Title

Structural Characterization of Microbial Pathogenesis: From Viruses to Parasites

Author

Stevens, Alexander

Publication Date

2024

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Los Angeles

Structural Characterization of Microbial Pathogenesis:

From Viruses to Parasites

A dissertation submitted in partial satisfaction of the requirement for the degree Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

by

Alexander Stevens

2024

ABSTRACT OF THE DISSERTATION

Structural Characterization of Microbial Pathogenesis:
From Viruses to Parasites

by

Alexander Stevens

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2024

Professor Hong Zhou, Co-Chair

Professor Jose Alfonso Rodriguez, Co-Chair

Diseases caused by pathogenic microbes impose a substantial economic and public health burden on the world. Advancements in cryogenic electron microscopy (cryoEM) have revolutionized our ability to describe the atomic structures driving microbial pathogenesis, while eliminating artefacts introduced by classical structure methods. CryoEM improves both the biological context of these structures and broadens the scope of structural investigations to include large and dynamic complexes, where subtle differences can yield insights critical to combating these microbes.

I began my thesis research by leveraging recent advancements in cryoEM image processing to address gaps in our understanding of the assembly and replication of

double stranded RNA (dsRNA) viruses. We began with the study of a dsRNA virus with minimal genomic complexity, which lacks the capacity for intercellular infection. The resulting 3.6 Å structure of the viral capsid informs us on the essential features required for dsRNA virus replication. We next turned to a more complex dsRNA virus, Aquareovirus (ARV), which poses a serious threat to aquaculture. By resolving the asymmetric structure of ARV capsid core—an intermediate state of the replication process—to 3.3 Å, we could use the subtle differences between the complete and core particles to suggest a mechanism that explains the previously observed phenomenon of transcriptional inhibition in the complete particle. Our findings in both viruses deepen our understanding of viral replication in dsRNA viruses.

Following this, we used cryoEM structures of the human cytomegalovirus (HCMV) capsid to guide the rational design of mutants targeting the protein-protein interface between the HCMV specific tegument protein, pp150, and the capsid. This structure guided mutagenesis allowed us to identify the potential mechanism of pp150 nuclear import and discover an attenuating mutation that slowed viral replication. Interestingly, this reduced replication rate did not compromise virion formation, suggesting such a mutation in clinical strains, may slow replication sufficiently for use as a vaccine strain, without affecting the virus's antigenic profile.

Lastly, we applied cryoEM to cytoskeletal elements from the common genitourinary parasite *T. vaginalis* (*Tv*). As *Tv* pathogenicity relies upon the multifunctional, motile flagella, we aim to characterize the flagellar cytoskeleton, namely the doublet microtubules of the axoneme. In doing so, we identified the minimally complex arrangement of microtubule inner proteins (MIPs) that facilitate motility through the

viscous environment of the host genitourinary tract. Furthermore, we identified a Novel MIP *Tv*-specific protein as a potential drug target.

This work demonstrates the versatility of cryoEM to resolve new high-resolution structures, identify novel drug targets, and guide the rational design of mutants to probe protein-protein interactions and inform on the mechanisms of pathogenesis.

The dissertation of Alexander Stevens is approved.

William M. Gelbart

Steven G. Clarke

Jose Alfonso Rodriguez, Co-Chair

Hong Zhou, Co-Chair

University of California, Los Angeles

2024

DEDICATION

For my Mom and Dad

Table of Contents

List of Figures

List of Tables

Acknowledgements

I extend my heartfelt gratitude to my advisor, Professor Hong Zhou, whose unwavering belief in me helped foster a belief in myself. Dr. Zhou leads by example, and his careful guidance and steadfast support have been instrumental in my pursuit of academic and professional goals. I am also deeply thankful to my doctoral committee members, Professors Jose Rodriguez, William Gelbart, and Steven Clarke, who have provided invaluable direction throughout my studies at UCLA and encouraged me to explore diverse projects.

I stand on the shoulders of those lab members who preceded me, including Dr. Yanxiang Cui, Dr. Sakar Shivakoti, and Dr. Ana Lucia Alvarez-Cabrera, whose work introduced me to cryoEM. I am also grateful for the fruitful scientific discussions and mentorship from my lab mates, particularly James Zhen and Johnathan Jih, for their broad scientific expertise and advice. Special thanks to Dr. Hui Wang for his expertise in electron tomography, Dr. Yuntao Liu for training me in single particle analysis, and Dr. Yao He, Dr. Xian Xia, Dr. Zhu Si, and Dr. Shiheng Liu for their guidance during data collection and processing. I also appreciate Johnathan's assistance in managing the data center and integrating different technologies which made me much more productive.

My gratitude extends to the facilities staff at CNSI, including Ivo Atanasov, Dr. David Strugatsky, Dr. Wong Hoi Hui, Dr. Sergey Ryazantsev, and Dr. Matthew Mecklenburg, who trained me in both molecular biology and electron microscopy, and guided me through numerous challenges. Their round-the-clock support was invaluable in the collection of much of the data described here.

During my studies, I had the privilege of collaborating with several labs, including those of Professor Patricia Johnson (UCLA) and Professor Qiyi Tang (Howard University). I am particularly grateful to Dr. Edward Wang from the Johnson lab for completing the T. vaginalis work initiated by Dr. Katherine Muratore, and to Ruth Cruz-Cosme and Najealicka Armstrong from the Tang lab for providing the biochemical and molecular biology expertise essential to our HCMV mutagenesis work.

I am fortunate to have received significant support from the Biochemistry and MIMG departments, as well as awards from programs funded by donations from Sidney Charles Rittenberg and Audree V. Fowler. I would like to thank Professors Patricia Johnson and Peter Bradley for overseeing the Interdisciplinary Microbial Pathogenesis Training Grant, which I was honored to be a part of for two years.

I would not have accomplished much of what is described without the outstanding work of my undergraduate students, Saarang Kashyap and Ethan Crofut. Training them has been a pleasure, and their tireless work ethic motivates me to be more productive.

Outside of the lab, I have been fortunate to receive the love and support of many family and friends. I am especially grateful to my girlfriend, Sara Brancati, whose unwavering support and understanding have been a constant source of strength throughout this journey. I also extend a heartfelt thanks to Sara's family, particularly her parents Russel and Elisa, who have welcomed me into their family and given me endless encouragement.

I am also deeply grateful to Dr. J.P. Nelson, who has always offered a compassionate ear and with whom I had the privilege of publishing an off-topic article in 2023. I'd also like to thank Simon Paige for his constant friendship and for encouraging me to pursue excellence in endeavors beyond academia. Few are willing to wake up at 5 AM for more work, but we keep at it.

I am also profoundly thankful to the Gilbreath family, who graciously took me in during high school and provided me with a home full of warmth and support during a trying year of my life. Their kindness and belief in me have left an indelible mark on my life that continues to inspire me today.

