining cryoET and STA, one can visualize the virion from the architectural organization of its viral proteins to the higher resolution *in situ* structure of its membrane-bound fusion protein.
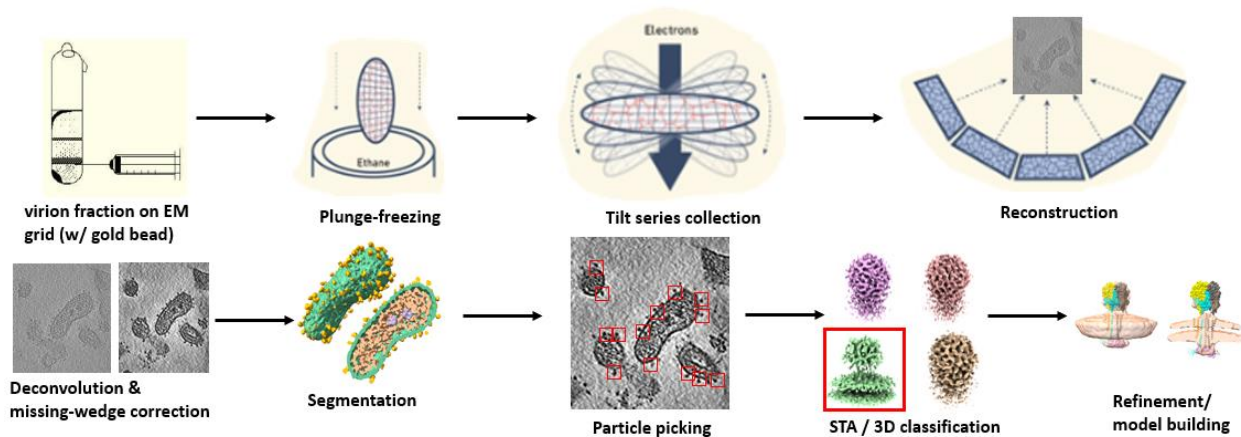
Figure 1.3: CryoET and STA workflow
CryoET imaging and STA data processing workflow

Credit: Technology Networks

1.1.3 Air-water interface adsorption phenomenon

In cryoEM, which involves blotting with filter paper to form a thin buffered layer of sample, it is often observed that embedded particles undergo aggregation and deformation at the air-water interface (AWI). CryoET revealed that a broad range of macromolecular complexes were adsorbed

to the AWI in cryoEM grid holes when these standard grid preparation steps were used (Figure 1.4) (12). This problem, also known as the AWI adsorption phenomenon, causes uneven distribution and preferred orientational positioning of the particles, resulting in directional resolution anisotropy—a major hindrance to determining high-resolution cryoEM structures (13, 14). It has been proposed that this phenomenon results from the particles' diffusion to the AWI due to their exposed hydrophobic regions (15), but a complete understanding remains elusive. This phenomenon is explored in greater depth in chapter 2 of this dissertation.

Sample # Name	Example cross-sectional schematic diagram	Sample # Name	Example cross-sectional schematic diagram	Sample # Name	Example cross-sectional schematic diagram	Sample # Name	Example cross-sectional schematic diagram
1* 32 kDa Kinase		14* Neural Receptor		27* IDE		38*† Apoferritin (0.5 mg/mL)	
4*† Hemagglutinin		17* Protein with Bound Lipids (deglycosylated)		30*† GDH		39*† Apoferritin with 0.5 mM TCEP	
5* HIV-1 Trimer Complex 1		18 Protein with Bound Lipids (glycosylated)		31*† GDH		40 Protein with Carbon Over Holes	
6* HIV-1 Trimer Complex 1		19* Lipo-protein		32*† GDH + 0.001% DDM (2.5 mg/mL)		41 Protein and DNA Strands with Carbon Over Holes	
7* HIV-1 Trimer Complex 2		20 GPCR		33*† DnaB Helicase-helicase Loader		42*† T20S Proteasome	
10* Stick-like Protein 1		21*† Rabbit Muscle Aldolase (1mg/mL)		34*† Apoferritin		43*† T20S Proteasome	
12* Stick-like Protein 2		22*† Rabbit Muscle Aldolase (6mg/mL)		35*† Apoferritin		44*† T20S Proteasome	
13* Neural Receptor		25* Protein in Nanodisc (0.58 mg/mL)		36*† Apoferritin		45*† Mtb Proteasome	
				37*† Apoferritin (1.25 mg/mL)		46 Protein on Streptavidin	

Noble et.al. 2018

Figure 1.4: AWI phenomenon demonstrated using cryoET
A selection of cross-sectional schematic diagrams of particles and ice behavior in grid holes from individual tomograms.

1.2 General structural organization of virions

A non-enveloped virus has a protein shell capsid as the outermost layer. The capsid provides protection for the viral genome. Most non-enveloped viruses consist of an outer symmetric capsid that can be a spherical icosahedron (16-18) or long helical filaments or rods (19, 20). Complex

nonenveloped viruses, namely bluetongue virus (BTV) and mammalian reovirus (MRV), have a capsid composed of multiple concentric protein layers (21, 22).

Simple enveloped viruses have the outermost lipid bilayer envelope, embedding viral proteins, and the underlying viral capsid (23, 24). The viral envelope proteins are anchored to the viral membrane through their transmembrane domains, which also have direct contact with the inner capsid (25-28). Similar to the nonenveloped viruses, the viral genome is encapsidated, and the outermost envelope and the underlying capsid layer are icosahedral.

In contrast, pleomorphic enveloped viruses, such as influenza viruses and coronaviruses (29, 30), have a matrix protein layer underneath the viral envelope that prevents the assembly of a closed symmetric protein shell (31-33). Aside from governing the overall shape and structural rigidity, the matrix protein layer serves as a scaffold to guide virion assembly and a mediator between the outer envelope and the inner nucleocapsid. Instead of a symmetric envelope protein shell, the membrane envelope is usually studded with spikes of membrane fusion glycoproteins. The inner viral nucleocapsid is usually made of the viral genome encapsidated by a virus-encoded nucleoprotein (NP) that assembles into helical filaments (34, 35).

1.2.1 Fusion mechanism of enveloped viruses

The glycoproteins of pleomorphic viruses form protruding spikes on the virion surface. Most of these glycoproteins are categorized into type I-III viral membrane fusion proteins, which the enveloped viruses must utilize to fuse with the host cell membrane to deliver their genetic material. As can be seen in Figure 1.5, the general model for this process involves the fusion pore formation between the viral and host membrane (36-42). This process is achieved by the fusion protein undergoing dramatic conformational changes (43-45).

1.2.1.1 Type I fusion protein

Type I fusion proteins, found in human immunodeficiency virus (HIV), Ebola virus (EBOV), SARS-CoV-2, influenza, and arenavirus, are usually synthesized as precursor polypeptide that is subsequently cleaved by host cell peptidases, generating the mature prefusion state trimeric glycoprotein complex (GPC) (46). The short stable peptide (SSP), which directs the intracellular trafficking of glycoprotein during its post-translational cleavage and modification, is usually degraded following the final localization of GPC (38). The transition from fusogenic trimer GPC to stable postfusion trimer GPC is triggered by receptor binding or pH change (47-49). The prefusion trimer has a head and a stem domain at the distal and proximal end, respectively (Figure 1.5). The head domain contains the receptor-binding domain that recognizes the host cell surface receptors, while the stem domain contains helices and a hydrophobic fusion peptide that directly follows the precursor peptide's cleavage site. As the GPC transitions from pre to postfusion state, the head domain detaches from the stem domain, allowing the stem helices to undergo conformational changes to form an extended coiled-coil with the fusion peptide fully exposed at its distal end (49). The fully exposed fusion peptides then anchor to the host cell membrane. The coiled-coil structures spontaneously fold back on itself to form a stable six-helix bundle (6HB). The formation of 6HB brings the virion and host membranes close, creating a fusion pore through which the viral genome enters the host cell cytoplasm (Figure 1.5) (38, 50).
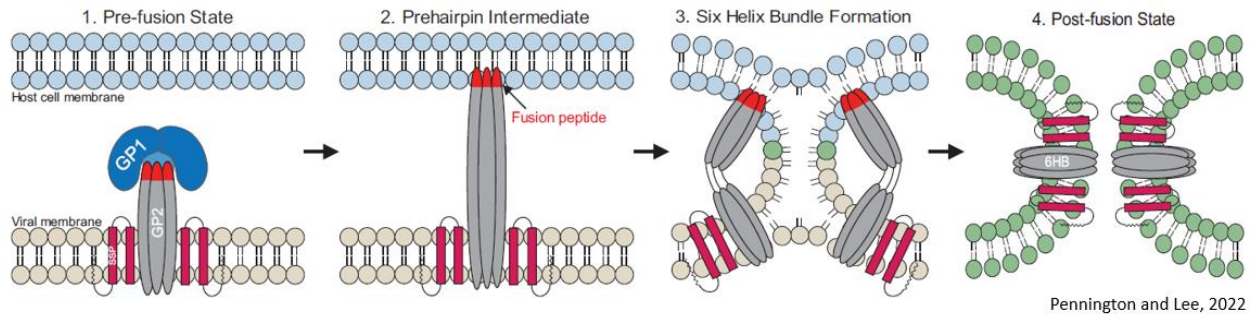
Figure 1.6: Membrane fusion illustration
Membrane fusion illustration of Lassa virus (LASV), a member of arenavirus family.

1.2.1.2 Arenavirus

Arenaviruses that infect humans belong to the Mammarenavirus genus and are divided into two groups: the Old World (OW) and the New World (NW) arenaviruses. Lassa virus (LASV), a highly pathogenic virus endemic to West Africa with high mortality that affects an estimated 100,000 to 300,000 individuals annually, and lymphocytic choriomeningitis virus (LCMV) are OW arenaviruses, while several South American hemorrhagic fever viruses, including Machupo (MACV) and Junin viruses (JUNV), are NW viruses. Chapter 3 of the dissertation focuses on LCMV, which is widely used as an experimental model for the study of viral persistence and pathogenesis of arenavirus, especially the highly pathogenic LASV. Although often asymptomatic, LCMV infections occur worldwide and can cause isolated fever, meningitis, and encephalitis (51, 52). Despite LASV's high pathogenicity, no FDA-approved vaccines or antiviral drugs are available against LASV or any other arenaviruses.

While the conformational change undergone by the arenavirus during membrane fusion follows the same framework as other class one fusion proteins, the pH-induced activation in the late endosome is unique, partly due to the unusual retention of the SSP in the mature GPC (Figure 1.5) and the SSP's direct role as a pH sensor (53, 54). Typical SSPs are 18-30 amino acids long, with a positively charged N-terminus, central hydrophobic region, and polar C-terminus that is

degraded after successful translocation of the proteins (55, 56). In contrast, the arenavirus SSP is ~58 amino acids long and is proposed to have two antiparallel transmembrane helices connected by an ectodomain loop, as shown in Figure 1.5. The association of arenavirus SSP with the glycoprotein throughout the membrane fusion process has been shown to influence the pH sensing mechanism and the overall stability of the intermediate conformation of glycoprotein (57). Despite these findings, there is an extensive knowledge gap regarding the arenavirus SSP. The structural findings that reveal the localized interactions between the SSP and the glycoprotein may be helpful in the development of new therapeutics against the arenavirus.

1.3 Dissertation outline

This thesis is organized into 4 chapters as follows: Chapter 1 provides the background on cryoEM, cryoET, AWI adsorption phenomenon, general structural organization of virions, the fusion mechanism of pleomorphic viruses, and the distinct features of the arenavirus glycoprotein. Chapter 2 provides the theoretical formulation that explains the AWI adsorption phenomenon introduced in Chapter 1 and the experimental evidence to show how different surfactants can alleviate this problem on groEL and membrane protein using cryoEM and cryoET. Chapter 3 showcases the architectural organization of viral proteins in LCMV using cryoET and the *in situ* structure of the mature GPC with its transmembrane domain intact, revealing the GP-SSP interface. Chapter 4 summarizes the work presented in the previous chapters and discusses the prospective outlook of the current work.

1.4 References

1. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

2. T. L. Blundell, A. K. Chaplin, The resolution revolution in X-ray diffraction, Cryo-EM and other Technologies. *Prog Biophys Mol Biol* **160**, 2-4 (2021).

3. M. V. de Ruiter, R. Klem, D. Luque, J. Cornelissen, J. R. Caston, Structural nanotechnology: three-dimensional cryo-EM and its use in the development of nanoplatforms for in vitro catalysis. *Nanoscale* **11**, 4130-4146 (2019).

4. X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-590 (2013).

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

6. Y. Cheng, N. Grigorieff, P. A. Penczek, T. Walz, A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438-449 (2015).

7. M. Guaita, S. C. Watters, S. Loerch, Recent advances and current trends in cryo-electron microscopy. *Curr Opin Struct Biol* **77**, 102484 (2022).

8. M. Turk, W. Baumeister, The promise and the challenges of cryo-electron tomography. *FEBS Lett* **594**, 3243-3261 (2020).

9. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76 (1996).

10. Y. T. Liu *et al.*, Isotropic reconstruction for electron tomography with deep learning. *Nat Commun* **13**, 6482 (2022).

11. K. E. Leigh *et al.*, Subtomogram averaging from cryo-electron tomograms. *Methods Cell Biol* **152**, 217-259 (2019).

12. A. J. Noble *et al.*, Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7** (2018).

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

14. E. D'Imprima *et al.*, Protein denaturation at the air-water interface and how to prevent it. *Elife* **8** (2019).

15. R. M. Glaeser, B. G. Han, Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys Rep* **3**, 1-7 (2017).

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

17. M. G. Rossmann *et al.*, Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145-153 (1985).

18. X. Dai *et al.*, In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112-116 (2017).

19. A. Klug, The tobacco mosaic virus particle: structure and assembly. *Philos Trans R Soc Lond B Biol Sci* **354**, 531-535 (1999).

20. D. W. Banner, C. Nave, D. A. Marvin, Structure of the protein and DNA in fd filamentous bacterial virus. *Nature* **289**, 814-816 (1981).

21. P. Roy, Bluetongue virus structure and assembly. *Curr Opin Virol* **24**, 115-123 (2017).

22. M. Pan *et al.*, Asymmetric reconstruction of mammalian reovirus reveals interactions among RNA, transcriptional factor micro2 and capsid proteins. *Nat Commun* **12**, 4176 (2021).

23. R. J. Kuhn *et al.*, Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717-725 (2002).

24. S. V. Pletnev *et al.*, Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* **105**, 127-136 (2001).

25. R. Zhang *et al.*, 4.4 A cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J* **30**, 3854-3863 (2011).

26. L. Chen *et al.*, Implication for alphavirus host-cell entry and assembly indicated by a 3.5A resolution cryo-EM structure. *Nat Commun* **9**, 5326 (2018).

27. S. Sun *et al.*, Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *Elife* **2**, e00435 (2013).

28. S. Mukhopadhyay *et al.*, Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* **14**, 63-73 (2006).

29. T. Noda, Native morphology of influenza virions. *Front Microbiol* **2**, 269 (2011).

30. B. W. Neuman *et al.*, Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J Virol* **80**, 7918-7928 (2006).

31. J. Peukes *et al.*, The native structure of the assembled matrix protein 1 of influenza A virus. *Nature* **587**, 495-498 (2020).

32. L. Selzer, Z. Su, G. D. Pintilie, W. Chiu, K. Kirkegaard, Full-length three-dimensional structure of the influenza A virus M1 protein and its organization into a matrix layer. *PLoS Biol* **18**, e3000827 (2020).

33.   A. J. Battisti *et al.*, Structure and assembly of a paramyxovirus matrix protein. *Proc Natl Acad Sci U S A* **109**, 13996-14000 (2012).

34.   M. Luo, J. R. Terrell, S. A. McManus, Nucleocapsid Structure of Negative Strand RNA Virus. *Viruses* **12** (2020).

35.   M. Jamin, F. Yabukarski, Nonsegmented Negative-Sense RNA Viruses-Structural Data Bring New Insights Into Nucleocapsid Assembly. *Adv Virus Res* **97**, 143-185 (2017).

36.   F. M. Hughson, Structural characterization of viral fusion proteins. *Curr Biol* **5**, 265-274 (1995).

37.   B. Eschli *et al.*, Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. *J Virol* **80**, 5897-5907 (2006).

38.   J. M. White, S. E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**, 189-219 (2008).

39.   M. Kielian, F. A. Rey, Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* **4**, 67-76 (2006).

40.   W. Wang, Z. Zhou, L. Zhang, S. Wang, G. Xiao, Structure-function relationship of the mammarenavirus envelope glycoprotein. *Virol Sin* **31**, 380-394 (2016).

41.   J. Zhu *et al.*, Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem Biophys Res Commun* **319**, 283-288 (2004).

42.   A. S. Dimitrov, S. S. Rawat, S. Jiang, R. Blumenthal, Role of the fusion peptide and membrane-proximal domain in HIV-1 envelope glycoprotein-mediated membrane fusion. *Biochemistry* **42**, 14150-14158 (2003).

43.   S. C. Harrison, Viral membrane fusion. *Virology* **479-480**, 498-507 (2015).

44.   J. Bentz, H. Ellens, D. Alford, An architecture for the fusion site of influenza hemagglutinin. *FEBS Lett* **276**, 1-5 (1990).

45.   J. M. White, G. R. Whittaker, Fusion of Enveloped Viruses in Endosomes. *Traffic* **17**, 593-614 (2016).

46.   H. N. Pennington, J. Lee, Lassa virus glycoprotein complex review: insights into its unique fusion machinery. *Biosci Rep* **42** (2022).

47.   W. Song, M. Gui, X. Wang, Y. Xiang, Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **14**, e1007236 (2018).

48.   J. Gao, M. Gui, Y. Xiang, Structural intermediates in the low pH-induced transition of influenza hemagglutinin. *PLoS Pathog* **16**, e1009062 (2020).

49.   D. J. Benton, S. J. Gamblin, P. B. Rosenthal, J. J. Skehel, Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150-153 (2020).

50.   S. Urata, J. Yasuda, Molecular mechanism of arenavirus assembly and budding. *Viruses* **4**, 2049-2079 (2012).

51.   P. B. Jahrling, C. J. Peters, Lymphocytic choriomeningitis virus. A neglected pathogen of man. *Arch Pathol Lab Med* **116**, 486-488 (1992).