Finally, I express my deepest gratitude to my family—my brother Calvin, who taught me patience; my father Richard, who inspired me to strive for greatness; and my mother Brenda, who taught me how to find joy in life. Their constant support has instilled in me the belief that with hard work and determination, one can fully pursue their dreams.

Chapter 2 is adapted from **Stevens, A.**, Muratore, K., Cui, Y., Johnson, P. J. & Zhou, Z. H. Atomic Structure of the Trichomonas vaginalis Double-Stranded RNA Virus 2. *MBio* 12, 1–17 (2021). ASM authors also retain the right to reuse the full article in their dissertation or thesis. This project is supported partly by grants from the US NIH (DE028583/DE025567/AI094386 to Z.H.Z. and R01AI103182/R33AI119721 to P.J.J). K.A.M. received support from NIH Ruth L. Kirschstein National Research Service Award AI007323. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by UCLA and grants from the NIH (1S10OD018111 and 1U24GM116792) and the National Science Foundation (DBI-1338135 and DMR-1548924). Z.H.Z. and P.J. conceived the project; K.M. prepared samples; Y.C. recorded

cryoEM images and reconstructed the 3D maps; A.S. analyzed the map and built atomic models; Z.H.Z. and A.S. interpreted results and wrote the paper; all authors edited and approved the paper.

Chapter 3 is adapted from **Stevens, A**., Cui, Y., Shivakoti, S. & Zhou, Z. H. Asymmetric reconstruction of the aquareovirus core at near-atomic resolution and mechanism of transcription initiation. *Protein Cell* 14, 544–548 (2023). All Oxford journal authors retain the right "… to include the article in full or in part in a thesis or dissertation, provided that this is not published commercially". This work was supported in part by grants from the National Institutes of Health (AI094386 to Z.H.Z.). A.S. received support from NIH Ruth L. Kirschtein National Research Service Award AI007323. We acknowledge the use of resource at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057, 1S10OD018111 and U24GM116792) and NSF (DBI-1338135 and DMR-1548924) and at the UCLA AIDS Institute supported in part by UCLA-CDU CFAR (NIH AI152501), the James B. Pendleton Charitable Trust and the McCarthy Family Foundation. Z.H.Z. designed and supervised the project; S.S. prepared sample and made cryoEM grids; Y.C. performed cryoEM imaging and 3D reconstruction; A.S. built the atomic models, interpreted the structures, made the figures, and wrote the paper with input from Z.H.Z.; All authors reviewed and approved the paper.

Chapter 4 is adapted from **Stevens, A.,** Cruz-Cosme, R., Armstrong, N., Tang, Q. & Zhou, Z. H. Structure-guided mutagenesis targeting interactions between pp150 tegument protein and small capsid protein identify five lethal and two live-attenuated HCMV mutants. *Virology* 596, 110115 (2024). From Elsevier journals "Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.” We thank Sergey Ryazantsev & Chunni Zhu for the cell embedding and thin-sectioning service. This work was supported by grants from the US National Institutes of Health (R01DE028583 to ZHZ). A.S. Received support from NIH Ruth L. Kirschstein National Research Service Award AI007323. We acknowledge the use of resources in the Electron Imaging Center for NanoSystems (EICN) supported by UCLA, National Science Foundation (DMR-1548924 and DBI-1338135) and National Institutes of Health (S10RR23057 and U24GM116792). We acknowledge support from the UCLA AIDS Institute, the UCLA Brain Research Institute Electron Microscopic Core, the James B. Pendleton Charitable Trust, and the McCarthy Family Foundation. Alex Stevens: Overall experimental design and execution of the electron microscopy-related components, data analysis and result interpretation, manuscript drafting and figure preparation, paper editing. Ruth Cruz-Cosme: BAC-related experiments, co-IP, and western blot assays. Najealicka Armstrong: molecular cloning to generate plasmids and immunofluorescence assays. Qiyi Tang: Design and supervision of BAC-related experiments, co-IP experiments, fluorescence microscopy, and manuscript editing. Z. Hong Zhou: Overall project design and supervision of execution, research fund acquisition and management, manuscript draft and editing.

Chapter 5 is adapted from a manuscript submitted to Nature Communications on August 24, 2024: **Stevens, A**. *et al.* Structures of Native Doublet Microtubules from Trichomonas vaginalis Reveal Parasite-Specific Proteins as Potential Drug Targets. *bioRxiv* (2024) doi:doi.org/10.1101/2024.06.11.598142. Authors retain copyright of their manuscripts submitted to *bioRxiv*. We thank Yanxiang Cui for his preliminary work on

this project, along with Xian Xia and Kent Hill for their expert advice in the development of this work. This work was supported in part by grants from the National Institutes of Health (R01GM071940 to Z.H.Z. and R01AI103182/R33AI119721 to P.J.J.). K.A.M. and A.S. received support from NIH Ruth L. Kirschtein National Research Service Award AI007323. We acknowledge the use of resources at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057, 1S10OD018111) and NSF (DBI-1338135 and DMR-1548924). We acknowledge support from the UCLA AIDS Institute, the James B. Pendleton Charitable Trust, and the McCarthy Family Foundation. Z.H.Z and P.J. designed and supervised the project. K.A.M., S.E.W., and A.S. prepared samples. S.E.W. conducted mass spectrometry work. A.S. and S.K. performed cryoEM imaging and prepared 3D reconstructions. Under the guidance of Z.H.Z., A.S., S.K., and E.H.C. built the atomic models, interpreted the structures, and made the figures and wrote the paper; all authors reviewed and approved the paper.

Vita

Education

University of California, Los Angeles

Los Angeles, CA

PhD, Biochemistry, Molecular, and Structural Biology Expected August 2024

- GPA: 3.83

Arizona State University 2019

Tempe, AZ

B.S., Biochemistry (Medicinal Chemistry)

- GPA: 3.79

Publications: Original Research

1. **Stevens, A**., Cruz-Cosme, R., Armstrong, N., Tang, Q. & Zhou, Z. H. Structure-guided mutagenesis targeting interactions between pp150 tegument protein and small capsid protein identify five lethal and two live-attenuated HCMV mutants. *Virology* **596**, 110115 (2024).
2. **Stevens, A**.*, Cui, Y.*, Shivakoti, S*. & Zhou, Z. H. Asymmetric reconstruction of the aquareovirus core at near-atomic resolution and mechanism of transcription initiation. *Protein Cell* **14**, 544–548 (2023).
3. **Stevens, A**.*, Muratore, K.*, Cui, Y.*, Johnson, P. J. & Zhou, Z. H. Atomic Structure of the Trichomonas vaginalis Double-Stranded RNA Virus 2. *MBio* **12**, 1–17 (2021).
4. **Stevens, A**.*, Kashyap, S.*, Crofut, E. H., Wang, S. E. & Katherine, A. Structures of Native Doublet Microtubules from Trichomonas vaginalis Reveal Parasite-Specific

Proteins as Potential Drug Targets. *bioRxiv* (2024). Submitted to Nature Communications.

(* indicates co-first authorship)

Honors and Awards

Sidney Charles Rittenberg Award, UCLA	2024
Whitcome Pre-doctoral Fellowship, UCLA	2023-2024
Dissertation Year Fellowship, UCLA	2023-2024
Audree V. Fowler fellow in Protein Science, UCLA	2022-2023
NIH T32 Interdisciplinary Training in Microbial Pathogenesis	2021-2023
George M. Bateman Memorial Scholarship, ASU	2016

1. Chapter 1: Introduction

1.1. Understanding microbial pathogenesis through structure

James Watson and Francis Crick, best known for interpreting Rosiland Franklin's X-ray diffraction data to propose the DNA double helix [1], were also the first to propose that viral capsids—proteinaceous shells containing nucleic acids—are composed of many identical subunits arranged into regular, often helical or icosahedral, shapes [2] (Fig. 1A). Later structural studies using X-ray crystallography, including those of tobacco mosaic virus (TMV) and tomato bushy stunt virus (TBSV) confirmed their hypothesis [3,4]. Further crystallographic investigations, such as Stephen Harrison's work on southern bean mosaic virus (SBMV), revealed a shared 'jelly-roll' fold between TBSV and TBMV that suggested a common ancestor amongst these RNA viruses [5,6]. Michael Rossmann and colleagues later identified this jelly-roll fold in the structure of human rhinovirus 14 (HRV14) [7], and linked this conserved structure to host-receptor recognition and immune evasion [8]. This laid the groundwork for the development of broadly effective anti-HRV drugs that targeted these structural motifs [9].

1.2. X-ray crystallography gives way to cryoEM

Though X-ray crystallography was integral to early high-resolution structural studies of pathogens, progress was often slow, hampered by the challenges of crystallizing large, often dynamic samples [10]. By contrast, transmission electron microscopy (TEM), while theoretically more versatile, had significant limitations for use in high-resolution structure determination but could determine global structures of large symmetric complexes to about 1 nanometer resolution [11,12]. Indeed, TEM had demonstrated utility for resolving symmetric virus structures of human papilloma virus and TBSV which were resolved by R.A. Crowther and colleagues at the MRC [13]. Unfortunately, drying artefacts, data collection and processing remained major roadblocks to widespread adoption of TEM for the study of pathogen structures.

This changed when Jacques Dubochet demonstrated the cryogenic preservation of specimens in thin films of vitreous ice, which protected them within the vacuum of the TEM column and paved the way for cryogenic electron microscopy (cryoEM) [14-16]. Although data collection and processing remained challenging, structural biologists recognized the contextual significance of specimens preserved in near-native, frozen-hydrated conditions. Dubochet quickly applied his innovative cryoEM technique to viruses [17], work that was soon replicated by others including Venkatar Prasad from Wah Chiu's Lab [18], and Peter Metcalf at EMBL [19], though resolutions remained limited to the molecular scale.

Less than a decade later, Bettina Bottcher from Crowther's Lab and James Conway from Alisdair Stevens' lab published cryoEM structures of Hepatitis B virus showing secondary structure details [20,21]. By the late 1990s, well established crystallographers such as Michael Rossmann and Richard Kuhn began to incorporate cryoEM into their work, developing a hybrid workflow to insert high-resolution X-ray structures of individual viral capsid subunits into the low resolution reconstructions of the complete capsids generated by cryoEM [22-24]. Harrison and others quickly adopted this approach [25], improving the biological relevance of their structures and abrogating the need for challenging crystallization trials, though cryoEM the resolution remained limited to secondary structure determination.

In 2008, Hong Zhou reported the first near-atomic resolution cryoEM structures from an icosahedral virus capsid that enabled de-novo atomic model building [26], and repeated this feat in a filamentous virus shortly thereafter [27], kicking off what has come to be known as the resolution revolution in cryoEM. Recent advancements in single-particle cryoEM helped it largely supplant X-ray crystallography and become the premier technique for the structural characterization of large biological complexes [28,29]. Key innovations, including the development of direct electron detectors and sophisticated algorithms for image processing helped cryoEM structures reach near-atomic and eventually atomic resolution [30-33].

1.3. State-of-the-art cryoEM

The modern cryoEM workflow (Fig. 1) begins with isolation of the sample, which is then frozen in a thin layer of vitreous ice on a transparent cryoEM grid [14,16,17]. Within the vacuum of the TEM column, electrons are accelerated to relativistic speeds, and focused via electromagnetic lenses onto the sample [34,35]. At the specimen level, many electrons collide with the sample, scattering inelastically and imparting energy that rapidly degrades the molecular integrity of the fragile biological samples [36]. However, some electrons scatter elastically passing through the sample harmlessly, while accumulating spatial information [36]. While the inelastically charged electrons are filtered out based on energy loss, the elastically scattered electrons are focused to form an image containing the spatial information of the particles under examination.

Two-dimensional (2D) images are recorded as movie frames across the cryoEM grid, capturing many copies of the identical specimens in various orientations. As electron scattering induces sample motion, software is used to correct for this motion blur, aligning and averaging the frames to create a single micrograph [37]. The orientation of each particle in three-dimensional (3D) space is determined using complex algorithms, and identical orientations are averaged to improve the signal-to-noise ratio [38]. Symmetrical specimens, such as helical or icosahedral virus capsids, benefit from reduced angular search space needed to determine their orientation [13,39]. Once their orientations are determined, the 2D images are merged to generate a complete 3D density map [40], which corresponds to the structure of the 3D specimen and, if the resolution is sufficient, enables us to determine how the amino acid residues are arranged throughout the protein. Following the days of crystallography, cryoEM has fostered an explosion in structural studies characterizing pathogens and elucidating the mechanisms underpinning their pathogenesis.