52.   M. B. Mets, L. L. Barton, A. S. Khan, T. G. Ksiazek, Lymphocytic choriomeningitis virus: an underdiagnosed cause of congenital chorioretinitis. *Am J Ophthalmol* **130**, 209-215 (2000).

53.   E. L. Messina, J. York, J. H. Nunberg, Dissection of the role of the stable signal peptide of the arenavirus envelope glycoprotein in membrane fusion. *J Virol* **86**, 6138-6145 (2012).

54.   K. Briknarova, C. J. Thomas, J. York, J. H. Nunberg, Structure of a zinc-binding domain in the Junin virus envelope glycoprotein. *J Biol Chem* **286**, 1528-1536 (2011).

55.   N. Zheng, L. M. Gierasch, Signal sequences: the same yet different. *Cell* **86**, 849-852 (1996).

56.   G. von Heijne, Signal sequences. The limits of variation. *J Mol Biol* **184**, 99-105 (1985).

57.   J. York, J. H. Nunberg, Role of the stable signal peptide of Junin arenavirus envelope glycoprotein in pH-dependent membrane fusion. *J Virol* **80**, 7775-7780 (2006).

Chapter 2: Theoretical framework and experimental solution for the air-water interface adsorption problem in cryoEM

Joon S. Kang[1,2,#], Xueting Zhou[1, #, £], Yun-Tao Liu[1,3], Kaituo Wang[1,3,%], Z. Hong Zhou[1,2,3,*]

[1]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA.

[2]Molecular Biology Institute, UCLA, Los Angeles, CA 90095, USA

[3]California NanoSystems Institute, UCLA, Los Angeles, CA 90095, USA

Keywords

#These authors contributed equally to this work.

*Correspondence to Dr. Z. H. Zhou: email: Hong.Zhou@UCLA.edu; Cell Phone: 1-310-983-1033

£**Current address**: Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

%**Current address**: Department of Biomedical Sciences, University of Copenhagen, Copenhagen N, Denmark

2.1 Abstract

As cryogenic electron microscopy (cryoEM) gains traction in the structural biology community as a method of choice for determining atomic structures of biological complexes, it has been increasingly recognized that many complexes that behave well under conventional negative-stain electron microscopy tend to have preferential orientation, aggregate or simply mysteriously "disappear" on cryoEM grids. However, the reasons for such misbehavior are not well understood, limiting systematic approaches to solving the problem. Here, we have developed a theoretical formulation that explains these observations. Our formulation predicts that all particles migrate to the air-water interface (AWI) to lower the total potential surface energy — rationalizing the use of surfactant, which is a direct solution to reduce the surface tension of the aqueous solution. By performing cryogenic electron tomography (cryoET) on the widely-tested sample, GroEL, we demonstrate that, in a standard buffer solution, nearly all particles migrate to the AWI. Gradually reducing the surface tension by introducing surfactants decreased the percentage of particles exposed to the surface. By conducting single-particle cryoEM, we confirm that suitable surfactants do not damage the biological complex, thus suggesting that they might provide a practical, simple, and general solution to the problem for high-resolution cryoEM. Applying this solution to a real-world AWI adsorption problem with a more challenging membrane protein, namely, the ClC-1 channel, has resulted in to its near-atomic structure using cryoEM.

2.2 Introduction

Cryogenic electron microscopy (cryoEM) has become a tool of choice for determining the atomic structures of biological complexes. The general strategy in sample preparation involves three steps: First, the sample is placed on continuous carbon film, negatively stained, and evaluated by transmission electron microscopy (TEM) to optimize the purity and concentration of the sample. Second, a droplet of the optimized sample is applied to various cryoEM grids (*e.g.*, holey carbon or gold foil grids, graphene film grids, lacey carbon film grids, *etc*.) and blotted with filter paper to form a thin layer of liquid containing the sample. Third, the blotted grid is plunged into liquid ethane to flash-freeze the thin layer of liquid, embedding the sample in vitreous ice (1). After completing these steps, it is often observed that embedded particles of the sample (*e.g.*, Fig. 2.1) undergo aggregation and deformation at the air-water interface (AWI) (2) (Fig. 2.2 B), which can be surprising and disappointing. Cryogenic electron tomography (cryoET) revealed that a broad range of macromolecular complexes was adsorbed at the AWI in cryoEM grid holes when these standard grid preparation steps were used (3). This problem, also known as the AWI adsorption phenomenon, causes an uneven distribution and preferred orientations of the particles, resulting in directional resolution anisotropy (4)—a major obstacle in high-resolution cryoEM structure determination (5, 6). This AWI adsorption phenomenon was reported several decades ago (7-9), affecting membrane and non-membrane proteins (3). It has been proposed that this phenomenon results from the particles' diffusion to the AWI due to their exposed hydrophobic regions (10). However, a complete understanding and elimination of this phenomenon remain elusive.
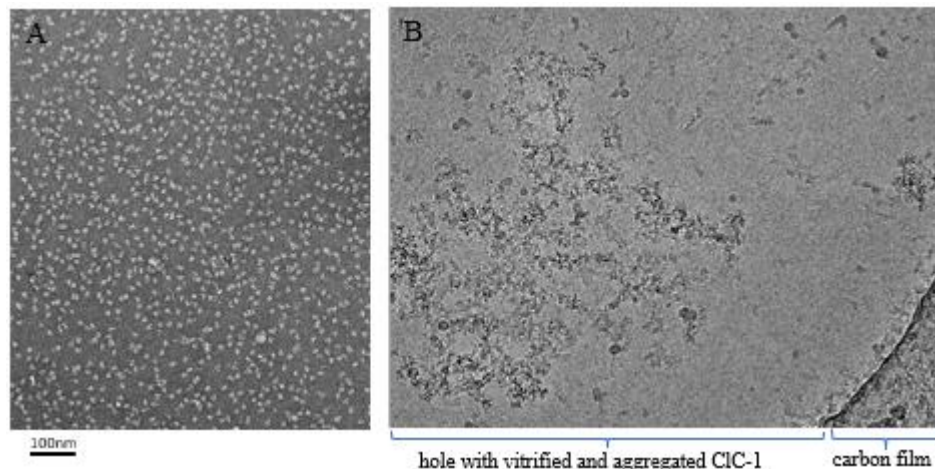
Figure 2.1: Example of the AWI adsorption problem in cryoEM
(A) Negative stain EM micrograph of purified human ClC-1 protein particles, showing intact particles and optimal particle distribution prior to vitrification. (B) CryoEM micrograph of frozen-hydrated human ClC-1 protein particles across holey carbon film, showing the typical aggregation and deformation problem occurring at the AWI in the absence of adequate surfactant.

Despite the lack of understanding, numerous attempts have been made to develop practical methods to address the problem. For example, using a faster plunge device (such as Spotiton), whose elapsed time between applying the sample to grids and plunging into liquid ethane is in the order of 100 ms (much less than the traditional plunge of more than 1s), can relieve preferred orientations problem caused by the AWI phenomenon (11, 12). Also, using continuous thin layers of amorphous carbon (13) or graphene (6, 14, 15) has been found to mitigate the AWI adsorption problem. Although these methods may improve the cryoEM result, they still have limitations. For instance, using a fast nano-dispenser does not provide direct evidence of eliminating AWI absorption problems (12). Using a thin layer of amorphous carbon as a support layer on cryoEM grids can adsorb and thus immobilize some particles, but it also adds background noise to cryoEM images (16). A more widely used method is to include surfactants or detergents in the sample solution to improve particle preservation on cryoEM grids, especially for membrane proteins. For example, zwitter-ionic detergent (CHAPSO) has been reported to reduce the AWI adsorption of

bacterial RNA polymerase (17) and human erythrocyte catalase (18), as well as alleviate the preferred orientation problem. Regrettably, a specific surfactant may prove to be effective for one sample type but fail to yield similar results for others. Therefore, a deeper understanding of the mechanism of AWI is required to better utilize surfactants in cryoEM sample preparation.

In this study, we developed a theoretical framework to explain the AWI adsorption phenomenon in the context of surface energy. We also experimentally examined the effects of different surfactants on cryoEM sample preparation using the standard test sample, GroEL, using cryoET and cryoEM. According to this framework, particles migrate to AWI to decrease the overall area of the liquid exposed to the air, thereby minimizing the overall potential energy of the system. This framework rationalizes using surfactants to alleviate the AWI problem: They reduce the surface energy of an aqueous solution, the overall potential energy, and the particles' tendency to migrate to AWI. To experimentally test the theory's predictability, we used cryoET to visualize particle distribution across the sample depth on the cryoEM grid from one AWI to the other. Our cryoET results demonstrated the effectiveness of three surfactants [NP40, DDM, and fluorinated fos-choline 8 (FFC8)] against the AWI adsorption problem. We also conducted single-particle cryoEM and 3D reconstruction with GroEL to demonstrate that these surfactants did not adversely affect the protein structures at near-atomic resolution. We further showed the effectiveness of FFC8 in drastically improving particle distribution and generating a near-atomic resolution structure of a more challenging membrane protein, the ClC-1 channel.

2.3 Results

2.3.1 Theoretical formulation of the air-water interface problem in cryoEM

All matter, except for ideal gases, is held together by molecules that exhibit varying degrees of attraction to one another. In the bulk of matter, the intermolecular forces equilibrate each other.

However, at the surface of matter, molecules are not fully surrounded by their neighbors, which results in a net inward force pointing into the bulk (19). This net inward force increases as the exposed surface area of matter increases. To achieve equilibrium, work must be done to counterbalance this net inward force; The greater the exposed surface area of matter, the greater the work to counterbalance the net inward force. The ratio between the work and surface area, or work per unit surface area, can be represented by (γ):

$$\gamma \quad = \frac{\Delta \text{Work}}{\Delta \text{Area}} = \frac{\text{Joules}}{m^2} \tag{1}$$

which is also known simply as *surface energy* in the field of surface science and fluid dynamics. Surface energy is due to the attraction of electrical charges around the molecules. Surface energy is then defined as the work required to disrupt these intermolecular attractions and pull apart molecules to a certain extent of area.

As surface energy is applied to the matter to stretch its surface, the surface of the matter counteracts the surface increase through the tangential tension force. This isotropic surface stress associated with deformation is called surface tension. The terms "surface tension" and "surface energy" refer to the same, as demonstrated in the following analysis, although they may not necessarily have the same value:

$$Surface\ energy\ (\gamma) \quad = \frac{\Delta\ \text{Work}}{\Delta\ \text{Area}} = \frac{\text{Joules}}{m^2}$$

$$= \frac{\text{Newton x m}}{m^2}$$

$$= \frac{\text{Newton}}{m}$$

$$= \frac{\Delta\ \text{Tangential tension force}}{\Delta\ \text{Length}} = Surface\ tension \tag{2}$$

For liquid, the two values are the same because as the surface deforms, the intermolecular distance at the surface does not change. As such, two terms are often used interchangeably for liquids.

23

However, for a solid, the surface molecules remain constant, so the work required for surface deformation is dependent on the intermolecular distance. As a result, this work is not the same as the work required for creating a new surface. As our paper deals with both solid and liquid surfaces, only the term "surface energy" will be used henceforth for simplicity.

The surface energy of water is one of the highest (72 J/m^2 at room temperature and atmospheric pressure) because of a strong hydrogen bond between water molecules (19). Interestingly, in a cryoEM grid, the solid particles are primarily drawn to the liquid surface, despite the high surface energy there. Such behavior is also known as the air-water interface (AWI) adsorption phenomenon. To explain this phenomenon, which at a glance, seems to defy the laws of physics, we provide time-lapsed schematics of three main stages and the mathematical equation that accounts for the area of all interfaces and the respective surface energies. The equation calculates the overall potential energy for each stage, allowing for direct comparison among the stages.

In the first stage, immediately after the sample is applied to a grid, particles are within the sample's bulk volume, suspended across a grid hole away from the AWI (Fig. 2.2A, left). Since the surface energy (γ) is defined as the work required to build a unit area, the work (in Joules) needed to build a surface is the product of γ $\left(\text{in } \frac{\text{Joules}}{\text{area}}\right)$ and the area created. Based on Navier-Stokes equations, which describe the motion of viscous fluid substances (20), the overall potential energy (Ω in Joules) for all the surfaces along the vertical axis can be calculated as the following:

$$\Omega_{\text{bulk}} = 2A_1\gamma_{\text{lg}} + 2A_2\gamma_{\text{ls}} \tag{3}$$

where γ_{lg} is the surface energy generated by the buffer solution (liquid) interfacing with the surrounding air (gas); A_1 is the liquid surface area, which is equivalent to the size of the grid hole; γ_{ls} is the surface energy at the top and bottom surfaces of the particle (solid) interfacing with the

24

buffer solution (liquid); A_2 is the top or bottom surface area of the solids, which is essentially half of the combined areas of all particles.

In the second stage, a particle adsorbs to one of the AWIs, which we call the single adsorption state (Fig. 2.2A, center). The overall potential energy of the single adsorption state can be calculated as

$$\Omega_{\text{single adsorption}} = A_1 \, \gamma_{lg} + A_2 \gamma_{ls} + A_2 \gamma_{sg} + (A_1 - A_2)\gamma_{lg}$$

$$= A_1 \, \gamma_{lg} + A_2 \gamma_{ls} + A_2 \gamma_{sg} + A_1 \gamma_{lg} - A_2 \gamma_{lg}$$

$$= 2A_1 \gamma_{lg} + A_2 \gamma_{ls} + A_2 \gamma_{sg} - A_2 \gamma_{lg} \tag{4}$$

where a new surface energy, γ_{sg}, is generated by the particle's top surface. As the particles occupy the top surface, the working surface area for γ_{lg} decreases, resulting in the reduction of the total work exerted by the surface energy. The change in the overall potential energy from $\Omega_{\text{single adsorption}}$ to Ω_{bulk}, can be expressed as,

$$\Omega_{\text{bulk}} - \Omega_{\text{single adsorption}} = A_2 \gamma_{ls} + A_2 \gamma_{sg} - A_2 \gamma_{lg} \tag{5}$$

which can be further simplified using Young's equation,

$$\gamma_{sg} = \gamma_{sl} + \gamma_{lg} \cos \theta, \tag{6}$$

where θ is the equilibrium contact angle that the liquid forms with the solid.

In the case of single adsorption, the contact angle between the liquid and solid is 180°, simplifying the equation to

$$\gamma_{sg} = \gamma_{sl} - \gamma_{lg}. \tag{7}$$

Using this relationship, the equation (5) simplifies to

$$\Omega_{\text{bulk}} - \Omega_{\text{single adsorption}} = 2A_2 \gamma_{sg}, \tag{8}$$

indicating a decrease in the overall potential energy as the system transitions from the bulk state to the single adsorption state.

Because the particle adsorption to the surface leads to a more energetically favorable state, the remaining particle side will likely follow the same path, leading to a double adsorption state (Fig. 2.2A, right). The overall potential energy of the surface of double adsorption can be expressed as

$$\Omega_{\text{double adsorption}} = 2A_1\,\gamma_{lg} + 2A_2\gamma_{sg} - 2A_2\gamma_{lg} \tag{9}$$

Comparing the potential energy between $\Omega_{\text{single adsorption}}$ and $\Omega_{\text{double adsorption}}$,

$$\Omega_{\text{single adsorption}} - \Omega_{\text{double adsorption}} = A_2\gamma_{ls} - A_2\gamma_{sg} + A_2\gamma_{lg} \tag{10}$$

Using the relationship established in equation (7), the above difference is simplified to

$$\Omega_{\text{single adsorption}} - \Omega_{\text{double adsorption}} = 2A_2\gamma_{lg} \tag{11}$$

indicating a decrease in the overall potential energy as the system transitions from the single adsorption state to the double adsorption state. This explains why the ice thickness of a properly blotted cryoEM grid tends to be the same as the particle thickness. Indeed, the adsorption phenomenon in AWI, despite the particles' migration to the surface, i.e., a high surface energy region, is logical as it generates the most energetically favorable state.

Figure 2.2: Side-view schematics of a hole of cryoEM grid, highlighting particle behavior and distribution, in relation to AWI

(A) Time-lapsed illustration of AWI adsorption of particles in the buffer. First, particles stay in the bulk volume (Left panel). A_1 denotes the surface area of the buffer layer at the AWI, i.e., the size of the grid hole. There are two A_1's, one at the top and one at the bottom. After a short period of time, particles adsorb to one of the two AWIs, reaching a more favorable energy state (one-side AWI adsorption) (Center panel). A_2

denotes the overall sum of the area of particles' surfaces exposed to the air, upon migrating to the AWI. At the top AWI, the overall A_1 decreases as the particles occupy the surface, unlike the unoccupied A_1 at the bottom AWI. Eventually, particles adsorb to the remaining AWI as the buffer layer thins, reaching the most favorable energy state (two-side AWI adsorption) (Right panel). At the bottom AWI, the overall A_1 also decreases due to the same particles occupying that surface. (B) Three common problems encountered by the sample particles at AWI, namely, particle deformation, denaturation, and preferred orientation. (C) Time-lapsed illustration of how surfactants reduce or even eliminate AWI adsorption of particles during sample preparation. First, the sample-surfactant mixture is loaded from the top side, denoted by the bulge (first panel). Surfactants migrate more quickly than the sample particle (see explanation in Section 2.3) to occupy the two AWIs, blocking sample particles from migrating to the surfaces. Upon blotting, some particles and surfactants are removed by the filter paper (second panel), exposing a second surface at the blotted side (third panel), which would also be occupied by the surfactant molecules, preventing the remaining sample particles from migrating to the free surface (last panel).
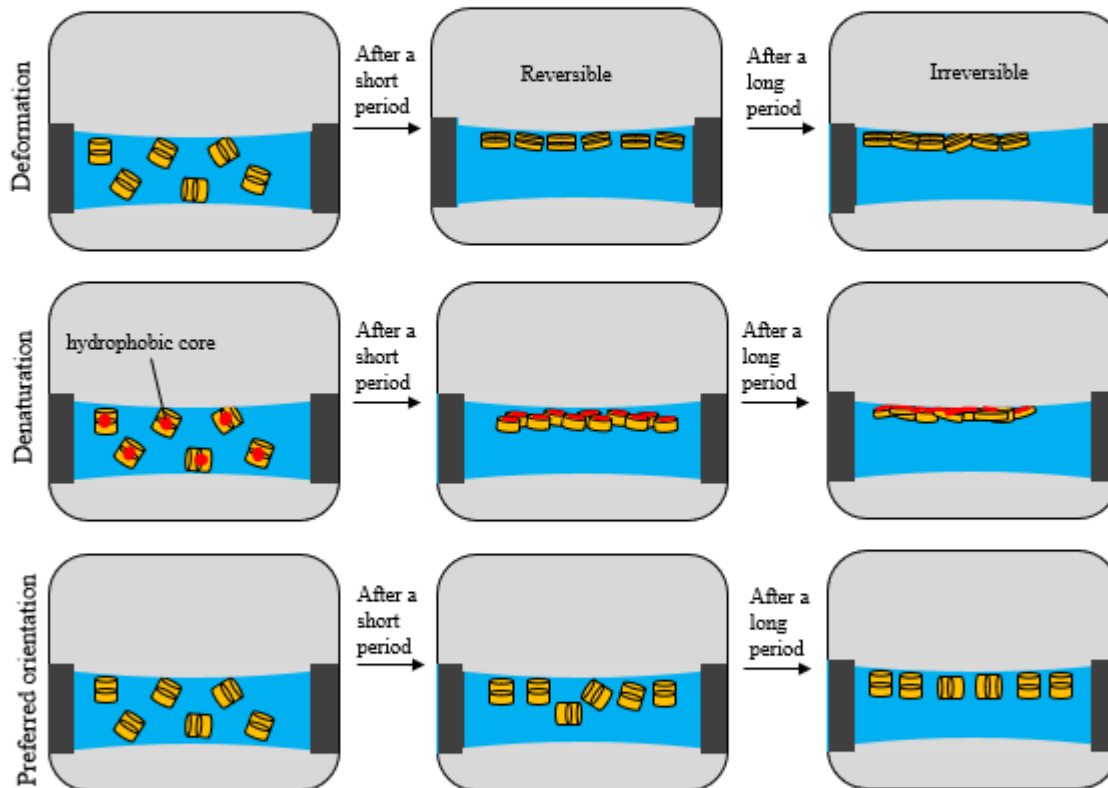
2.3.2 How is sample aggregation related to AWI

As commonly observed in electron microscopy, some particles aggregate at the AWI. Protein aggregation is the aftermath of protein denaturation that causes the exposed hydrophobic core of the individual proteins to come together (21), the event prevalent at the AWI (6, 22). To explain the connection between AWI and protein denaturation, we adopted a coarse-grained structure-based model that describes the force exerted on the particles at the AWI (23).

$$F_i^{wa} = q_i A \frac{\exp(-z_i^2/2W^2)}{\sqrt{2\pi}W} \tag{12}$$

In this model, A refers to the amplitude of the depth of the potential in the effective contact interaction between two residues. A higher value of A indicates a greater degree of pinning of a molecule to the water surface; W is the width of the interface; q_i is the hydropathy index of amino acid, ranging from -4.5 for the polar arginine to 4.5 for the most hydrophobic isoleucine. The AWI is centered at z=0, where Δ z indicates protein deformation. This phenomenological model suggests that the force exerted at the AWI denatures protein, leading to the readjustment of the protein orientation to balance the hydropathy-related forces. By comparing proteins with varying degrees of hydrophilicity, others have demonstrated that the AWI adsorption occurrence is proportional to

both the protein's hydrophobic residues and the protein deformation events (24); To equilibrate the hydropathy-related forces, the hydrophilic and hydrophobic residues of the protein pull toward the bulk and AWI, respectively, resulting in protein deformation at the AWI.

Since there are charged molecules in the sample, we also considered electrostatic forces. Toward this end, we explored the concept of Debye-Hückel length (25), which measures how far the charge-carrying species' net electrostatic effect persists in a solution. This length is defined as

$$\lambda_D = \left(\frac{\epsilon k_B T}{\sum_{j=1}^{N} n_j^0 q_j^2}\right)^{\frac{1}{2}} \tag{13}$$

where ϵ is the relative static permittivity of solvent, k_B is Boltzmann's constant, T is the temperature, n_j is the mean concentration of charges of the species j, and q_j is the charge of the species j. The Debye length in phosphate-buffered saline (PBS) at room temperature, a commonly used solution in biology and cryoEM research, is 0.7 nm (26), which is approximately the diameter of α helix secondary structure in proteins and significantly smaller than the size of folded proteins and their domains. As such, we concluded that the electrostatic force is negligible when considering molecular interactions at the AWI.

Lastly, we consider protein denaturation in relation to energy. In the energy landscape, unfolding events involve a series of small, destabilizing (uphill) steps, with small bumps (activation barriers) and dips (local minima), which are driven by the constant, reversible "sub-globally cooperative unfolding/refolding" events (10, 27) that occur at the molecular level. A native protein initially adsorbs to the AWI to reach a more energetically favorable state. After reaching a transitional state, i.e., the peak of the energy barrier, the protein spontaneously proceeds downhill and unfolds (10). Hydrogens within the native protein can exchange with solvent as the native hydrogen bonds are transiently broken. The cytochrome c (Cytc) results show five sub-

globally cooperative units, called foldons, that fold and unfold stepwise, whose amide hydrogen exchange is demonstrated by small local structural fluctuations that break one hydrogen bond at a time (27). Results from the unfolding of RNase H showed that it begins with the local fluctuations of the least stable protons on the side of the helix that is more solvent-exposed before the global unfolding. Both events place amide groups in proton exchange-competent forms in the solvent (28). These transient events expose additional hydrophobic groups of the sample, causing them to adsorb to the AWI irreversibly. With no activation barrier at the surface, a spontaneous, non-reversible "catastrophic" denaturation occurs, leading to the unfolded proteins. Exposed internal regions of these unfolded proteins are typically hydrophobic, which promotes sample aggregation, a hallmark of the AWI problem.

2.3.3 Surfactant application in cryoEM sample preparation alleviates the AWI adsorption problem by shifting the equilibrium

A surfactant is an amphipathic surface-active molecule, with its polar head in the water and non-polar tail exposed to the air. Surfactants have been widely used in cryoEM to alleviate the AWI adsorption phenomenon. Our formulation (Equation 7) predicts that all particles would migrate to the AWI as that would lower the potential energy of the system (*i.e.*, make it more negative Ω). It also predicts ways to solve or alleviate the AWI problem: by eliminating such aqueous surfaces of the system or, if impossible, by eliminating sample particles' access to the surface, and decreasing the surface energy. For the latter, surfactants can lower the surface energy (*i.e.*, γ_{lg} above), thus rationalizing the empirical use of surfactants in alleviating the AWI problem (29). Surfactants reduce the surface energy of the aqueous solution because when surfactant molecules populate the water surface, they replace the existing hydrogen bonds between water molecules (Fig. 2.2 C). As

a result, surface energy (γ_{lg}) gradually decreases until reaching the minimum at the surfactant's critical micelle concentration (CMC), the concentration above which surfactants form micelles (29). Importantly, thanks to their hydrodynamic radius, surfactants will migrate faster to the surface than the sample molecules, which typically are much larger macromolecular complexes, protecting the sample molecules from denaturation. In the extreme scenario, wherein all sites on the AWI are occupied by the surfactant molecules, the exposed surfaces accessible to sample molecules are entirely eliminated. This situation leads to an ideal condition where an aqueous surface is present for the macromolecular complexes to be visualized using cryoEM.

2.3.4 Experimental confirmation of the air-water interface adsorption problem

CryoEM sample is embedded in vitreous ice, which helps retain near-native high-resolution features of the sample. The aqueous buffer, which becomes vitreous ice through flash freezing, serves as a double-edged sword in that it is crucial for providing a "near-native" environment for the sample, but its existence introduces AWI adsorption phenomenon, causing denaturation, deformation, and preferred orientations of the sample. To address these potential problems in cryoEM sample preparation, the behavior of a widely tested non-membrane protein, GroEL, was observed. We used cryoET to reveal the 3D structure of the embedding ice in the hole of the holey carbon grids to assess the GroEL distribution throughout the grid thickness.

After generating 3D reconstructions using the IMOD package (30), we could assess the particle distribution throughout the ice thickness, as shown in Figure 2.3. The slice of tomograms (about 5nm thick) was extracted from two AWIs and the middle region of the ice. Without surfactant (Fig. 2.3A1-2.3A3), most particles were observed at the AWIs, indicating that GroEL particles, like other complexes, are prone to AWI adsorption problems. More specifically, most

31

particles were on the opposite surface (Fig. 2.3A1, Top slice) of the sample-applied surface (Fig. 2.3A3, bottom slice).

Next, we analyzed the side view from our reconstructed tomogram to better understand the particle distribution (Fig. 2.3A4-5). The degree of deviation away from the middle was expressed in percentage; The middle plane and the surfaces were set to 0% and 100%, respectively (Fig. 2.3A6). Thirty percent of the particles were in the 80~100% range, and 33% were in the 60~80% range. Taken together, our results indicated that GroEL suffers from the AWI adsorption problem.
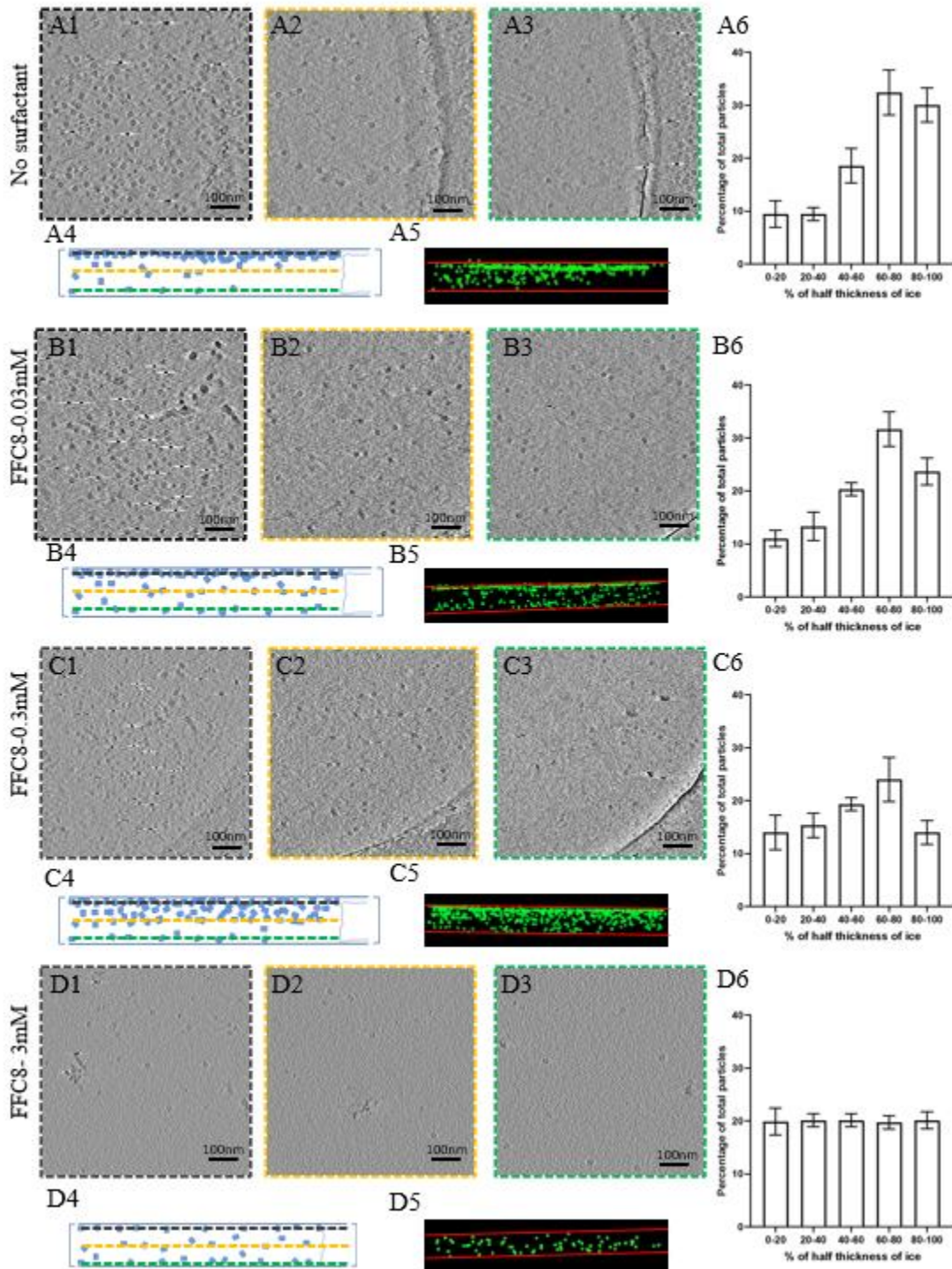
Figure 2.3: GroEL particle distribution in the vitrified sample with different concentrations of FFC8

(A-D) Each alphabetized panel (A-D) contains 6 numericized sub-sections (*e.g.*, A1 through A6 or D1 through D6), organized in the same manner for all panels. Panels A1 through A6 pertain to no-surfactant condition, and B through D pertain to different FFC8 concentrations as indicated. The first three sub-sections (1-3) show the top, center, and bottom slice of the tomogram from cryoET, respectively, viewed

33

from the top of the grid. Sub-section 4 shows the schematic of sub-sections (1-3), viewed from the side of the grid. The dashed border (black, orange, and green) in sub-sections 1-3 corresponds to the dashed horizontal line (black, orange, and green) in sub-section 4. Sub-section 5 shows the actual particle distribution of the tomogram, viewed from the side, with green dots representing GroEL particles and red horizontal lines representing the AWIs. Sub-section 6 shows the histogram of particle distribution across the entire ice thickness. X-axis is broken into ranges based on the magnitude of deviation (represented in %) from the center of the ice (*e.g.,* 0 % representing no deviation from the center of the ice and 100% representing the greatest deviation from the center, i.e., either side of AWI). Y-axis represents the number of particles (represented as the percentage of the total population) residing in each ice region.

2.3.5 Different surfactants at varying concentrations can alter particle distribution in ice.

Upon confirming the AWI absorption problem in GroEL, we experimentally validated the effectiveness of surfactants. Our first surfactant of choice was FFC8, which can help form a thinner buffer layer for cryoEM samples (31-33) than the non-fluorinated counterparts. Also, FFC8 has a relatively high CMC (3mM) compared to many other surfactants, whose CMCs are generally less than 1mM (34), alleviating the micelle-induced perturbation of image contrast. As a result, FFC8 tolerates a broader range of applicable concentrations. Lastly, FFC8 contains a polar tail that is both hydrophobic and oleophobic, reducing the likelihood of biomolecule denaturation (35).

We prepared three samples of 2 mg/mL GroEL, each with a different concentration of FFC8 (1x CMC, 1:10 CMC, and 1:100 CMC FFC8) added immediately prior to freezing the grid, and cryoET tilt series were collected and reconstructed. The number of particles in three tomogram slices (~5nm thick), including two slices for AWI layers and one slice for the middle layer, was compared among the samples in the above-mentioned concentrations (Fig. 2.3B1-2.3B3, 2.3C1-2.3C3, 2.3D1-2.3D3). As the FFC8 concentration increased, the distribution bias toward the AWI decreased (Fig. 2.3B4-2.3B6, 2.3C4-2.3C6). At CMC, particles were evenly distributed throughout the ice thickness (Fig. 2.3D4-2.3D6). The overall particle number significantly decreased at CMC, which indicated a more even particle distribution, as shown in our calculation from the previous

section. Our findings are in line with the previous study that highlighted surfactants' effectiveness in relieving the oversaturation of the field of view (36).