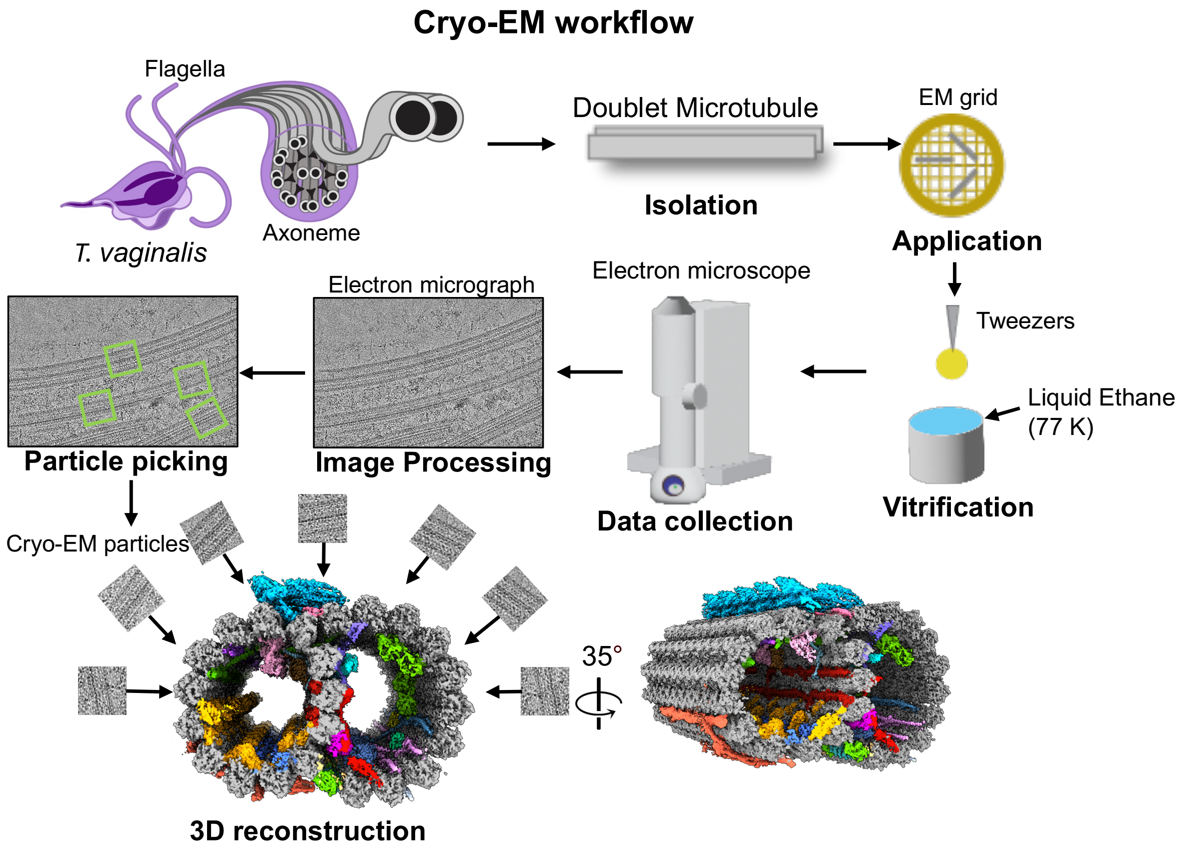


Figure 1- 1. typical CryoEM workflow as illustrated using T. vaginalis doublet microtubules.

1.4. Viral genome structures and cryoEM

In addition to less rigorous sample preparation requirements, cryoEM offers distinct advantages over X-ray crystallography inherent to the single particle data processing approach. In X-ray crystallography, specimens are packed into a regular crystalline lattice, which relies solely on the identical external features of the specimens without consideration for internal contents [41]. By contrast, cryoEM averages many individual particles to achieve high-resolution structures, allowing each particle's orientation to be determined. This capability is especially valuable for resolving structures of the internal features that do not conform to the same symmetry as the particle exteriors that determine crystalline packing, such as describing the genomic organization within viral capsids. Wah Chiu's group demonstrated this feat was possible for bacteriophage, even with the limited resolution of cryoEM in 2005 [42]. Since then, structures describing genome organization have been determined for a variety of virus types including those with double and single stranded nucleic acid genomes [42-46].

1.5. Double stranded RNA viruses

Double stranded RNA (dsRNA) viruses infect all manner of organisms, and exact a significant socioeconomic toll, despite their segmented genomes being alien to DNA-based life and thus a potent inducers of anti-viral responses [47,48]. This dynamic necessitates strategies unique amongst viruses, including continuous concealment of their genomes in multi-functional, icosahedral capsids, alongside viral RNA-dependent RNA-polymerases (RdRps) which synthesize plus-strand viral messenger RNAs (mRNAs) of each genome segment to be exported into the cytosol for translation and to act as templates for complementary strand synthesis within the newly formed capsids [47,48]. Thus, viral replication depends upon transcriptional considerations and, consequently, the RdRp containing TECs. In addition to their transcriptional roles, viral RdRps also assist in viral assembly, binding template mRNA to seed capsid formation, though it is unclear what role, if any, they play in the assembly of less complex viruses [49,50]. Our studies thus aimed to derive the mechanisms of assembly in comparatively simple systems and transcription initiation in more complex viruses from their respective capsid structures.

1.6. Revealing asymmetry: dsRNA viruses and cryoEM

Work by our lab and others established methods to resolve the asymmetric structures of icosahedral virus capsids and the genomes within them [42,43,51,52]. To do this, capsids are initially aligned based on their icosahedral symmetry which enables determination of the cartesian coordinates from the 2D micrographs corresponding to specific features on the 3D capsids. In the case of most dsRNA viruses, the TECs are located beneath the five-fold vertices of the icosahedron, so the 12 vertices from each particle can be located on the micrographs and extracted as individual particles, or sub-particles, with five-fold symmetry [43,51]. The particles are then be aligned based on the asymmetrically oriented TECs beneath each 5-fold sub-particle, and these orientations can be applied to the entire capsid to resolve its true asymmetric structure [43].

Leveraging these cryoEM techniques, we resolved the high-resolution structures of both a prototypical dsRNA virus from *T. vaginalis* and the economically important Aquareovirus (ARV). From the *T. vaginalis* virus (TVV), we discovered a minimally complex capsid lacking many of the stabilizing interactions found in extracellularly transmitted dsRNA viruses [53]. Further, we identified the putative site of viral mRNA egress and TEC attachment. These findings provide insights into the features of dsRNA viruses that are required for propagation within this system of minimal genomic density and complexity while leading an entirely intracellular lifestyle.

In the case of ARV, we structurally characterized the first transcriptionally inactive uncoated core particle of a reovirus from the *Spinareoviridae* family, representing its cytoplasmic state preceding transcription [54]. Comparison with previously resolved ARV structures suggested a novel mechanism of transcriptional initiation amongst multilayered members of the *Reoviridae* family of dsRNA viruses, whereby the outer capsid layers impede the elongation phase of mRNA transcription so that it does not begin prematurely [28,54-56]. This work helped to explain why the complete, multilayered reoviruses were incapable of mRNA synthesis prior to uncoating [57]. These findings were also supported by small confirmational changes to the TEC which appeared to prime the mRNA exit channel for transcriptional activity. Additionally, because the genome segments of ARV are partially organized by the TECs, our asymmetric reconstructions revealed the arrangement of the ARV core and showed no difference compared to the complete particle. Supporting the notion that these subtle structural changes undergone during capsid uncoating may facilitate but do not initiate transcription. In this study we employed cryoEM to inform on the assembly and transcriptional mechanisms that drive replication in this important class of viruses.

1.7. Herpesviruses: a quiet menace

Herpesviruses constitute a large family of dsDNA viruses characterized by a high degree of species specificity and the ability to establish both lytic and latent infection [58]. Among the eight known human herpesviruses, there is a broad range of tissue tropisms and associated diseases, ranging from the relatively mild chicken pox caused by varicella zoster virus (VSV) to more severe conditions such as cancer and multiple sclerosis in cases of Kaposi's Sarcoma herpesvirus and Ebstein Barr Virus respectively [58,59]. Given their medical relevance and ubiquity [60], significant efforts have been made to better comprehend these viruses and develop vaccines and therapies to mitigate their impact.