Aside from the fluorinated surfactant, we also tested non-ionic surfactants such as dodecyl maltoside (DDM) and NP-40, commonly used for solubilizing membrane proteins for cryoEM studies (37). Like FFC8 at CMC, the addition of NP-40 at CMC (0.30mM) and DDM at CMC (0.17mM) resulted in even particle distribution of GroEL (Fig. 2.4). These results indicated that DDM and NP40, surfactants of choice for membrane proteins, can also be effective on cytosolic proteins, such as GroEL, suggesting a broad spectrum of surfactant efficacy in improving the particle distribution. In summary, using reconstructed cryoET tomograms, we determined that the particle distribution in ice is directly related to surfactant concentration, reaching near elimination of AWI absorption phenomenon at CMC.
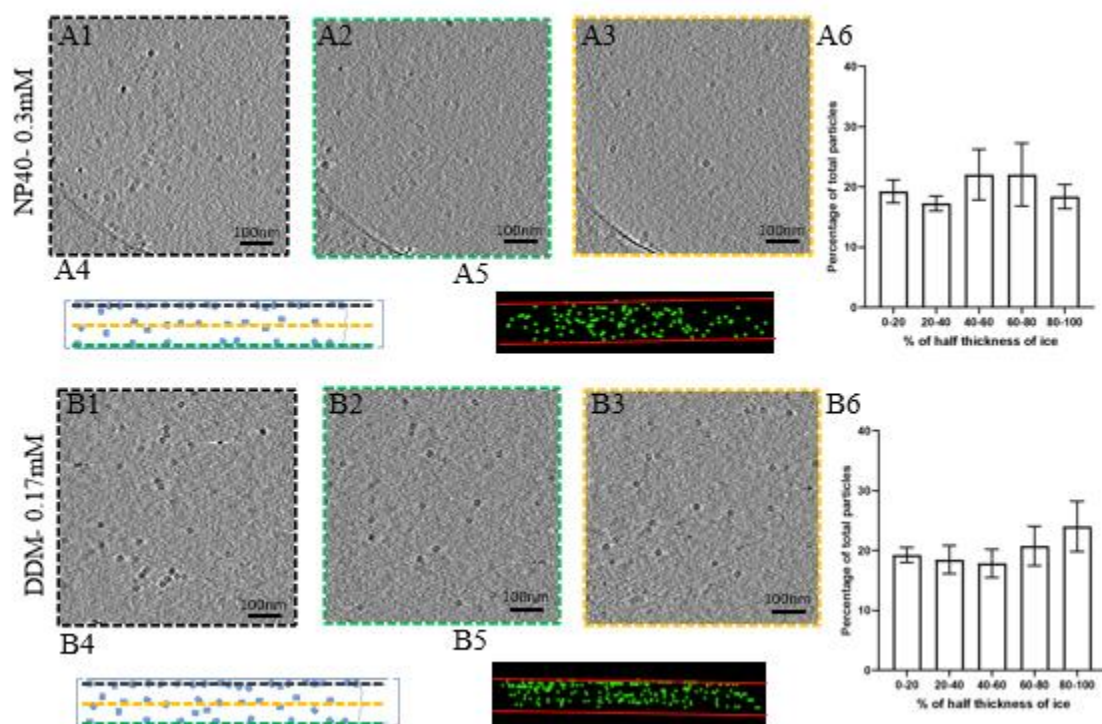

Figure 2.4: GroEL particle distributions in vitrified samples with different surfactants at their respective CMC
(A-B) Each alphabetized panel contains 6 numericized sub-sections (*e.g.*, A1 through A6 or B1 through B6), organized in the same fashion for all panels. Panel A1 through A6 pertain to NP40 at CMC condition

35

and B1 through B6 pertain to DDM at CMC condition as indicated. The first three sub-sections (1-3) show the top, center, and bottom slice of the tomogram from cryoET, respectively, viewed from the top of the grid. Sub-section 4 shows the schematic of sub-sections (1-3), viewed from the side of the grid. The dashed border (black, orange, and green) in sub-sections 1-3 corresponds to the dashed horizontal line (black, orange, and green) in sub-section 4. Sub-section 5 shows the actual particle distribution of the tomogram, viewed from the side, with green dots representing GroEL particles and red horizontal lines representing the AWIs. Sub-section 6 shows the histogram of particle distribution across the entire ice thickness. The X-axis is broken into ranges based on the magnitude of deviation (represented in %) from the center of the ice (*e.g.,* 0 % representing no deviation from the center of the ice and 100% representing the greatest deviation from the center, i.e., either side of AWI). The Y-axis shows the number of particles (represented as the percentage of the total population) residing in each ice region.

2.3.6 Impact of commonly used surfactants on high-resolution cryoEM

Surfactant at high concentration is commonly used in molecular biology research to denature protein, namely, sodium dodecyl sulfate in SDS-PAGE. However, the denaturing ability of surfactants is undesirable in structural studies. To test if the surfactants harm the protein structure, we conducted a single-particle cryoEM analysis to assess the structural integrity of GroEL in the presence of surfactants. Approximately 500 to 700 micrographs were collected for the conditions above-mentioned (at CMC for conditions with surfactants). Subsequently, 3D reconstructions derived from the extracted particles were compared (Fig. 2.5). 3D classification into four classes generated one "good" class average for all conditions, which was selected for the 3D refinement. The number of particles was normalized to 20,936 for all conditions. We were able to reach 3.3 Å, 3.5 Å, 3.5 Å, and 3.3 Å resolution for no-surfactant, DDM, FFC8, and NP40 conditions, respectively. Chain L of the X-ray crystal structure [PDB: 1MNF (38)] was fitted into the corresponding cryoEM density of all reconstructions to validate the structural integrity (Fig. 2.5) at the level of amino-acid side chains (bottom row of Fig. 2.5). Overall, these surfactants at CMC did not induce structural damage of the particles.
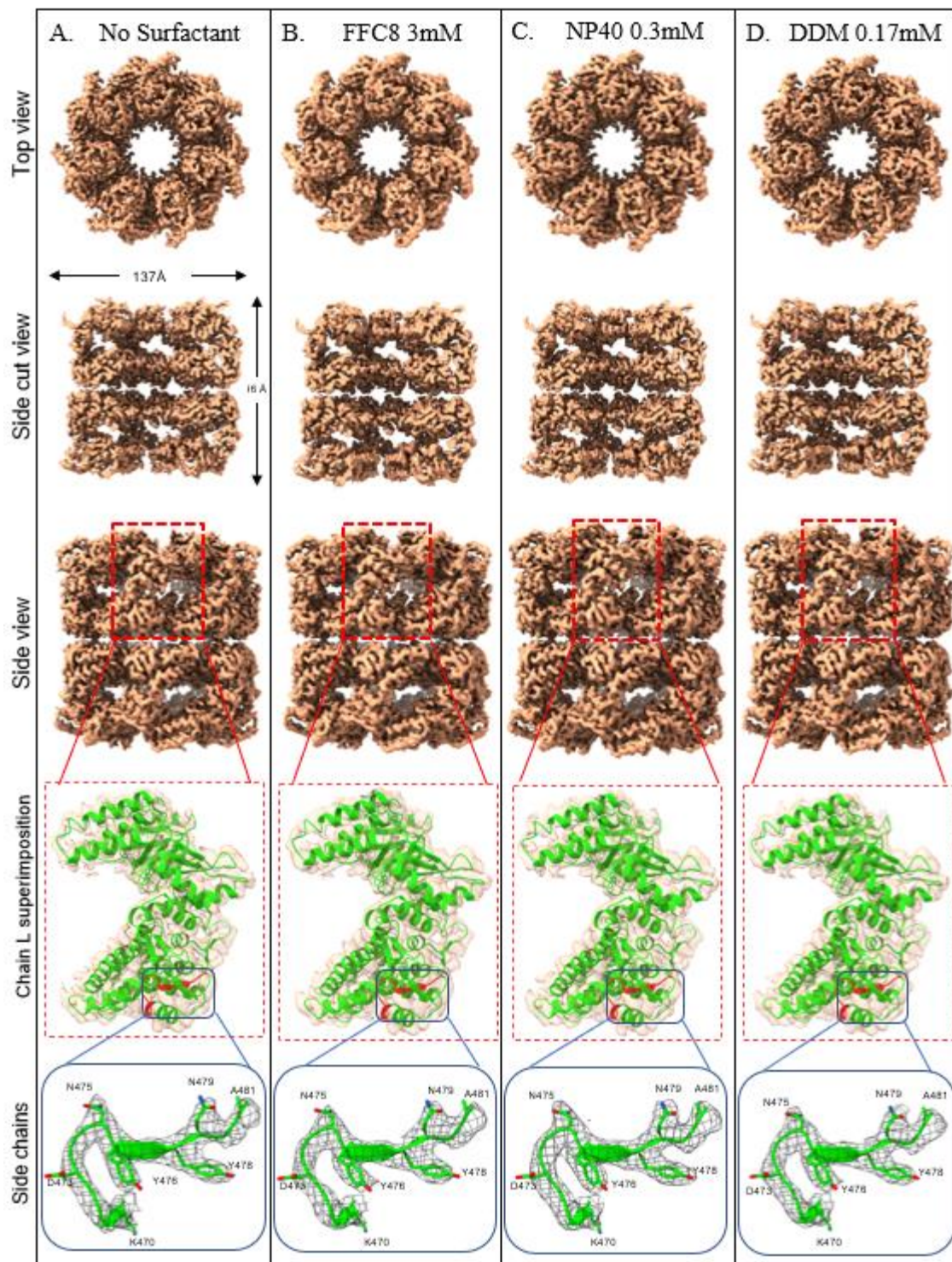
Figure 2.5: Near-atomic cryoEM reconstructions of GroEL, demonstrating no structural damage introduced by the surfactants at the indicated concentrations

(A-D) Various surface views of the cryoEM reconstructions of GroEL in different buffer conditions, in the absence of any surfactant (A) at 3.3 Å resolution, and the presence of 3.0 mM FFC8 (B) at 3.5 Å resolution, 0.3mM NP40 (C) at 3.3 Å resolution, and 0.17mM DDM (D) at 3.5 Å resolution. The 4th row shows the

semi-transparent zoom-in view of the boxed region in the 3rd row, where a ribbon diagram of chain L of the X-ray crystal structure (92) (PDB: 1MNF, shown in green ribbon representation) was fitted. The 5th row shows the zoom-in view of the boxed region in the 4th row. The structure colored in red within the box (amino acid residues 470 to 481) is now shown with the amino-acid residues and the cryoEMdensities are displayed as wireframes.

Having demonstrated that surfactants can prevent AWI adsorption while maintaining the structural integrity of a stable, soluble protein, GroEL, we investigated whether this approach could provide a solution to a more challenging, real-world problem, namely, ClC-1, as shown in Figure 1. ClC-1 is a membrane channel essential for maintaining Cl$^-$ permeability across the plasma membrane of skeletal muscle fibers, accounting for 80% of the resting membrane conductance in humans (39). Initially, the common phenomena that arose when transitioning the samples from the negative-stain EM grid to the cryoEM grid, such as particle aggregation and deformation, were also observed in ClC-1 (Fig. 2.1). The application of FFC8 at CMC (3mM), prior to vitrification, led to mono-dispersion of ClC-1 particles (Fig. 2.6A), suggesting that the samples no longer migrated to the AWI, thus preventing deformation and subsequent aggregation. This step was essential in producing good quality particles and the eventual 3.6 Å overall resolution cryoEM density map of the transmembrane domain, enabling *de novo* atomic model building (40) (Fig. 2.6B). Taken together, surfactant application is an effective and practical solution to alleviating the AWI adsorption problem in cryoEM.
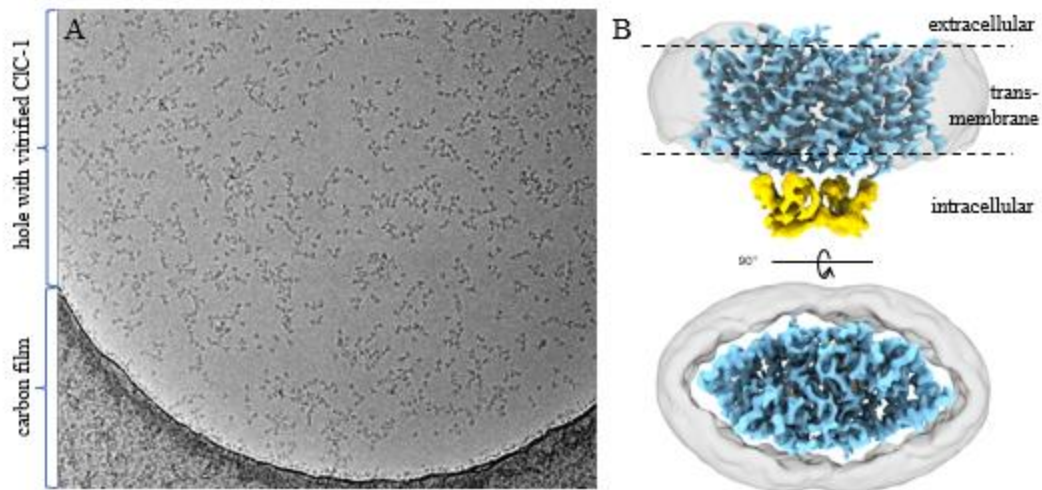
Figure 2.6: Application to real-world AWI adsorption problem

(A) A representative micrograph of single-particle cryoEM of purified human ClC-1 protein after applying FFC8 (at CMC, 3 mM). As can be compared with Figure 1, adding surfactant effectively solved the protein aggregation and deformation problem. (B) Two orthogonal shaded surface views of the resultant cryoEM map at an overall resolution of 3.6 Å, enabling *de novo* atomic modeling of the membrane domain (blue) of the complex (40). The intracellular domain is shown in yellow, and the detergent belt is in semi-transparent gray.

2.4 Discussion

As cryoEM becomes widely recognized in the structural biology community as a tool of choice for determining atomic structures of biological complexes, the cases of these complexes encountering problems of preferential orientation, aggregation, or mysterious "disappearance" on cryoEM grids—all resulting from AWI adsorption phenomenon of particles— have been well-recognized. However, the reasons for such misbehavior are not well-understood, limiting systematic approaches to addressing these problems. Here, we have proposed theoretical formulations that explain the AWI adsorption phenomenon and the associated problems. According to our formulation, the particles in the bulk solution migrate to the buffer surface to minimize the overall potential energy within the grid hole. As a result, the buffer surfaces become

fully occupied by the particles. Based on this explanation, we experimentally examined the effect of surfactant application on the problems associated with the AWI adsorption phenomenon since surfactants can lower the surface energy (γ_{lg}) by occupying the buffer surface to decrease the overall potential energy. To this end, we conducted cryoET to construct tomograms of the entire grid thickness to visualize the particle distribution in the hole of cryoEM grids. Indeed, our cryoET showed that all tested surfactants at CMC reduced the AWI adsorption of particles; FFC8 at CMC resulted in the most even distribution. Without surfactant, most particles were on the opposite surface to where the sample was applied. We conjectured that the particles began migrating to the opposite surface when the sample-applied surface was blotted, and that most particles had already finished migrating before the grid plunged, given that the particles can adsorb to the AWI in a matter of 0.1 s (2, 3, 41). Subsequent single-particle cryoEM confirmed that adding surfactant at CMC does not adversely affect the GroEL structure, as shown by near-atomic resolution reconstructions that are on par with that of the no-surfactant particles.

A question that arises naturally is what happens to the particles in the bulk volume, such as in the Eppendorf tube during long-term storage. Our theory suggests that particles would migrate to the surface until the surface is fully occupied by these particles. At this point, given the much smaller number of sites available on the surface area as compared to the bulk volume, the remaining particles would stay in the bulk volume, essentially eliminating the AWI problem for these particles. That is why the AWI adsorption problem has escaped observation until the advent of cryoEM, in which the particles are exposed to a greater surface and smaller bulk volume, a situation created by sample blotting. Our study rationalizes using surfactants in cryoEM sample preparation and experimentally demonstrates its effectiveness in lowering the surface energy (γ_{lg}) and alleviating the AWI adsorption problem. However, potential concerns associated with the use

of surfactants, such as micelle formation that can increase background noise and particle denaturation, still persist. Thus, other alternatives to surfactants are also worth exploring. For one, introducing a layer of 2D particle arrays on the AWI can be a barrier to protect other particles from AWI adsorption (10). Another alternative is the new class of surfactants called amphipols (42, 43), which enable handling membrane proteins in a detergent-free solution. Recently, a support film using 2D crystals of hydrophobin HFBI was developed. The hydrophilic side of HFBI adsorbs sample particles via electrostatic interactions to protect them from the AWI, enabling thin ice formation for enhanced data collection (44). Other advanced engineering approaches aim to improve the cryoEM result quality without the use of surfactants. For example, novel equipment has been developed to reduce the time before the sample plunges into liquid ethane (12), and innovative support film (45), like monolayer graphene-supported film and modified graphene/graphene oxides, have been developed to enhance cryoEM sample quality. More recently, *in situ* structures of ribosomes and expressomes were solved inside microbial cells via cryoET (46-48), avoiding the AWI problem altogether. However, at the time of this writing, such successes are still predominantly associated with large complexes such as ribosomes or ribosome-containing super-complexes. Consequently, high-resolution structures will continue to depend on practical solutions to resist or neutralize the potential energy at AWI, as discussed here through both theoretical consideration and experimental evidence.

2.5 Materials and methods

2.5.1 Sample preparation

Chaperonin 60 lyophilized powder (GroEL, 1 mg) from *Escherichia coli* was purchased from Sigma-Aldrich (Cat # C7688, Sigma-Aldrich, St. Louis, MO, USA). GroEL powder was solubilized in 50mM Tris-HCl (pH7.5), 10mM KCl, an 10mM MgCl$_2$. Subsequently, GroEL

concentration was determined by NanoDrop Microvolume UV-Vis Spectrophotometers based on Protein A280. For short-term storage, 2mg/mL GroEL was stored at 4 °C and used within 2 weeks. For long-term storage, stock solutions were stored at -80 °C and thawed before use.

2.5.2 Grids preparation for cryoEM

GroEL solution was prepared as 2mg/mL and mixed with 1:70 diluted 5-nm diameter fiducial gold beads in an 8:1 volume ratio. FEI Vitrobot Mark IV was used to make vitrified samples. The sample was applied to the carbon side of 200 mesh Cu Quantifoil 100 holey carbon films (R 2/1), which were beforehand glow discharged by Gatan Plasma System SOLARUS. Tested surfactants (concentrations shown in Table 1) were directly mixed with the sample right before freezing. The mixture was applied to the grid. The grids were blotted with filter paper to remove the extra sample and then plunged into liquid ethane. Grids were stored in liquid nitrogen.

2.5.3 Single-particle cryoEM data collection

The grids were loaded into a Titan Krios electron microscope (Thermo Fisher Scientific) equipped with a Gatan imaging filter (GIF), and cryoEM images were recorded on a post-GIF Gatan K3 Summit direct electron detection camera operated in super-resolution electron-counting mode. The magnification was 81,000 ×, with a pixel size of 1.1038 Å/pixel at the specimen level. Data collection was facilitated by SerialEM (103). The dosage rate was set to 30 electrons/Å2 on the sample level, and the exposure time for each frame was 0.2 seconds. The targeted under-focus value was 1.8 μm – 2.2 μm. In total, 500-700 micrographs were collected.

2.5.4 Single-particle cryoEM data processing

We followed the computational steps under the framework of WARP (50) for the preprocessing, which included motion correction for the frame alignment and estimation of local defocus and resolution. Particle-picking was performed based on a machine-learning algorithm (BoxNet) in WARP, generating, on average, 173,000 extracted particles with a box size of 400 pixels (pixel size: 1.1038 Å) for all conditions. The extracted particles were classified with the "2D Classification" tool on the Relion GUI with the following parameters: Number of classes: 32, Regularisation parameter T: 2, Number of iterations: 25, Mask diameter: 300 Å. After 2D classification, "good" class averages with clear top/side views of GroEL, as well as those that fit the general dimensions of GroEL were manually selected for 3D classification, using "Subset selection" tool on the Relion GUI. The 3D classification was conducted with the following parameters: Symmetry: D7, Number of classes: 4, Regularisation parameter T: 4, Number of iterations: 25, Mask diameter: 230 Å. After 3D classification, "good" class averages were manually selected for 3D refinement using the "Subset selection" tool on the Relion GUI. Particle images belonging to the selected "good" class averages were normalized to 20,936 particles. The normalized particles were subjected to a D7 symmetry refinement with a mask diameter of 200 Å, using "3D auto-refine" on the Relion GUI. Subsequent fitting of the crystal structure and images were created in Chimera X (51).

2.5.5 CryoET data collection

Single particle micrographs were collected on the FEI Tecnai TF20 at 200 kV equipped with a TIETZ F415MP 16-megapixel CCD camera. TEM Imaging & Analysis (TIA) was used to acquire data. All tilt series were collected from -50° to +50° with 2° increments using FEI TEM Batch Tomography software with -6 μm nominal defocus. The cumulative dose count was 50~60 e-/Å2

per tilt series. The pixel size was 4.4 Å/pixel at the specimen level with the 50,000× nominal magnification for imaging.

2.5.6 CryoET data processing

The tilt series were reconstructed by the Etomo component of the IMOD software package (30) to 3D tomograms. "Build Tomogram" was chosen to start the reconstruction process. Tilt series images were pre-processed by removing the outlier pixel values in the data files. Coarse alignment was done using the fiducial seeding model, followed by a fine alignment. 10 to 15 gold beads were picked for each tilt series, and a mathematical model for specimen movements was used to predict the gold beads' positions. The mean residual error was reduced during fine alignment by fixing big residuals. Positioning tomogram thickness was set to 1000 nm to include the top and bottom AWI. The tilt axis and Z shift were also computed and adjusted to create the final alignment. The final tomograms were built using SIRT with 6 iterations. Finally, the whole tomogram was rotated around the x-axis to make the air-water interface roughly parallel to the field of view.

GroEL particles were selected and saved as a mod file. Three points on the same plane were used to identify the AWI and the middle plane. The coordinates of all particles were recorded and used to calculate the distance to each plane. Particle distribution was visualized as a 3D model by IMOD.

2.6 Acknowledgments

2.7 Author Contributions

ZHZ initialized and supervised the project; XZ performed cryoET, and JSK and XZ processed the data; JSK, XZ, YL, and KW performed cryoEM imaging; JSK performed single-particle reconstruction; ZHZ, JSK, and XZ wrote the paper; all authors edited and approved the paper.

2.8 Competing Interests

The authors declare no competing interests.

2.10 References

1.   J. Dubochet *et al.*, Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129-228 (1988).

2.   K. A. Taylor, R. M. Glaeser, Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *Journal of Structural Biology* **163**, 214-223 (2008).

3.   A. J. Noble *et al.*, Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7** (2018).

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

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

6.   E. D'Imprima *et al.*, Protein denaturation at the air-water interface and how to prevent it. *eLife* **8**, e42747 (2019).

7.   D. E. Graham, M. C. Phillips, Proteins at liquid interfaces: I. Kinetics of adsorption and surface denaturation. *Journal of Colloid and Interface Science* **70**, 403-414 (1979).

8.   F. MacRitchie, Desorption of proteins from the air/water interface. *Journal of Colloid and Interface Science* **105**, 119-123 (1985).

9.   G. Narsimhan, F. Uraizee, Kinetics of adsorption of globular proteins at an air-water interface. *Biotechnology Progress* **8**, 187-196 (1992).

10.   R. M. Glaeser, B.-G. Han, Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophysics Reports* **3**, 1-7 (2017).

11. T. Jain, P. Sheehan, J. Crum, B. Carragher, C. S. Potter, Spotiton: a prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *J Struct Biol* **179**, 68-75 (2012).

12. A. J. Noble *et al.*, Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat Methods* **15**, 793-795 (2018).

13. R. C. Williams, R. M. Glaeser, Ultrathin carbon support films for electron microscopy. *Science* **175**, 1000-1001 (1972).

14. R. S. Pantelic *et al.*, Graphene: Substrate preparation and introduction. *J Struct Biol* **174**, 234-238 (2011).

15. C. J. Russo, L. A. Passmore, Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nat Methods* **11**, 649-652 (2014).

16. C. J. Russo, L. A. Passmore, Progress towards an optimal specimen support for electron cryomicroscopy. *Curr Opin Struct Biol* **37**, 81-89 (2016).

17. J. Chen, A. J. Noble, J. Y. Kang, S. A. Darst, Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *Journal of Structural Biology: X* **1**, 100005 (2019).

18. S. Chen, J. Li, K. R. Vinothkumar, R. Henderson, Interaction of human erythrocyte catalase with air-water interface in cryoEM. *Microscopy (Oxford, England)* **71**, i51-i59 (2022).

19. I. M. Hauner, A. Deblais, J. K. Beattie, H. Kellay, D. Bonn, The Dynamic Surface Tension of Water. *The Journal of Physical Chemistry Letters* **8**, 1599-1603 (2017).

20. R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis* (AMS Chelsea Pub., 2001).

21.　E. Koepf, R. Schroeder, G. Brezesinski, W. Friess, The film tells the story: physical-chemical characteristics of IgG at the liquid-air interface. *European Journal of Pharmaceutics and Biopharmaceutics* **119**, 396-407 (2017).

22.　J. Wiesbauer, R. Prassl, B. Nidetzky, Renewal of the Air–Water Interface as a Critical System Parameter of Protein Stability: Aggregation of the Human Growth Hormone and Its Prevention by Surface-Active Compounds. *Langmuir* **29**, 15240-15250 (2013).

23.　M. Cieplak, D. B. Allan, R. L. Leheny, D. H. Reich, Proteins at Air–Water Interfaces: A Coarse-Grained Model. *Langmuir* **30**, 12888-12896 (2014).

24.　Y. Zhao, M. Chwastyk, M. Cieplak, Topological transformations in proteins: effects of heating and proximity of an interface. *Scientific Reports* **7**, 39851 (2017).

25.　B. J. Kirby, *Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices* (Cambridge University Press, 2010).

26.　C. H. Chu *et al.*, Beyond the Debye length in high ionic strength solution: direct protein detection with field-effect transistors (FETs) in human serum. *Sci Rep* **7**, 5256 (2017).

27.　H. Maity, M. Maity, M. M. G. Krishna, L. Mayne, S. W. Englander, Protein folding: The stepwise assembly of foldon units. *Proceedings of the National Academy of Sciences* **102**, 4741-4746 (2005).

28.　A. K. Chamberlain, T. M. Handel, S. Marqusee, Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **3**, 782-787 (1996).

29.　W.-C. Liao, J. L. Zatz, Surfactant solutions as test liquids for measurement of critical surface tension. *Journal of Pharmaceutical Sciences* **68**, 486-488 (1979).

30. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76 (1996).

31. R. M. Glaeser, Proteins, Interfaces, and Cryo-Em Grids. *Curr Opin Colloid Interface Sci* **34**, 1-8 (2018).

32. T. E. T. Hughes *et al.*, Structural basis of TRPV5 channel inhibition by econazole revealed by cryo-EM. *Nat Struct Mol Biol* **25**, 53-60 (2018).

33. Z. L. Johnson, J. Chen, Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. *Cell* **168**, 1075-1085 e1079 (2017).

34. Â. S. Inácio *et al.*, In vitro surfactant structure-toxicity relationships: implications for surfactant use in sexually transmitted infection prophylaxis and contraception. *PLoS One* **6**, e19850-e19850 (2011).

35. K. H. Park *et al.*, Fluorinated and hemifluorinated surfactants as alternatives to detergents for membrane protein cell-free synthesis. *Biochem J* **403**, 183-187 (2007).

36. S.-H. Roh *et al.*, Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. *Proceedings of the National Academy of Sciences* **114**, 8259-8264 (2017).

37. K. R. Vinothkumar, Membrane protein structures without crystals, by single particle electron cryomicroscopy. *Current opinion in structural biology* **33**, 103-114 (2015).

38. J. Wang, L. Chen, Domain Motions in GroEL upon Binding of an Oligopeptide. *Journal of Molecular Biology* **334**, 489-499 (2003).

39. T. Stauber, S. Weinert, T. J. Jentsch, Cell biology and physiology of CLC chloride channels and transporters. *Compr Physiol* **2**, 1701-1744 (2012).

40. K. Wang *et al.*, Structure of the human ClC-1 chloride channel. *PLoS Biol* **17**, e3000218 (2019).

41. K. Naydenova, C. J. Russo, Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature Communications* **8**, 629 (2017).

42. M. Liao, E. Cao, D. Julius, Y. Cheng, Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112 (2013).

43. C. Tribet, R. Audebert, J.-L. Popot, Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proceedings of the National Academy of Sciences* **93**, 15047 (1996).

44. H. Fan *et al.*, A cryo-electron microscopy support film formed by 2D crystals of hydrophobin HFBI. *Nature Communications* **12**, 7257 (2021).

45. Y. Han *et al.*, High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proc Natl Acad Sci U S A* **117**, 1009-1014 (2020).

46. D. Tegunov, L. Xue, C. Dienemann, P. Cramer, J. Mahamid, Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells. *Nat Methods* **18**, 186-193 (2021).

47. F. J. O'Reilly *et al.*, In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554-557 (2020).

48. P. C. Hoffmann *et al.*, Structures of the eukaryotic ribosome and its translational states in situ. *Nature Communications* **13**, 7435 (2022).

49. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).

50.   D. Tegunov, P. Cramer, Real-time cryo-electron microscopy data preprocessing with

        Warp. *Nat Methods* **16**, 1146-1152 (2019).

51.   T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and

        analysis. *Protein Sci* **27**, 14-25 (2018).

Chapter 3: Architectural organization and *in situ* fusion protein structure of lymphocytic choriomeningitis virus

Joon S. Kang[1,2,3], Kang Zhou[1], Hui Wang[1,3], Sijia Tang[4], Kristin Van Mouwerik Lyles[5], Ming Luo[5,6], and Z. Hong Zhou[1,2,3,*]

1. California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA

2. Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

3. Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

4. Institute of Biomedical Sciences, Georgia State University, Atlanta, GA 30302, USA

5. Department of Chemistry, Georgia State University, Atlanta, GA 30302, USA

6. Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, USA

*Corresponding author: Hong.Zhou@UCLA.edu; phone: 1-310-694-7527

KEYWORDS   Arenavirus, lymphocytic choriomeningitis virus, cryogenic electron tomography, in situ structures, spike proteins, virion, prefusion

3.1 Abstract

Arenaviruses exist globally and can cause hemorrhagic fever and neurological diseases, exemplified by the zoonotic pathogen lymphocytic choriomeningitis virus (LCMV). The structures of individual LCMV proteins or their fragments have been reported, but the architectural organization and the nucleocapsid assembly mechanism remain elusive. Importantly, the *in situ* structure of arenavirus fusion protein complex GPC as present on the virion prior to fusion, particularly with its integral stable signal peptide (SSP), has not been shown, hindering efforts such as structure-based vaccine design. Here, we have determined the *in situ* structure of LCMV proteins and their architectural organization in the virion by cryogenic electron tomography (cryoET). The tomograms reveal the global distribution of GPC, matrix protein Z, and the contact points between the viral envelope and nucleocapsid. Subtomogram averaging yielded the *in situ* structure of the mature GPC with its transmembrane domain intact, revealing the GP2-SSP interface and the endodomain of GP2. The number of RNA-dependent RNA polymerase (RdRp) L molecules packaged within each virion varies, adding new perspectives to the infection mechanism. Together, these results delineate the structural organization of LCMV and offer new insights into its mechanism of LCMV maturation, egress, and cell entry.

3.2 Importance

The impact of COVID-19 on public health has highlighted the importance of understanding zoonotic pathogens. Lymphocytic choriomeningitis virus (LCMV) is a rodent-borne human pathogen that causes hemorrhagic fever. Herein, we describe the *in situ* structure of LCMV proteins and their architectural organization on the viral envelope and around the nucleocapsid.

The virion structure reveals the distribution of surface glycoprotein complex (GPC) and the contact points between the viral envelope and the underlying matrix protein, as well as the association with the nucleocapsid. The morphology and sizes of virions, as well as the number of RNA polymerase L inside each virion vary greatly, highlighting the fast-changing nature of LCMV. A comparison between the *in situ* GPC trimeric structure and prior ectodomain structures identifies the transmembrane and endo domains of GPC and key interactions among its subunits. The work provides new insights into LCMV assembly and can inform future structure-guided vaccine design.

3.3 Introduction

Lymphocytic choriomeningitis virus (LCMV), a rodent-borne human pathogen, is a member of the Arenaviridae family in the Bunyavirales order. Globally distributed by the dispersion of the rodent reservoir, LCMV is a widely used experimental model for viral pathogenesis studies (1, 2). Although typically asymptomatic in healthy individuals, LCMV infection poses a threat to immunosuppressed patients, children, and pregnant women. For example, LCMV can cause aseptic meningitis in immunocompromised individuals, while prenatal infection can cause abortion or severe congenital disabilities (3, 4).

LCMV is a pleomorphic enveloped virus that contains segmented negative-strand RNA genomes with one large (L) segment of ~7.5 kb and one small (S) segment of ~3.5 kb. Each segment contains two open reading frames (ORF) encoding two gene products in opposite orientations, allowing for the expression of early and late genes (5). The L segment encodes the large RNA-dependent RNA polymerase (RdRp) (L, 200 kDa) and the multi-functional matrix protein (Z, 11 kDa). The S segment encodes the viral nucleoprotein (NP, 63 kDa) and precursor glycoprotein complex (GPC, 75 kDa). Since its RNA is negative sense, LCMV relies on nascent

RdRp for its genomic replication. Because their ORFs are in a 5`-3` orientation, L and NP are produced first as early products. These early gene products replicate the viral gene to make the late gene products, Z and GPC. Initially expressed as a single polypeptide, GPC is translocated to the endoplasmic reticulum, where a stable signal peptide (SSP, aa 1-58) is cleaved from its N terminus by the host Spase (6). Then, GPC is translocated to Golgi, where it is additionally cleaved by the host convertase subtilisin kexin isozyme 1 (SKI-1)/site 1 protease (S1P), yielding peripheral glycoprotein 1 (GP1, aa 59-265) and transmembrane glycoprotein 2 (GP2, aa 266-498). GP1 is responsible for receptor engagement (7-10), and GP2 is responsible for membrane fusion (11-13). LCMV GPC belongs to the class one viral fusion protein, which also includes GPCs of human immunodeficiency virus (HIV), Ebola virus (EBOV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Because these viruses are enveloped, they must fuse with the host membrane to release their genetic material through the fusion pore (6, 14).

When the class one fusion protein concludes its final translocation to the membrane, its stable signal peptide (SSP) is generally degraded and does not remain as part of the mature GPC (15). However, compared to other class one fusion proteins, the LCMV SSP remains associated with the mature GPC—a unique feature of arenaviruses. After its cleavage at the ER, SSP is myristoylated and rearranged to translocate its C-terminus tail to the cytosolic side of the membrane (16, 17). This allows SSP's interaction with GP2 and retention in the mature GPC (18). In addition to its conventional role in directing intracellular trafficking, arenavirus SSP is an integral third structural subunit, alongside GP1 and GP2, of mature GPC necessary for the proper maturation of GPC and pH-dependent fusion process (19-22). Arenavirus SSP is proposed to traverse the membrane twice, with two antiparallel transmembrane helices connected

55

by an eight-amino acid ectodomain loop (23, 24), but the complete structure of mature GPC with SSP intact remains unknown (12, 25, 26).

To address this, we obtained 3D structure of the whole virion using cryogenic electron tomography (cryoET) and the structure of the prefusion state of LCMV GPCs by conducting subtomogram averaging (STA). We reconstructed a 3D model of the mature GPC with its transmembrane region intact, which reveals the GP2-SSP interface and GP2 endodomain. We also used our whole virion tomogram to fit the previously resolved models of individual proteins (12, 25-35) to establish the architectural framework of the virion and unveil the interactions among GPC, Z, NP, and L proteins. Together, these findings provide a roadmap to understanding the structural organization of LCMV and its mechanism of maturation, egress, and cell entry.

3.4 Results

3.4.1 Structural organization of the LCMV viral proteins

In the virion, L and S segments of the viral genome are encapsidated by NP in the nucleocapsids. Each nucleocapsid is associated with the L protein, and both nucleocapsids are copackaged via interactions with membrane-associated Z proteins that also interact with GPs studded in the viral envelope. Structures of individual proteins from Arenavirus have been solved previously by X-ray crystallography and cryo-electron microscopy (cryoEM) (12, 25-35). However, the architectural organization of these proteins in the virion and the assembly mechanism of NP and RNA into the nucleocapsid are poorly understood. To address this, we used cryoET to reconstruct the 3D model of the LCMV virion at molecular resolution. The simultaneous iterations reconstruction technique (SIRT)-like implementation was used to analyze the tomograms of virion particles (Figure 3.1A). The tomograms were denoised and corrected for missing-wedge artifacts to improve the visualization of the low-resolution features (Figure 3.1B).

A virion of interest was segmented from the corrected tomogram to generate a complete map of a fully intact virion, revealing the external and internal structural organization (Figure 3.1C, 3.1D). The external view features the pleomorphic viral envelope and the random distribution of GPs throughout the envelope. The fitting of the GP model (PDB 8DMI) suggests that our virion carries prefusion state GPCs (Figure 3.1E). Internally, the nucleocapsid is mostly packed randomly, although the periphery forms a structured layer underneath the viral envelope. At the core, two L polymerases are wrapped around the nucleocapsids (Figure 3.1D).