1.8. High-pressure herpesvirus capsids and cryoEM

Like dsRNA viruses, herpesviruses possess large icosahedral capsids which can be structurally characterized by cryoEM [61-63]. However, unlike dsRNA virus capsids, which assemble around mRNA in the cytosol of infected cells, herpesvirus capsids assemble within the nucleus around a core of scaffolding proteins [58,64]. These capsids are then filled with their exceptionally large genomes—ranging from 120 to 240 kilobases—through asymmetrical portals at one of their five-fold vertices. This packaging occurs under pressures exceeding 20 atmospheres [65], with the capsid walls being reinforced by external tegument proteins [61,63,66]. Amongst the herpesviruses, Human Cytomegalovirus is particularly interesting as it has one of the highest genomic densities, packaging a genome 51% larger than HSV-1 into a capsid only 17% larger [67].

Our lab previously determined that the HCMV specific tegument protein pp150 forms a net-like tegument layer around the HCMV capsid exterior, replacing most of the reinforcing tegument proteins observed in other herpesviruses [61]. This suggests pp150 plays a critical role in stabilizing the capsid against its exceptionally large genome [61]. Building on this, we aimed to characterize the interaction between the capsid and pp150 by introducing mutations at the interface between pp150 and its binding partners.

During this study, we thoroughly characterized the interface between pp150 and the associated major and small capsid proteins (MCP and SCP, respectively). We focused our mutagenesis efforts at the interface between pp150 and the SCP, and discovered most mutations at this small interface were lethal to the virus [68]. Additionally, we identified a mechanism by which pp150 may be trafficked to the nucleus—a question that has eluded researchers for decades. Our mutagenesis work also identified a single amino acid mutation in pp150 that slowed viral replication without significantly impacting capsid formation and maturation. This suggested a novel pathway for attenuated vaccine development, wherein virus replication can be tuned not by whole gene deletion, as has been attempted before [69-71], but by targeted mutations that preserve the complete antigenic signature of the virus.

Our work enhances our understanding of HCMV assembly and demonstrates the potential for structure to guide molecular biology researchers towards more effective treatments for disease.

1.9. Structure informs drug design

Like viral capsids, many complexes within pathogenic organisms, particularly those of the cytoskeleton, are composed of repeating subunits. One prominent example is the motile flagella—large cytoskeletal organelles which protrude from the surface of eukaryotic cells, and facilitate directional fluid flow and cellular locomotion through coordinated flagellar beating [72]. In *Trichomonas vaginalis* (*Tv*), flagella are integral to parasite locomotion, intercellular communication, and establishing the host pathogen interface, making them essential to the host-pathogen interactions [73].

Flagella structure is provided by evolutionarily conserved axonemes which are circular arrangements of dynein-decorated doublet microtubules (DMTs) [74-76]. Most motile flagella employ axonemes with a “9+2” architecture, wherein nine DMTs encircle a central pair of singlet microtubules connected via radial spokes. Axoneme bending, driven by dynein-mediated sliding of the DMTs, produces the characteristic flagellar beating motion. DMTs are assemblies of two protofilament rings: the A-tubule with 13 protofilaments and the B-tubule with 10. The protofilaments are composed tubulin α/β heterodimers [77,78], forming the basic 8 nm repeat of DMTs. Microtubule inner proteins (MIPs) bind the luminal surface of DMTs to help resist the mechanical stresses of axoneme bending and ensure MIP architecture repeats every 48 nm [74].

Despite the conserved axoneme morphology, the diversity of flagellar motion observed across different organisms suggests species-specific variations to axoneme architecture, as seen in structural studies of DMTs from the biflagellate *Chlamydomonas reinhardtii* algae and mammalian bovine trachea cilia [74,75]. *Tv* has four anterior flagella and lone recurrent flagellum that drives the undulating membrane [73]. While limited structural studies of trichomonad axonemes have shown they conform to the typical “9+2” architecture, the lack of high-resolution structures has hindered a more comprehensive understanding of these pathogenically relevant organelles [73]. To address this, we sought to leverage advancements in cryoEM, particularly in helical particle processing, to determine which MIPs are present, whether there are any novel MIPs, and how they may contribute to axonemal function in *Tv*.

Previous studies of the doublet microtubules from non-pathogenic organisms established methods to resolve these structures to near atomic resolution [74,79-81]. Similar to the alignment strategies used for icosahedral viruses, DMT particles can be aligned using larger features such the tubule protofilaments [74]. Given particles are picked every 8 nm, the MIPs, which run the length of these tubules and repeat every 48 nm, are initially averaged and appear to repeat every 8 nm. This artefact is resolved by sorting particles with aligned periodicity into the same classes. Thus, starting with six million 8 nm-long particles will be halved into three million 16-nm long particles and one million 48 nm-long particles.

Applying this workflow to *Tv*-DMTs we successfully resolved the first high resolution structure of the DMTs from this significant parasite [82]. In addition to the well conserved MIPs, we identified several *Tv*-specific proteins that lack homologues in other organisms. The novel MIPs included one which seemingly stabilized the junction between the A- and B-tubules and appeared to bind a ligand in a novel binding pocket. Additionally, we identified the first reported instance of true 96 nm periodicity within the MIPS of a DMT. These structural insights provide a deeper understanding of flagellar motility within this pathogenic organism and demonstrate cryoEM's capacity to identify and inform on drug design to develop improved therapies.

1.10. Overview

This dissertation is divided into six chapters, each examining an important architectural feature of pathogenic microbes. Chapter 1 introduces cryoEM and single particle analysis and its implications for study in each of the microbes investigated here. Chapters 2-4 delve into unique applications of cryoEM and SPA to inform on genome management strategies amongst double strand nucleic acid viruses. Chapter 2 provides the first detailed structural characterization of a minimally complex dsRNA virus, while Chapter 3 demonstrates cryoEM's ability to elucidate the transcriptional activation mechanism of the most complex families of dsRNA viruses, and Chapter 4 illustrates the capacity for cryoEM structures to inform on the rational design of live-attenuated herpesvirus mutants. Finally, Chapter 5 identifies several novel proteins from locomotive flagella of T. vaginalis including one potential therapeutic target. Chapter 6 provides a brief conclusion to this dissertation and a description of future of cryoEM and its related methodologies.

1.11. References

1. Watson, J.D. & Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).
2. Crick, F.H.C. & Watson, J.D. Structure of Small Viruses. *Nature* **177**, 473-475 (1956).
3. Caspar, D.L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology* **27**, 1-24 (1962).
4. Franklin, R.E. & Holmes, K.C. Tobacco mosaic virus: application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. *Acta Crystallographica* **11**, 213-220 (1958).
5. Abad-Zapatero, C. et al. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* **286**, 33-39 (1980).