Viral protein interactions were examined in greater detail by segmenting the virion map into regions of GP, viral envelope, nucleocapsids, and L polymerases. The most prominent type of interface was found between the viral envelope layer and the underlying nucleocapsid layer (Figure 3.1E). These interfaces were displayed as small pieces of densities that bridge the two layers. These densities seem to provide the anchorage for the nucleocapsids and the viral envelope, conferring structural rigidity to the nucleocapsids. To better discern these densities and those underneath, the viral envelope was removed, and previously resolved structure models were fitted into the remaining density (Figure 3.1E, F). The Z monomer (PDB 5I72) fitted relatively well into the bridging density. The density underneath the Z proteins was fitted with the NP monomers (PDB 3T5N). The NP monomers directly contacting the viral envelope are spaced apart (~13 Å) and stacked in an organized manner (Figure 3.1F, G). Moving away from the viral envelope, the nucleocapsids become less orderly and more tightly packed.
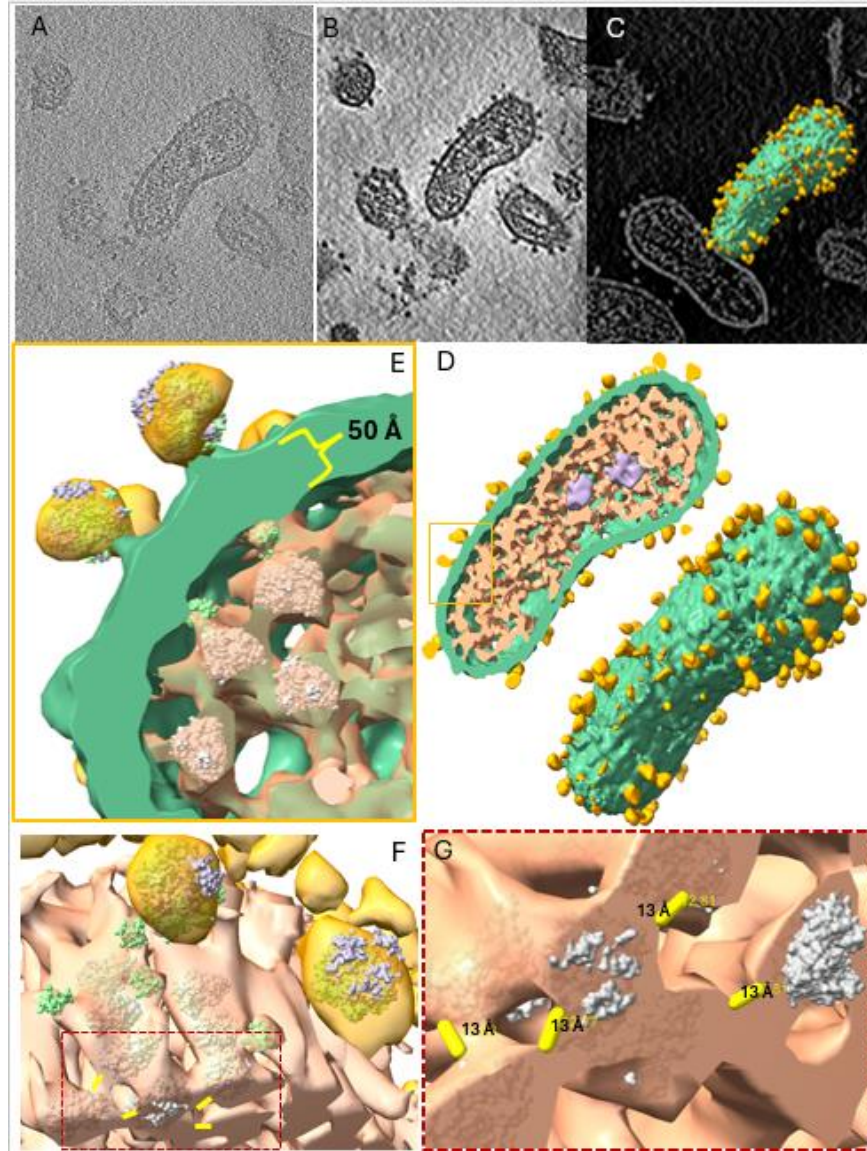
Figure 3.1: The architectural organization of LCMV: GP-Z-NP interface and peripheral NP arrangement
(A) XY density slice view of a cryoET tomogram containing a virion of interest. (B) XY slice view of the tomogram from (A) after deconvolution and missing wedge correction. (C) Chimera slice view of (B) with the segmented virion of interest. (D) Right: segmentation map of the virion from (C) Left: Cut-open view of the virion. (E) Zoom-in view of the yellow box region in (D) with prefusion trimeric ectodomain of GP (8DMI), Z (5I72), NP (monomer of 3T5N) fitted into the density. The viral envelope thickness indicated in yellow (F) Slightly rotated along the x-axis view of (E) with the viral envelope removed (G) Zoom-in view of the dotted burgundy red box region in (F) with the distance between the NPs indicated in yellow using tape tool in Chimera. Gold: GP, Green: Viral envelope, Peach: Nucleocapsid, Purple: L

58

Next, the interaction between the nucleocapsid and L was examined more closely. To begin, two dense regions shown in the denoised tomogram were segmented, as they were determined to represent the L polymerases (Figure 3.2A, inset cyan region). Then, the innermost layer surrounding the aforementioned segmented region was extracted, and the previously resolved L-Z complex (PDB 7X6V) was fitted into the segmented region (Figure 3.2B). The first interesting finding is that both dense regions are larger than the L-Z complex, indicating that the region contains additional proteins that may be associated with the L-Z. Another finding is that there are empty spaces (~50 to 90 Å) between the core complex and the nearest nucleocapsid layer around the core complex (Figure 3.2B, C). While there exists an empty space between the core complex and the nucleocapsid, there are also tight contacts between the two structures, locking the core complex in place.
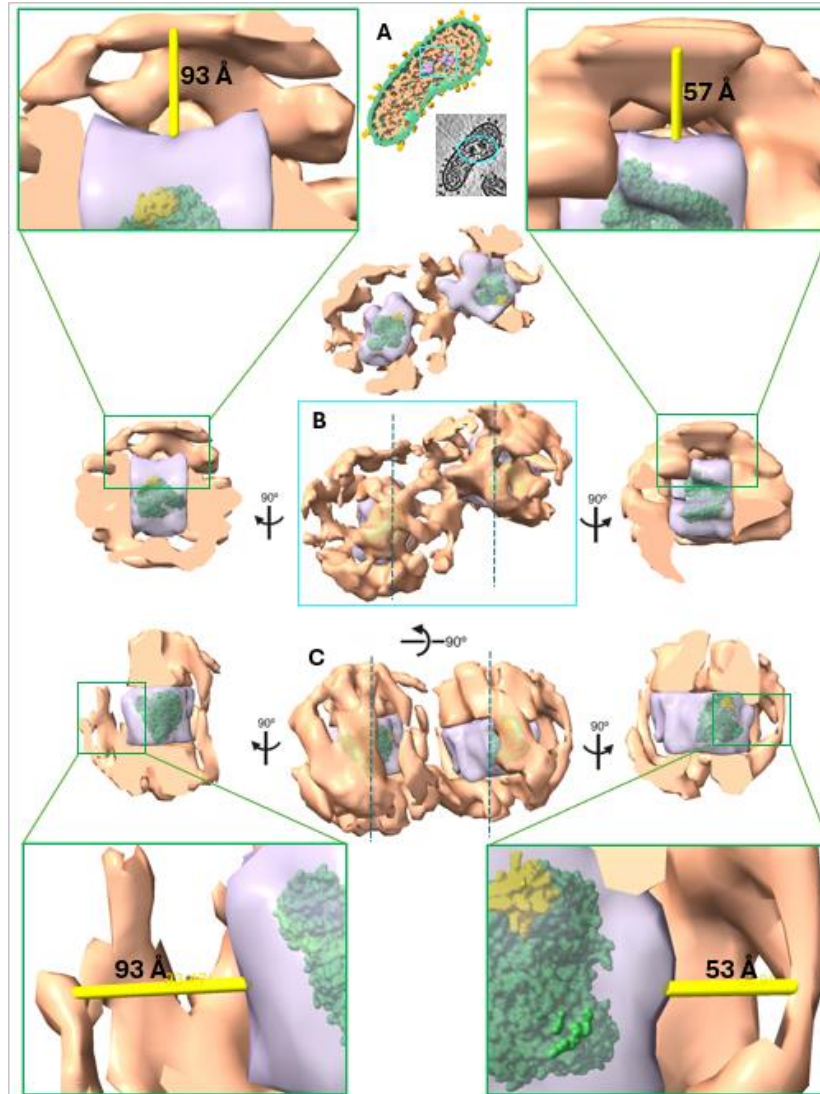
Figure 3.2: L-nucleocapsid interactions
(A) Cut-open view of the segmentation map from Figure 1D. The Inset shows the denoised virion reconstruction with the polymerase region highlighted in a cyan circle. (B) Zoom-in view of the cyan-boxed region in (A); (Top) (B) with the front nucleocapsid layer removed; (Left, Right) Left, or Right, L-Z region (fitted with green-yellow structure models) rotated 90° along the axis indicated on (B) with the front nucleocapsid layer removed; (Top Left, Right) Zoom-in view of the green-boxed region on the Left and right of (B) with the distance between the L-Z complex and the layer of nucleocapsid measured using the Chimera tape tool. (C) 90° rotated view of (B) as indicated by the arrow; (Left, Right) Left, or Right, L-Z region rotated 90° along the axis indicated on (C) with the front nucleocapsid layer removed; (Bottom Left, Right) Zoom-in view of the green-boxed region on the Left and Right of (C) with the distance between the L-Z complex and the layer of nucleocapsid measured using the Chimera tape tool.

While examining the anchorages earlier in Figure 1, we discovered a large number of distributed overlaps between the Z monomer and the GPC above. To examine this relationship further, we first identified all contact points between the viral envelope and the nucleocapsid with the surface markers and moved the markers up along their longitudinal axis to the outer surface layer (Figure 3.3A). Then, the contact point markers that coincided with GPCs were distinguished from those that did not (Figure 3.3A, B). The GPCs coinciding with the contact point markers were further distinguished from those that did not (Figure 3.3A, B). The total contact points were 180, of which 108 (60%) coincided with GPCs (Figure 2.3C). The total GPCs were 135, of which 108 (80%) coincided with the contact points (Figure 3.3D). These findings suggest that more than half of the contact points are found directly underneath GPCs and that GPCs are closely associated with the Z proteins, which serve as the anchorage of the nucleocapsid to the viral envelope.
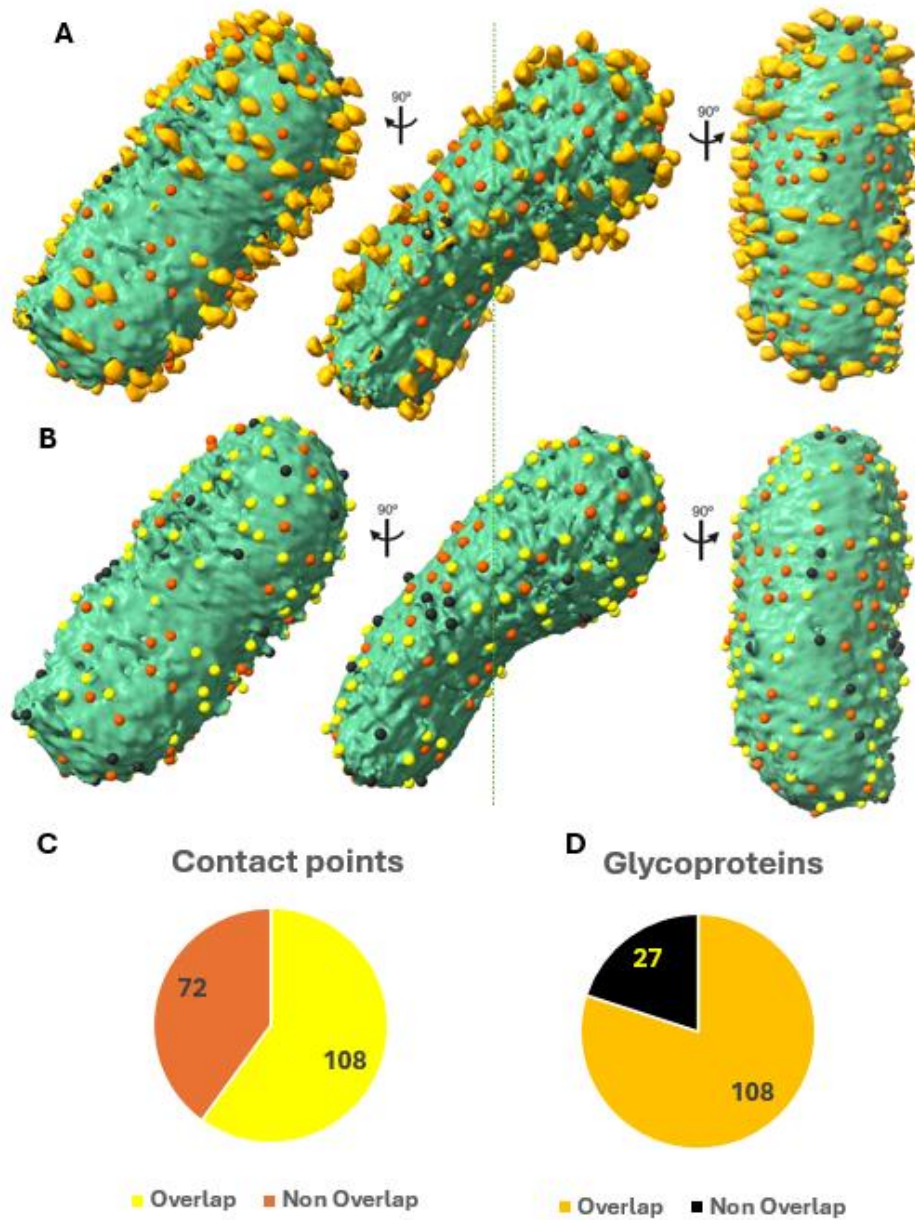
Figure 3.3: Global distribution of contact points and how it relates to glycoprotein
(A, B) (Center) Segmentation map of a virion from Figure 1D with the distribution of contact points between the nucleocapsid and the membrane shown as surface markers. Yellow markers indicate the contact points that coincide with the GPCs. Orange markers indicate the contact points that do not coincide with the GPCs. Black markers indicate GPCs that do not coincide with the contact points; (Left, Right) The view rotated 90° along the axis indicated on the center view. (B) Segmentation map of a virion from (A) with GPCs removed to show the yellow surface markers among others, which indicate the overlap between the contact points and GPC. (C) Pie Chart depicting the distribution of nucleocapsid-membrane contact points that do/do not coincide with the GPs. (D) Pie chart depicting the distribution of GPs that do/do not coincide with the nucleocapsid-membrane contact points. The color code follows the marker color code in (A, B).

3.4.2 *In situ* structure of LCMV GPC

The prefusion state GPC of LCMV is a trimer with each monomer subunit existing as a tripartite complex that consists of GP1, GP2, and SSP. Arenavirus SSP is an integral structural subunit of GPC that is necessary for the proper maturation and pH-dependent fusion process of GP. Retention of arenavirus SSP is proposed to be achieved via the interaction of SSP with the C-terminus region of GP2. Structurally, arenavirus SSP is modeled to comprise two antiparallel transmembrane helices connected by an eight-amino acid ectodomain loop. Despite these findings, the available GPC-related structures of LCMV are the prefusion state GPC trimer structure with each monomer consisting of GP1 and GP2 ectodomain and the postfusion state GPC trimer structure with each monomer consisting of GP2 ectodomain. Previous preparations were achievable by removing the hydrophobic domains, such as the GP2 TM domain and SSP domain (12, 25, 26).

CryoET, on the other hand, can be used to resolve structural features in the native environment, namely within the cell, by incorporating a subtomogram averaging (STA) approach. In STA, structures present in multiple copies, i.e., isotropic structures within the tomograms, can be extracted (subtomograms), aligned, and averaged, thus improving the signal-to-noise ratio and resolution. By taking advantage of STA, we resolved the *in situ* structure of membrane-bound LCMV GPC. GPCs of each virion were extracted as individually boxed particles from the earlier reconstructed tomograms. The box size was set to contain only one GPC and its underlying viral bilayer envelope. The GPC particles were subjected to RELION classification and refinements. RELION classification was set to generate 4 different classes, of which one class (5835 particles, 16.7 % of the extracted particles, data not shown) displayed the GPC ectodomain with the viral bilayer envelope underneath. The particles in this class were

selected for the subsequent rounds of refinement, generating a final C3 symmetrized reconstruction with a 12.7 Å resolution (Figure 3.4A). The ectodomain fitted well with the previously resolved model (PDB 8DMI), the prefusion state GPC (Figure 3.4A). The map was sliced halfway along the z-axis to examine the internal density more closely (Figure 3.4C). The thickness of the bilayer was 50 Å, matching the membrane thickness measured earlier on the segmentation map (Figure 3.1D) and indicating that the two layers represent the leaflets of the phospholipid bilayer (Figure 3.4C). Regarding the small density directly underneath the bottom leaflet, the C3 symmetrized Z monomers (PDB 5I72) fitted well into this region, corroborating our earlier findings of the distributed overlaps between GPC and Z (Figure 3.4B, C, G).

Next, the full-length LCMV GPC sequence was used to generate an AlphaFold2 (36) predicted model, which was utilized to build the TM domain and the C-terminal domain of GP2 (Figure 4B). For the most part, there was no significant deviation in our GP2 model from that of PDB 8DMI (Figure 3.4D). However, a deviation was observed in the region between F405 and I426 (Figure 4E). While 8DMI modeled this entire region as an ectodomain helix, AlphaFold2 modeled the region from F405 to E412 as a flexible string, lengthening the region by ~16 Å, and the remainder as a helix (Figure 3.4E). The lengthening of the region between F405 and E412 allowed us to bend this region to fit it into the narrow channel that connects the ectodomain and the top leaflet (Figure 3.4D, E). In this way, the string penetrates into the top leaflet, and the subsequent region from A413 to I426 becomes the first TM helix domain that spans the top leaflet (Figure 3.4C, E). The remaining helix domain extends perpendicular to the bilayer and crosses the central density between the bilayers (Figure 3.4C). A kink is introduced in the middle of the TM helix due to the cyclic P433 (Figure 3.4C, F). The second TM helix extends from L436 to L453, spanning the bottom leaflet of the bilayer. Following the TM domain, the

cytoplasmic tail begins at K457, followed by the zinc-binding domains 1 and 2 (ZBD1, 2) (Figure 3.4G).

Next, the Alphafold2 SSP model was utilized to complete the GPC modeling. The SSP model began with the helix 1 domain from G2 to I32, followed by an ectodomain from K33 to T40, and finished with the helix 2 domain from C41 to G54 (Figure 3.4F). The longer helix 1, instead of the shorter helix 2, is positioned next to the GP2 helix as it better fits our density map (Figure 3.4C). While the domain designation for the SSP ectodomain loop and helix 2 follows the previous finding, the helix 1 domain in our model deviates from the previous finding (6). Previously, the region from G2 to T6 and from M7 to D16 were designated as the myristoylation motif and cytoplasmic loop, respectively. However, our model suggests that the aforementioned regions are part of helix 1 to accommodate the ~50 Å thick bilayer (Figure 3.4C).

With the full GPC modeled, we next examined the model at the residue level to suggest potential interactions between SSP helix 1 and GP2 TM domain. Notwithstanding caution warranted for interpreting models based on cryoET structure of limited resolution, we propose residues D16, E17, N20, I24, and I27 of the SSP helix 1 domain and R422, I426, and Q429 of the GP2 TM domain as potential residues contributing to the two domain's hydrophobic interaction as they point toward each other in our model (Figure 3.4F). Residues I417, L421, and R428 of GP2 point toward the neighboring GP2, proposing their potential role in inter-subunit interaction. In summary, our model suggests that SSP retention is achieved by the hydrophobic interaction between its helix 1 domain and GP2's TM helix domain.
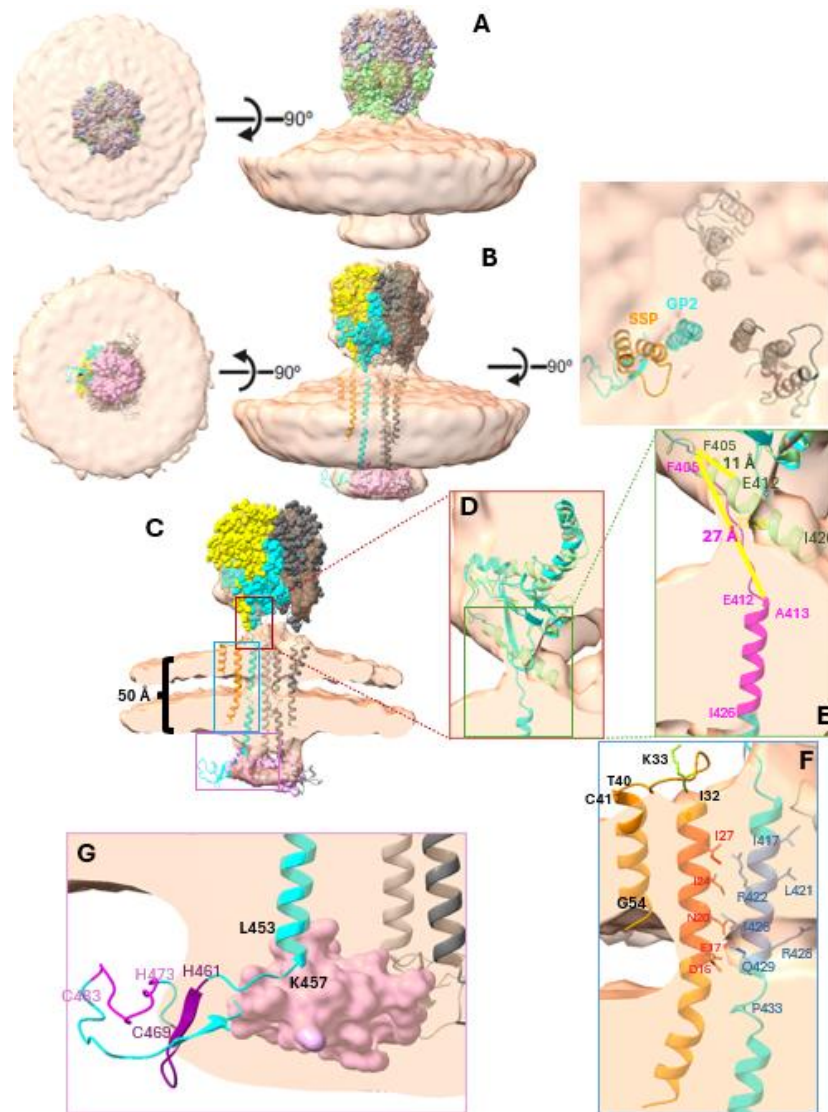
Figure 3 4 in situ structure of GPC using subtomogram averaging

(A) STA reconstruction (Peach) fitted with 8DMI in space-filling model representation. GP1 is in gray, and GP2 is in light green. (Left) Top view. (B) STA reconstruction fitted with our model and C3-symmetrized monomers of 5I72. GP1 is in yellow, GP2 is in cyan, SSP is in orange, and the Z monomer is in pink. Other GPCs are in dark gray and light gray. Ectodomain in space-filling model representation and TM domain and C-terminal domain in cartoon representation. (Left) Bottom view; (Right) Top view with the ectodomain removed. (C) Cut-open view of (B). The threshold has been adjusted from Figures A and B to better define the bilayer and the TM domain density. (D) Zoom-in view of the burgundy-boxed region in (C). It shows the superimposition of 8DMI GP2 (Light Green) and our modeled GP2 domain (Cyan). The models are in cartoon representation. (E) Zoom-in view of the green-boxed region in (D). The region between F405 and I426 of our model is highlighted in magenta. The distance between F405 and E412 of both models is measured using the Chimera tape tool. (F) Zoom-in view of the blue-boxed region in (C). The hydrophobic residues are labeled. SSP helix 1 residues that are interacting with GP2 are highlighted in darker orange, and the GP2 residues that are interacting with SSP helix 1 and neighboring GP2 are colored in darker blue. Ectodomain and helix 2 residues of SSP are highlighted in orange. (G) Zoom-in view of the purple-colored region in (C). ZBD1 and 2 of GP2 are highlighted in dark purple and light purple, respectively. One Z monomer is displayed for clarity.

66

3.4.3 Variation in the number of L polymerases packaged in LCMV

Arenaviruses have two segments of the RNA genome, L and S, each encoding two gene products. It has been assumed that each segment is encapsidated by polymeric NP subunits, and a nascent L polymerase is bound to 3' and 5'ends of each segment, reminiscent of a pearl necklace with a pendant. This general description suggests that each virion contains two nucleocapsids and two nascent L polymerases. To better understand the nucleocapsid-polymerase relationship, we examined each virion in our reconstructed tomograms (Figure 3.5A-D). Our 151 reconstructed tomograms contained 450 virions in total. The virions were grouped by the number of polymerases packaged, and the histogram was constructed (Figure 3.5F). Interestingly, our findings revealed that not all virions were packaged with two polymerases (Figure 3.5A-E). In fact, most virions were packaged with no polymerase, followed by those with 1 polymerase (Figure 3.5F). Virions with 2 polymerases belonged to the third most abundant group. Many virions were packaged with more than two Ls, with one virion with up to 20 Ls. All virions had mature GPCs studded on their membranes, regardless of the number of Ls packaged.

To find out the relationship between the virion size and the number of polymerases packaged in the virion, we measured the diameter of the virions belonging to each group assigned in the histogram. For the pleomorphic virions, the longest distance was measured, as indicated in the figures (Figure 3.5A, D). Virions with no or one L were generally smaller than those with more Ls, but had the greatest size variation, ranging from 36 to 140 nm and 50 to 200 nm, respectively (Figure 3.5G). Virions with 10 or more Ls were some of the largest virions, ranging from 330 nm to 376 nm, which corresponds to a virion with 20 Ls (Figure 3.5D, Right). Virions with 2 Ls had the least size variation, ranging from 157 to 170 nm in diameter (Figure 3.5C, 5G).
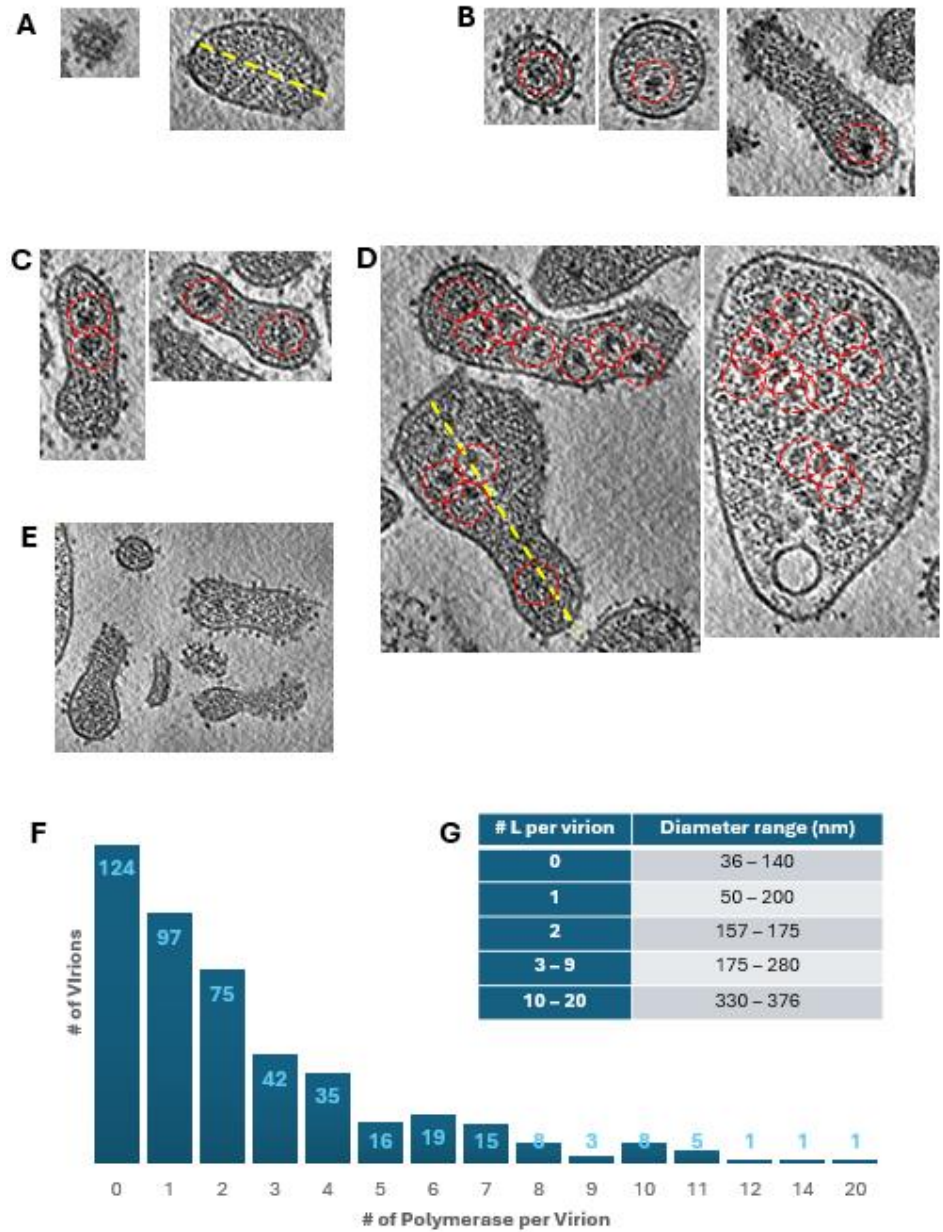
Figure 3.5 Denoised and missing-wedge corrected tomograms of LCMV virions with different number of L
(A-E) XY slice view of denoised tomograms, highlighting virions with various sizes and numbers of polymerases. Red dashed circles indicate the polymerase complex. (A) Virions with 0 polymerase. (B) Virions with 1 polymerase (C) Virions with 2 polymerases (D) Virions with more than 2 polymerases (E) Bird's eye view of polymerase-packaged virions coexisting with other virions packaged with no polymerase. (F) Histogram of the number of virions versus the number of polymerases packaged in a virion. For example, there are 124 virions packaged with no polymerase. (G) Table showing the size variation of virion in relation to the number of polymerase(s) packaged in the virion. The diameter was measured as indicated in the dashed yellow line in (A) and (D).

3.5 Discussion

In this study, we have used cryoET and STA to determine the architectural organization and *in situ* protein structures of LCMV. Our tomogram reconstruction indicates that the contact points between Z proteins and nucleocapsids are randomly distributed throughout the envelope, as Z proteins are associated with GPCs, which are also randomly distributed. This finding fits well with the previous findings on the subcellular colocalization and biochemical interaction between Z and GPC (37). Our model suggests that Z bridges between GPC and the nucleocapsid. More specifically, myristoylation of the N-terminal glycine of Z facilitates Z's interaction with the hydrophobic region of GP and the envelope membrane, and the nucleocapsid anchors to Z to associate with the envelope. In this way, the pattern of association among GP, envelope membrane, Z, and nucleocapsid governs the structural integrity and shape of LCMV. Indeed, the vesicular stomatitis virus (VSV) has a distinct bullet shape due to its unique association pattern of the viral proteins, albeit sharing the same fundamental three-layering system with LCMV. VSV, like LCMV, is a negative-strand RNA virus (NSV) with a double layer of matrix protein (M) residing between the glycoprotein (G)-studded membrane and the nucleocapsid that is composed of the nucleocapsid protein (N) and the encapsidated viral RNA. However, due to its unique organization of N-RNA, N-N, and N-M-N interactions and double layering of M that contributes to the helical packaging of the nucleocapsid, VSV assembles into its distinct bullet shape (38). With regard to the organization of glycoproteins, G trimers of VSV form hexagon-hexagon or hexagon-pentagon tiles of super-complexes on the viral envelope, whereas GPC trimers of LCMV have random distributions across the envelope. The hexagonal pattern formed by prefusion G trimers outside the VSV envelope matches the lattice formed by the outer M sites inside, similar to the way that the random distribution of LCMV GPC matches that of Z, which

bridges the GPC and the nucleocapsid. Taken together, the pleomorphic shape of LCMV is attributed to the random distribution of GPC, which results in the random distribution of internal contact points between Z and the nucleocapsid.

The global representation of a virion revealed the overall spatial arrangement of nucleocapsids inside the virion and its association with the L protein. While NPs at the periphery of nucleocapsids are spaced and stacked in an organized manner due to their close association with Z, the inner nucleocapsids are without any structural order, though more tightly bound. The segmented region at the core of the nucleocapsid is larger than the dimension of the L-Z complex, suggesting that the core region contains other viral proteins that bind tightly around the L-Z complex, resulting in a larger complex. It may be additional Z proteins or a tightly bound NP proteins reminiscent of heterochromatin that may help prevent viral genome replication. It may also be the case that the anchoring of Z on the peripheral nucleocapsid plays a supporting role in preventing immediate viral genome replication.

Our STA of GPC and modeling suggest that the retention of SSP is mainly achieved via hydrophobic interactions between the TM domain of GP2 and the helix 1 domain of SSP. The model supports the findings from the screening assay that identified residues 434 to 437 of GP2 TM domain and residues 25 to 27 of SSP helix 1 as the main residues targeted by the inhibitors, suggesting their critical roles in stabilizing the GP2-SSP interface (22, 39-42). In addition, our model places K33 of the ectodomain loop of SSP proximal to GP2, allowing for SSP interaction with GP2 and, most importantly, in the extracellular environment, allowing for pH sensing. This fits well with the finding of the role of highly conserved K33 on the pH-dependent fusion activity of GPC (22). At the same time, our model also reveals new findings. First, a 50 Å thick bilayer suggests longer TM domains, opening additional residues as potential candidates for

screening, namely from residues 16 to 24 and from residues 417 to 429 of the SSP and GP2, respectively. Second, our model suggests that the previously identified ectodomain helix of GP2 (F405 to I426) may instead be the first helix of the TM domain and that the first half of this helix has to be a flexible, lengthened string so that the rest of the structure can span the bilayer. Without the bending and the extra length, the ZBD1 and ZBD2 domains would become part of the TM domain instead of the cytosolic domains. The deviation may have arisen due to the absence of the C-terminal domain and the different sample preparation methods in the previous structure study (26). Third, our model suggests that the previously identified 20 AA-long region of GP2 between positions 428 and 447 (6) can only span one leaflet of the bilayer. This leads to the second finding in regard to the proposed interaction between GP2 ZBD2 domain and the C-terminal end of SSP. While it is speculated that the conserved C57 of the SSP C-terminus may function as the fourth ligand of the tetrahedral coordination of zinc in the cytoplasmic ZBD2 domain of GP2 (18), that model seems unlikely as the SSP helix 2 can only span the top leaflet of the ~50 Å thick bilayer. Taken together, our model suggests infeasibility of the interaction between GP2 ZBD2 domain and the C-terminal end of SSP. Our model does show an interaction between GP2 ZBD2 and a Z monomer, which may help explain our finding on the distributed overlaps between GPC and Z.

The large variation in the number of Ls packaged in the virion discovered by our image analyses adds new perspectives to the infection mechanism of LCMV. First, the abundance of virions with various numbers of Ls but all with intact GPCs suggests that the egress of LCMV can occur at any point in time as long as the proper maturation and intracellular trafficking of the GPC is completed. With no strict requirement for packaging of the nucleocapsid-L complex, LCMV may lack strategic temporal regulation of its egress, such as lytic to lysogenic transition

seen in Human Immunodeficiency virus type 1 (HIV-1) (43). Second, the great abundance of virions with no or one L suggests that LCMV mainly relies on these types of virions for their survival. To overcome their lack of strategic egress regulation, LCMV may use degenerate virions to increase the probability of reassortment, which helps to evade the host immunity. In addition, virions with more than two Ls may be used as special units for expedited viral transcription and rapid expression of viral proteins upon host cell entry, achieving more efficient virion production. Third, an unrestricted strategy might be a preferred way for LCMV to evolve due to its simple architectural organization— other NSVs require more sophisticated assembly checking due to their complex structure. As a comparison, the influenza genome contains eight segments that encode 10 essential viral proteins and several strain-dependent accessory proteins (44, 45). Insofar as the polymerases are concerned, three of these segments encode for the subunits of the heterotrimeric polymerase of the influenza virus (46, 47). This suggests that if the segments were randomly packaged upon egress, there are 1 in 56 chances for influenza virus and 1 in 2 chances for LCMV for the proper polymerase assembly. If we consider all other viral components in the probability calculation, random packaging becomes more feasible in LCMV with its simpler structure.   Due to the limited resolution of our reconstructions, it was difficult to discern further details of the nucleocapsid arrangement near the core, the tightly packed complex around L, and the TM domain of GP2. It will be of great value to investigate the structural arrangement of the nucleocapsid via single particle analysis of the purified nucleocapsid or nucleocapsid-L complex, and conducting STA of GPC with a greater number of particles to generate a higher resolution reconstruction. Nevertheless, our work highlights the use of cryoET and STA in elucidating the assembly mechanism and critical contacts between subunits of the protein oligomers and different viral proteins, namely within the TM region of GP and the

72

interfaces of GP-Z-NP and NP-NP, by whole virion reconstruction and solving the *in situ* structure of the full-length GPC of LCMV.