6. Richardson, J.S. The Anatomy and Taxonomy of Protein Structure. in *Advances in Protein Chemistry*, Vol. 34 (eds. Anfinsen, C.B., Edsall, J.T. & Richards, F.M.) 167-339 (Academic Press, 1981).

7. Erickson, J.W. et al. Crystallization of a common cold virus, human rhinovirus 14: "isomorphism" with poliovirus crystals. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 931-934 (1983).

8. Rossmann, M.G. The Canyon Hypothesis. *Journal of Biological Chemistry* **264**, 14587-14590 (1989).

9. Hayden, F.G. et al. Efficacy and Safety of Oral Pleconaril for Treatment of Colds Due to Picornaviruses in Adults: Results of 2 Double-Blind, Randomized, Placebo-Controlled Trials. *Clinical Infectious Diseases* **36**, 1523-1532 (2003).

10. Sevvana, M., Klose, T. & Rossmann, M.G. Principles of Virus Structure. *Encyclopedia of Virology: Volume 1-5, Fourth Edition* **1-5**, 257-277 (2020).

11. De Rosier, D.J. & Klug, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. **217**(1968).

12. Henderson, R. & Unwin, P.N.T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28-32 (1975).

13. Crowther, R.A., DeRosier, D.J. & KLUG, A. The Reconstruction of a Three-Dimensional Structure from Projections and Its Application to Electron Microscopy. in *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* Vol. 317 319-340 (1970).

14. Dubochet, J. & McDowall, A.W. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. in *Journal of Microscopy* Vol. 124 3-4 (1981).

15. Dubochet, J., Lepault, J., Freeman, R., Berriman, J.A. & Homo, J.C. Electron microscopy of frozen water and aqueous solutions. in *Journal of Microscopy* Vol. 128 219-237 (1982).

16. Dubochet, J. et al. Cryo-electron microscopy of vitrified specimens. in *Quarterly Reviews of Biophysics* Vol. 21 129-228 (1988).

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

18. Prasad, B.V.V., Wang, G.J., Clerx, J.P.M. & Chiu, W. Three-dimensional structure of rotavirus. *Journal of Molecular Biology* **199**, 269-275 (1988).

19. Metcalf, P., Cyrklaff, M. & Adrian, M. The three-dimensional structure of reovirus obtained by cryo-electron microscopy. *The EMBO Journal* **10**, 3129-3136 (1991).

20. Böttcher, B., Wynne, S.A. & Crowther, R.A. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88-91 (1997).

21. Conway, J.F. et al. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **386**, 91-94 (1997).

22. Zhang, W. et al. Placement of the Structural Proteins in Sindbis Virus. *Journal of Virology* **76**, 11645-11658 (2002).

23. Kuhn, R.J. et al. Structure of Dengue Virus. *Cell* **108**, 717-725 (2002).

24. Mukhopadhyay, S., Kim, B.-S., Chipman, P.R., Rossmann, M.G. & Kuhn, R.J. Structure of West Nile Virus. *Science* **302**, 248-248 (2003).

25. Zhang, X. et al. Features of Reovirus Outer Capsid Protein μ1 Revealed by Electron Cryomicroscopy and Image Reconstruction of the Virion at 7.0 Å Resolution. in *Structure* Vol. 13 1545-1557 (2005).

26. Yu, X., Jin, L. & Zhou, Z.H. 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. in *Nature* Vol. 453 415-419 (2008).

27. Ge, P. & Zhou, Z.H. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proceedings of the National Academy of Sciences* **108**, 9637-9642 (2011).

28. Zhang, X., Jin, L., Fang, Q., Hui, W.H. & Zhou, Z.H. 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell* **141**, 472-482 (2010).

29. Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443-1444 (2014).

30. Milazzo, A.C. et al. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *Journal of Structural Biology* **176**, 404-408 (2011).

31. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. in *Nature* Vol. 587 152-156 (Springer US, 2020).

32. Scheres, S.H.W. Single-particle processing in RELION. in *Manuals* 1-21 (2013).

33. Scheres, S.H.W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. in *Journal of Structural Biology* Vol. 180 519-530 (Elsevier Inc., 2012).

34. Knoll, M. & Ruska, E. Das Elektronenmikroskop. *Zeitschrift f□r Physik* **78**, 318-339 (1932).

35. Franken, L.E., Grünewald, K., Boekema, E.J. & Stuart, M.C.A. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small* **16**(2020).

36. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics* **28**, 171-193 (1995).

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

38. Sigworth, F.J. Principles of cryo-EM single-particle image processing. 57-67 (2016).

39. Penczek, P., Radermacher, M. & Frank, J. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **40**, 33-53 (1992).

40. Van Heel, M. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111-123 (1987).

41. Hendrickson, W.A. Facing the phase problem. *IUCrJ* **10**, 521-543 (2023).

42. Jiang, W. et al. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**, 612-616 (2006).

43. Zhang, X. et al. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**, 531-534 (2015).

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

45. Zhou, K. et al. Atomic model of vesicular stomatitis virus and mechanism of assembly. *Nature Communications* **13**, 5980-5980 (2022).

46. Munke, A. et al. Primordial Capsid and Spooled ssDNA Genome Structures Unravel Ancestral Events of Eukaryotic Viruses. *mBio* **13**(2022).

47. Rushton, J. & Lyons, N. Economic impact of Bluetongue: a review of the effects on production. in *Veterinaria italiana* Vol. 51 401-6 (2015).

48. Ding, S.-W. & Voinnet, O. Antiviral Immunity Directed by Small RNAs. in *Cell* Vol. 130 413-426 (2007).

49. Roy, P. Bluetongue virus assembly and exit pathways. in *Advances in Virus Research* 1 edn Vol. 108 249-273 (Elsevier Inc., 2020).

50. Mata, C.P., Rodríguez, J.M., Suzuki, N. & Castón, J.R. Structure and assembly of double-stranded RNA mycoviruses. in *Advances in Virus Research* Vol. 108 213-247 (2020).

51. Liu, H. & Cheng, L. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science* **349**, 1347-1350 (2015).

52. He, Y. et al. In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating. in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 116 16535-16540 (2019).

53. Stevens, A., Muratore, K., Cui, Y., Johnson, P.J. & Zhou, Z.H. Atomic Structure of the Trichomonas vaginalis Double-Stranded RNA Virus 2. *mBio* **12**, 1-17 (2021).

54. Stevens, A., Cui, Y., Shivakoti, S. & Zhou, Z.H. Asymmetric reconstruction of the aquareovirus core at near-atomic resolution and mechanism of transcription initiation. *Protein & Cell*, 1-22 (2023).

55. Ding, K., Nguyen, L. & Zhou, Z.H. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. *Journal of Virology* **92**, e00774-18 (2018).

56. Wang, X. et al. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 7344-7349 (2018).

57. Farsetta, D.L., Chandran, K. & Nibert, M.L. Transcriptional Activities of Reovirus RNA Polymerase in Recoated Cores. *Journal of Biological Chemistry* **275**, 39693-39701 (2000).