3.6 Methods

3.6.1 Preparation of LCMV

BHK-21 cells were maintained with EMEM (ATCC), supplemented with 10% Fetal Bovine Serum, at 37 °C, 5% CO2. At 90% confluency, LCMV inoculum diluted in EMEM with 1% Fetal Bovine Serum was absorbed to BHK-21 cells for 1 hour, and the infected cells were maintained EMEM, supplemented with 10% Fetal Bovine Serum for 48 hours at 37 °C, 5% CO2. The supernatant was collected and clarified by centrifugation. LCMV was pelleted by centrifugation at 30,000 g for 2.5 hours at 4 °C. The virus in the pellet was resuspended in phosphate-buffered saline (PBS) and further purified by sucrose gradient centrifugation (20%-50%) at 31,200 rpm in Beckman SW 41 Ti rotor for 1.5 hours at 4 °C. The virus band was collected and concentrated in PBS. Purified LCMV was deactivated by UV radiation for 30 min before use.

3.6.2 Grid preparation for cryoET

The purified, deactivated LCMV sample was diluted 100-fold in PBS, ultracentrifuged for 1 hour at 100,000 x g, and decanted to remove sucrose. The viral pellet was air-dried overnight in a cold room to be loosened and resuspended in 50 μL PBS. In preparation for the cryoEM, we added a 10 nm gold bead solution to the sample in a 1:20 volume ratio. The sample was applied to the carbon side of 200 mesh Cu Quantifoil 100 holey carbon films (R 3.5/1), which were glow discharged by Gatan Plasma System SOLARUS beforehand. The grids were loaded onto the manual plunger, blotted with the filter paper for 5 seconds, and plunged into liquid ethane/propane. Grids were stored in liquid nitrogen.

3.6.3 CryoET data collection and drift correction

The initial sample quality assessment was conducted using negative stain EM with 2% phosphate tungsten pH 7.5 on the FEI Tecnai TF20 at 200 Kilovolt (kV) equipped with a Gatan K2 direct electron detection camera. The final imaging was conducted with a Titan Krios electron microscope (Thermo Fisher Scientific) equipped with a Gatan imaging filter (GIF), and images were recorded on a post-GIF Gatan K3 Summit direct electron camera operated in super-resolution electron-counting mode. The magnification was 53,000x, with a pixel size of 1.72 Å/pixel at the specimen level. Data collection was facilitated by SerialEM (48). Tomographic tilt series were collected between -60° and 60° with a 3° angular increment in the dose-symmetry scheme (49). The total dose of each tilt series was 121 e$^-$/Å^2. Tilt series movies were recorded in dose-fraction mode and binned 2 times with the graphics processing unit-accelerated program MotionCor2 (50) to generate a micrograph in a tilt series.

3.6.4 Tomogram reconstruction and segmentation

152 tilt series were collected and reconstructed following the standard protocol provided by IMOD (51). The final tomograms were binned to a pixel size of 13.52 Å, deconvoluted, and corrected for the missing-wedge artifacts with IsoNet (52) using defocus values estimated by CTFFIND4 (53). The corrected tomogram was segmented using plane/surface markers and the split map function in ChimeraX (54).

3.6.5 Glycoprotein particle picking, subtomogram averaging, and TM domain modeling

The corrected tomograms were used to pick the GP particles. 1568 particles were manually picked from 4 tomograms and used to train a neural network model for automatic picking using the "AI

Autopicking" tool in TomoNet (55). In total, 34,775 particles were picked from 151 tomograms. The coordinates and orientations of these particles were formatted and imported to RELION 4 (56) for further refinement. In RELION, the imported subtomograms binned to a pixel size of 6.76 Å were subjected to two rounds of "3D auto-refine." The refined subtomograms were binned to a pixel size of 3.38 Å and underwent one round of "3D classification," from which a class containing 5835 particles was selected for the subsequent two rounds of "3D auto-refine." The finalized C3-symmetrized reconstruction reported a 12.5 Å resolution, which is based on the gold standard refinement procedures and the 0.143 Fourier shell correlation (FSC) criterion. TM domain modeling was conducted using AlphaFold2 (36) and Coot (57). The full sequence of LCMV GPC was used to generate a model which was modified manually to fit our density map while minimizing the steric clash.

3.7 Acknowledgments

3.8 Author contributions

Conceptualization: ZHZ and ML. Methodology: ZHZ, KZ, HW, ST, KVL and ML. Investigation and modeling: JK, KZ, HW, ZHZ and ML. Supervision: ZHZ and ML. Writing—original draft: JK and ZHZ. Writing—review & editing: JK, ZHZ, KZ, HW and ML

3.10 References

1.   C. G. Albarino *et al.*, High diversity and ancient common ancestry of lymphocytic choriomeningitis virus. *Emerg Infect Dis* **16**, 1093-1100 (2010).

2.   M. B. Oldstone, A suspenseful game of 'hide and seek' between virus and host. *Nat Immunol* **8**, 325-327 (2007).

3.   D. J. Jamieson, A. P. Kourtis, M. Bell, S. A. Rasmussen, Lymphocytic choriomeningitis virus: an emerging obstetric pathogen? *Am J Obstet Gynecol* **194**, 1532-1536 (2006).

4.   J. F. Meritet *et al.*, A case of congenital lymphocytic choriomeningitis virus (LCMV) infection revealed by hydrops fetalis. *Prenat Diagn* **29**, 626-627 (2009).

5.   S. J. Hallam, T. Koma, J. Maruyama, S. Paessler, Review of Mammarenavirus Biology and Replication. *Front Microbiol* **9**, 1751 (2018).

6.   H. N. Pennington, J. Lee, Lassa virus glycoprotein complex review: insights into its unique fusion machinery. *Biosci Rep* **42** (2022).

7.   W. Cao *et al.*, Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science* **282**, 2079-2081 (1998).

8.   S. C. Smelt *et al.*, Differences in affinity of binding of lymphocytic choriomeningitis virus strains to the cellular receptor alpha-dystroglycan correlate with viral tropism and disease kinetics. *J Virol* **75**, 448-457 (2001).

9.   B. M. Sullivan *et al.*, Point mutation in the glycoprotein of lymphocytic choriomeningitis virus is necessary for receptor binding, dendritic cell infection, and long-term persistence. *Proc Natl Acad Sci U S A* **108**, 2969-2974 (2011).

10. M. N. Teng, P. Borrow, M. B. Oldstone, J. C. de la Torre, A single amino acid change in the glycoprotein of lymphocytic choriomeningitis virus is associated with the ability to cause growth hormone deficiency syndrome. *J Virol* **70**, 8438-8443 (1996).

11. B. Eschli *et al.*, Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. *J Virol* **80**, 5897-5907 (2006).

12. S. Igonet *et al.*, X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc Natl Acad Sci U S A* **108**, 19967-19972 (2011).

13. J. York, S. S. Agnihothram, V. Romanowski, J. H. Nunberg, Genetic analysis of heptad-repeat regions in the G2 fusion subunit of the Junin arenavirus envelope glycoprotein. *Virology* **343**, 267-274 (2005).

14. F. M. Hughson, Structural characterization of viral fusion proteins. *Curr Biol* **5**, 265-274 (1995).

15. J. M. White, S. E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**, 189-219 (2008).

16. S. S. Agnihothram, J. York, M. Trahey, J. H. Nunberg, Bitopic membrane topology of the stable signal peptide in the tripartite Junin virus GP-C envelope glycoprotein complex. *J Virol* **81**, 4331-4337 (2007).

17. J. York, V. Romanowski, M. Lu, J. H. Nunberg, The signal peptide of the Junin arenavirus envelope glycoprotein is myristoylated and forms an essential subunit of the mature G1-G2 complex. *J Virol* **78**, 10783-10792 (2004).

18.   K. Briknarova, C. J. Thomas, J. York, J. H. Nunberg, Structure of a zinc-binding domain in the Junin virus envelope glycoprotein. *J Biol Chem* **286**, 1528-1536 (2011).

19.   D. J. Burri *et al.*, The role of proteolytic processing and the stable signal peptide in expression of the Old World arenavirus envelope glycoprotein ectodomain. *Virology* **436**, 127-133 (2013).

20.   M. Froeschke, M. Basler, M. Groettrup, B. Dobberstein, Long-lived signal peptide of lymphocytic choriomeningitis virus glycoprotein pGP-C. *J Biol Chem* **278**, 41914-41920 (2003).

21.   S. Shankar *et al.*, Small-Molecule Fusion Inhibitors Bind the pH-Sensing Stable Signal Peptide-GP2 Subunit Interface of the Lassa Virus Envelope Glycoprotein. *J Virol* **90**, 6799-6807 (2016).

22.   J. York, J. H. Nunberg, Intersubunit interactions modulate pH-induced activation of membrane fusion by the Junin virus envelope glycoprotein GPC. *J Virol* **83**, 4121-4126 (2009).

23.   E. L. Messina, J. York, J. H. Nunberg, Dissection of the role of the stable signal peptide of the arenavirus envelope glycoprotein in membrane fusion. *J Virol* **86**, 6138-6145 (2012).

24.   A. A. Saunders *et al.*, Mapping the landscape of the lymphocytic choriomeningitis virus stable signal peptide reveals novel functional domains. *J Virol* **81**, 5649-5657 (2007).

25.   K. M. Hastie *et al.*, Crystal structure of the prefusion surface glycoprotein of the prototypic arenavirus LCMV. *Nat Struct Mol Biol* **23**, 513-521 (2016).

26.   A. Moon-Walker *et al.*, Structural basis for antibody-mediated neutralization of lymphocytic choriomeningitis virus. *Cell Chem Biol* **30**, 403-411 e404 (2023).

27. L. Brunotte *et al.*, Structure of the Lassa virus nucleoprotein revealed by X-ray crystallography, small-angle X-ray scattering, and electron microscopy. *J Biol Chem* **286**, 38748-38756 (2011).

28. K. M. Hastie *et al.*, Convergent Structures Illuminate Features for Germline Antibody Binding and Pan-Lassa Virus Neutralization. *Cell* **178**, 1004-1015 e1014 (2019).

29. K. M. Hastie *et al.*, Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. *Proc Natl Acad Sci U S A* **108**, 19365-19370 (2011).

30. K. M. Hastie *et al.*, Crystal Structure of the Oligomeric Form of Lassa Virus Matrix Protein Z. *J Virol* **90**, 4556-4562 (2016).

31. S. Li *et al.*, Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike. *PLoS Pathog* **12**, e1005418 (2016).

32. B. W. Neuman *et al.*, Complementarity in the supramolecular design of arenaviruses and retroviruses revealed by electron cryomicroscopy and image analysis. *J Virol* **79**, 3822-3830 (2005).

33. R. Peng *et al.*, Structural insight into arenavirus replication machinery. *Nature* **579**, 615-619 (2020).

34. X. Qi *et al.*, Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature* **468**, 779-783 (2010).

35. L. Volpon, M. J. Osborne, A. A. Capul, J. C. de la Torre, K. L. Borden, Structural characterization of the Z RING-eIF4E complex reveals a distinct mode of control for eIF4E. *Proc Natl Acad Sci U S A* **107**, 5441-5446 (2010).

36.   J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

37.   A. A. Capul *et al.*, Arenavirus Z-glycoprotein association requires Z myristoylation but not functional RING or late domains. *J Virol* **81**, 9451-9460 (2007).

38.   K. Zhou *et al.*, Atomic model of vesicular stomatitis virus and mechanism of assembly. *Nat Commun* **13**, 5980 (2022).

39.   Y. Liu *et al.*, Screening of Botanical Drugs against Lassa Virus Entry. *J Virol* **95** (2021).

40.   K. Tang, X. Zhang, Y. Guo, Identification of the dietary supplement capsaicin as an inhibitor of Lassa virus entry. *Acta Pharm Sin B* **10**, 789-798 (2020).

41.   J. York, J. H. Nunberg, Role of the stable signal peptide of Junin arenavirus envelope glycoprotein in pH-dependent membrane fusion. *J Virol* **80**, 7775-7780 (2006).

42.   X. Zhang, K. Tang, Y. Guo, The antifungal isavuconazole inhibits the entry of lassa virus by targeting the stable signal peptide-GP2 subunit interface of lassa virus glycoprotein. *Antiviral Res* **174**, 104701 (2020).

43.   J. Hokello, A. L. Sharma, M. Dimri, M. Tyagi, Insights into the HIV Latency and the Role of Cytokines. *Pathogens* **8** (2019).

44.   W. Chen *et al.*, A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med* **7**, 1306-1312 (2001).

45.   D. Dou, R. Revol, H. Ostbye, H. Wang, R. Daniels, Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front Immunol* **9**, 1581 (2018).

46.   M. Clifford, J. Twigg, C. Upton, Evidence for a novel gene associated with human influenza A viruses. *Virol J* **6**, 198 (2009).

47. S. Dey, A. Mondal, Unveiling the role of host kinases at different steps of influenza A virus life cycle. *J Virol* **98**, e0119223 (2024).

48. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).

49. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J Struct Biol* **197**, 191-198 (2017).

50. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

51. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76 (1996).

52. Y. T. Liu *et al.*, Isotropic reconstruction for electron tomography with deep learning. *Nat Commun* **13**, 6482 (2022).

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

54. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).

55. H. Wang, S. Liao, X. Yu, J. Zhang, Z. H. Zhou, TomoNet: A streamlined cryoET software pipeline with automatic particle picking on flexible lattices. *bioRxiv* 10.1101/2024.02.17.580557 (2024).

56. J. Zivanov *et al.*, A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11** (2022).

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

Chapter 4: Conclusion

It has been well-recognized that many biological particles tend to have preferential orientation, aggregate, or mysteriously disappear on cryoEM grids, but the reasons for such misbehavior, also known as the AWI adsorption phenomenon, are not well understood. To explain this phenomenon, we developed a theoretical formulation that explains these behaviors in the context of surface energy. According to this framework, particles migrate to AWI to decrease the overall area of the liquid exposed to the air, thus minimizing the overall potential energy of the system. This framework rationalized the use of surfactants to alleviate the AWI adsorption phenomenon, as they reduce the surface energy of the buffer layer, reducing the overall potential energy and thus minimizing the particles' tendency to migrate toward the AWI. We used cryoET to demonstrate the effectiveness of various surfactants by visualizing the improved groEL particle distribution across the entire depth of the cryoEM grid from one AWI to the other. We showed that the surfactants were most effective at their CMCs and that the fluorinated surfactants were the most effective. We also conducted SPA cryoEM on groEL to show that the surfactants do not adversely affect the protein structures at a near-atomic level. We further showed the effectiveness of fluorinated surfactant on a more challenging membrane protein, CLC-1.

Next, we shifted our gear to conducting a structural study on LCMV to showcase the benefits of using cryoET and sub-tomogram averaging on membrane proteins. At the time of this writing, the structures of individual LCMV proteins or their fragments have been reported but the architectural organization and the full-length *in situ* structure of these proteins remained to be solved. The main hurdles for resolving the full GPC were its TM domain and its associating TM SSP domain. Using cryoET and STA, we unveiled the architectural organization of viral proteins of LCMV virion and resolved the *in situ* intact GPC along with the SSP domain. The cryoET

revealed the global distribution of GPC, matrix proteins Z, and the contact points between the viral envelope and the underlying nucleocapsid to better understand the mechanism behind the shape and the structural integrity of LCMV. Also, cryoET revealed a great variation in the number of RdRps packaged within each virion, adding new perspectives to the infection mechanism. Finally, STA unveiled the GP2-SSP interface, highlighting potential key residues that can be considered for future structure-guided vaccine design.

Recent advances in cryoEM have made it possible to understand macromolecules structurally in isolated forms. However, the isolation process can be challenging and can disrupt their native structure. This dissertation highlighted the benefit of cryoET, which addresses these challenges. However, while cryoET allows direct freezing of thin bacteria and small cells that are thinner than 500 nm, it is not applicable to most biological samples, which are thicker than 500 nm. To overcome this limitation, cryogenic focused ion beam (cryo-FIB) milling has been developed to generate thin biological samples while retaining their vitreous state, which can then be imaged using cryoET. FIB fabrication removes all layers above and beyond a section of the cell, leaving a thin (50-300 nm thick) lamella of a cell. Cellular tomography using FIB milled lamella has been implemented to unveil the molecular environment of the cell and just recently, revealed viral assembly intermediates within the host cytoplasm. Major challenges are still present, but the continued development that addresses these challenges can evolve cellular tomography using FIB-milled lamella into a powerful structural imaging tool for understanding biological processes.

This dissertation also highlighted the effectiveness of surfactants in alleviating the AWI adsorption phenomenon. However, while the particle distribution improved, the preferred orientation problem still persisted during our investigation. Both SPA and STA are prone to this problem as the preferential orientation bias can debilitate the averaging step that relies on the

isotropic features of the extracted particles. Currently, there are biochemical and physical approaches to address this problem, such as chemically modifying the protein specimen, coating cryoEM grids with support films, using time-resolved vitrification devices, collecting data from thicker ice regions, and tilting the grids during image acquisition. However, these approaches can be labor-intensive and expensive and may even lead to unexpected outcomes. Just recently, a self-supervised deep-learning method called single-particle IsoNet (spIsoNet) has been developed, which addresses both anisotropic reconstruction and particle misalignment caused by the preferred orientation problem. Innovation in computational approach will enable 3D reconstruction from fewer particles and orientation views, improving the accuracy and efficiency of cryoEM-based data processing.

FIN