58. Brown, J.C. & Newcomb, W.W. Herpesvirus capsid assembly: Insights from structural analysis. *Current Opinion in Virology* **1**, 142-149 (2011).

59. Bjornevik, K. et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**, 296-301 (2022).

60. Virgin, H.W., Wherry, E.J. & Ahmed, R. Redefining Chronic Viral Infection. in *Cell* Vol. 138 30-50 (2009).

61. Yu, X., Jih, J., Jiang, J. & Zhou, Z.H. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. in *Science* Vol. 356 eaam6892 (2017).

62. Liu, Y.-T.T., Jih, J., Dai, X., Bi, G.-Q.Q. & Zhou, Z.H. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. in *Nature* Vol. 570 257-261 (Springer US, 2019).

63. Wang, J. et al. Structure of the herpes simplex virus type 2 C-capsid with capsid-vertex-specific component. *Nature Communications* **9**, 3668 (2018).

64. Tandon, R., Mocarski, E. & Conway, J. The A, B, Cs of Herpesvirus Capsids. in *Viruses* Vol. 7 899-914 (2015).

65. Brandariz-Nuñez, A., Liu, T., Du, T. & Evilevitch, A. Pressure-driven release of viral genome into a host nucleus is a mechanism leading to herpes infection. *eLife* **8**, 1-20 (2019).

66. Dai, X. & Zhou, Z.H. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. in *Science* Vol. 360 eaao7298 (2018).

67. Liu, F. & Zhou, Z.H. Comparative virion structures of human herpesviruses. 27-43 (Cambridge University Press, 2007).

68. Stevens, A., Cruz-cosme, R., Armstrong, N., Tang, Q. & Zhou, Z.H. Structure-Guided Mutagenesis Targeting Interactions between pp150 Tegument Protein and Small Capsid Protein Identify Five Lethal and Two Live Attenuated HCMV Mutants. *bioRxiv* (2024).

69. Adler, S.P. et al. Phase 1 Clinical Trial of a Conditionally Replication-Defective Human Cytomegalovirus (CMV) Vaccine in CMV-Seronegative Subjects. *The Journal of Infectious Diseases* **220**, 411-419 (2019).

70. Das, R. et al. Safety, efficacy, and immunogenicity of a replication-defective human cytomegalovirus vaccine, V160, in cytomegalovirus-seronegative women: a double-blind, randomised, placebo-controlled, phase 2b trial. *The Lancet Infectious Diseases* **23**, 1383-1394 (2023).

71. Murata, S., Oshima, N., Iwasa, T., Fukao, Y. & Sawata, M. Safety, Tolerability, and Immunogenicity of V160, a Conditionally Replication-Defective Cytomegalovirus Vaccine, in Healthy Japanese Men in a Randomized, Controlled Phase 1 Study. *Antibodies* **12**(2023).

72. Zhou, F. & Roy, S. SnapShot: Motile Cilia. in *Cell* Vol. 162 224-224.e1 (Elsevier, 2015).

73. Coceres, V.M. et al. Ultrastructural and Functional Analysis of a Novel Extra-Axonemal Structure in Parasitic Trichomonads. in *Frontiers in Cellular and Infection Microbiology* Vol. 11 1-21 (2021).

74. Ma, M. et al. Structure of the Decorated Ciliary Doublet Microtubule. *Cell* **179**, 909-922.e12 (2019).

75. Gui, M. et al. De novo identification of mammalian ciliary motility proteins using cryo-EM. in *Cell* Vol. 184 5791-5806.e19 (Elsevier Inc., 2021).

76. Li, J.B. et al. Comparative Genomics Identifies a Flagellar and Basal Body Proteome that Includes the BBS5 Human Disease Gene. in *Cell* Vol. 117 541-552 (2004).

77. Nogales, E., Wolf, S.G. & Downing, K.H. Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, 199-203 (1998).

78. Nogales, E., Whittaker, M., Milligan, R.A. & Downing, K.H. High-Resolution Model of the Microtubule. *Cell* **96**, 79-88 (1999).

79. Chen, Z. et al. De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking. *Cell* **186**, 5041-5053.e19 (2023).

80. Tai, L., Yin, G., Huang, X., Sun, F. & Zhu, Y. In-cell structural insight into the stability of sperm microtubule doublet. *Cell Discovery* **9**, 116-116 (2023).

81. Leung, M.R. et al. Structural specializations of the sperm tail. *Cell* **186**, 2880-2896.e17 (2023).

82. Stevens, A. et al. Structures of Native Doublet Microtubules from Trichomonas vaginalis Reveal Parasite-Specific Proteins as Potential Drug Targets. *bioRxiv* (2024).

2. Chapter 2: Atomic Structure Of The Trichomonas Vaginalis Double-Stranded RNA Virus 2

Alexander Stevens[a,b,c,*], Katherine Muratore[a,*], Yanxiang Cui[b,*], Patricia J. Johnson[a], Z. Hong Zhou[a,b,#]

[a]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA

[b]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

[c]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA

*Alexander Stevens, Katherine Muratore, and Yanxiang Cui contributed equally to this article.

[#]Corresponding author

Keywords: cryoEM, sub-particle reconstruction, double-stranded RNA virus, *Trichomonas vaginalis*, *Totiviridae*

2.1. ABSTRACT

Trichomonas vaginalis, the causative pathogen for the most common non-viral sexually transmitted infection worldwide, is itself frequently infected with one or more of the four types of small dsRNA *Trichomonas vaginalis* viruses (TVV1-4, genus *Trichomoniasisvirus,* family *Totiviridae*). Each TVV encloses a non-segmented genome within a single-layered capsid and replicates entirely intracellularly, like many dsRNA viruses, and unlike those in the *Reoviridae* family. Here we have determined the structure of TVV2 by cryoEM at 3.6 Å resolution and derived an atomic model of its capsid. TVV2 has an icosahedral, T=2*, capsid comprised of 60 copies of the icosahedral asymmetric unit—a dimer of the two capsid shell protein (CSP) conformers, CSP-A and CSP-B—typical of icosahedral dsRNA virus capsids. However, unlike the robust CSP-interlocking interactions such as the use of auxiliary “clamping” proteins amongst *Reoviridae*, only lateral CSP interactions are observed in TVV2, consistent with an assembly strategy optimized for TVVs’ intracellular-only replication cycles within their protozoan host. The atomic model reveals both a mostly negatively charged capsid interior, which is conducive to movement of the loosely-packed genome, and channels at the five-fold vertices, which we suggest as routes of mRNA release during transcription. Structural comparison of TVV2 to the *Saccharomyces cerevisiae* L-A virus reveals a conserved helix-rich fold within the CSP and putative guanylyltransferase domain along the capsid exterior, suggesting conserved mRNA maintenance strategies amongst *Totiviridae*. This first atomic structure of a TVV provides a framework to guide future biochemical investigations into the interplay between *Trichomonas vaginalis* and its viruses.

2.2. Importance

Trichomonas vaginalis viruses (TVVs) are double-stranded RNA (dsRNA) viruses that cohabitate in *Trichomonas vaginalis,* the causative pathogen of Trichomoniasis, the most common non-viral sexually transmitted disease worldwide. Featuring an unsegmented dsRNA genome encoding a single capsid shell protein (CSP), TVVs contrast with multi-segmented dsRNA viruses, such as the diarrhea-causing rotavirus, whose larger genome is split into 10 dsRNA segments encoding 5 unique capsid proteins. To determine how TVVs incorporate the requisite functionalities for viral replication into their limited proteome, we derived the atomic model of TVV2, a first for

TVVs. Our results reveal the inter-subunit interactions driving CSP association for capsid assembly and the properties that govern organization and maintenance of the viral genome. Structural comparison between TVV2 capsids and those of distantly related dsRNA viruses indicate conserved strategies of nascent RNA release and a putative viral guanylyltransferase domain implicated in the cytoplasmic maintenance of viral messenger and genomic RNA.

2.3. Introduction

Trichomonas vaginalis (Tv) is a flagellated, parasitic protozoan responsible for the most common non-viral sexually transmitted infection (STI) worldwide, trichomoniasis, which afflicts nearly 280 million people each year [1]. Tv infection is associated with increased rates of developing aggressive genitourinary cancers, adverse pregnancy outcomes, and transmitting other STIs including human immunodeficiency virus [2-10]. Clinical Tv isolates are frequently found infected with one or more strain of the double-stranded RNA (dsRNA) *Trichomonas vaginalis* viruses (TVVs), belonging to the *Trichomoniasisvirus* genus within totiviruses [11,12]. The four phylogenetically distinct TVV strains (TVV1-4) share little sequence identity (<28%), and do not transmit extracellularly but are found to cohabitate the same organism [11-17]. TVV infection has been reportedly linked to upregulation of cysteine proteases and the major surface antigen, P270 in Tv, thought to contribute to subversion of the host immune system [18,19]. However, whether these effects are due to the virus or the protozoan strain remains unclear. This poorly understood relationship between TVVs and their pathogenic host has been compounded by the lack of detailed structural information which could help to interpret both clinical and laboratory observations.

TVV1-4 enclose 4.6-4.9 kbp, unsegmented genomes in single-layered icosahedral capsid shells (capsids) [11]. These TVV genomes contain two overlapping open reading frames encoding the 678-746 amino acid (aa) capsid shell protein (CSP) (74-82 kDa), and RNA-dependent RNA polymerase (RdRp) [11,14,15,17]. Due to ribosomal frameshifting, the RdRp complex is expressed as a C-terminal fusion protein (CSP-pol) of 1429-1481 aa's (combined, 159 to 165 kDa) [11,14,15,17]. Similarly, many other cytosol-bound dsRNA viruses exhibit this simplistic morphology, including members of the *Chrysoviridae*, *Partitiviridae*, and *Quadriviridae* families [20-22]. Because dsRNA is a powerful inducer of eukaryotic anti-viral defenses [23], the capsids of dsRNA viruses often continuously conceal the viral genome to prevent detection by the host. In Tv the burden of antiviral activity may fall on the putative RNA interference (RNAi) pathway of the protist, requiring passive dsRNA detection and providing limited downstream antiviral consequence [23-25]. As the dsRNA genomes remain inside the capsid, internal RdRp complexes are necessary to synthesize positive-stranded viral messenger RNA (mRNA) to be exported to the cytosol via capsid spanning channels of varying complexity. An earlier cryoEM study of the TVV1 virion resolved the icosahedral viral capsid to 6.7-5.5 Å resolution, and observed an icosahedral, T=2*, capsid composed of 60 asymmetric dimers and measuring 450-375 Å in diameter [26]. The authors identified pores spanning the entire capsid at the five-fold icosahedral (I5) vertices that were of reasonable diameter to act as channels for newly transcribed viral mRNA and, potentially, dsRNA [26]. However, the limited resolution and lack of available homology models precluded building an atomic model of the viral capsid.

Considering the prevalence and limited understanding of TVV infection among the ubiquitous *T. vaginalis*, we sought to elucidate TVV structures using cryoEM and single particle reconstruction techniques. The resulting cryoEM structure of TVV2 allowed us to characterize the structural features governing capsid assembly, mechanisms of genome sequestration, sites of mRNA release, and strategies of cytoplasmic maintenance of the viral mRNA. Additionally, comparisons of the TVV2 capsid with those of other dsRNA viruses featuring similar and distinct architectural characteristics offer insights into TVVs position in the evolutionary landscape of dsRNA viruses.

2.4. Results

2.4.1. 3D reconstruction of TVV2 virions.

TVV particles obtained from cellular subfractionation of Tv (strain G3) were imaged at a Titan Krios Electron microscope at 300 keV and nominal magnification of ×105,000 (see Methods below). From 2,034 micrographs, 2,493 particles were boxed and used for icosahedral reconstruction of the viral capsid to 4 Å (Fig. 1A and S1), based on the 0.143 “gold-standard” Fourier shell correlation coefficient (FSC) [27]. Sub-particle reconstruction about the I5 vertices improved the resolution of the decameric complex to ~3.6 Å by the same FSC criteria (Fig. S1). Some flexible loops along the capsid exterior appeared less well-defined than the ordered helices of the CSP cores, but the arrangement of 10 CSPs surrounding the I5 vertex is consistent with other totiviruses [26,28-30]. Using the molecular modeling software Coot [31], and guided by secondary structural element predictions from Phyre2 [32], TVV2 CSP residues 37-700 were built into a single decamer subunit, denoted capsid shell protein conformer A (CSP-A), nearest the I5 vertex. CSP-A was used as a homology model for building residues 37-701 into the alternative subunit (CSP-B) situated further from the I5 vertices. This A-B pair constitutes the principle icosahedral asymmetric unit (IAU) of the viral capsid (Fig. 1A and B), and model quality can be observed based on fit in the cryoEM map (Fig. 1C).

2.4.2. TVV2 capsid architecture and atomic models of the CSP conformers.

Like other totiviruses, the TVV2 viral capsid exhibits a T=2* icosahedral capsid shell, and measures ~430 Å across at its widest near the I5 vertices, down to ~350 Å at its narrowest between the icosahedral two-fold (I2) symmetry axes. The TVV2 capsid is composed of 60 IAUs, organized into 12 decameric units centered about the I5 vertices, enclosing the viral genome. Within these decamers (Fig. 2A), five CSP-As are seated nearest the I5 vertex, crowding out five CSP-Bs which sit partially intercalated between the flanking CSP-A's while extending between five-fold axes to contact CSP-B's from two other decamers forming the icosahedral three-fold (I3) symmetry axes. Both conformers measure approximately 110 Å by 55 Å, with thicknesses varying from ~45 Å near CSP centers to ~20 Å near the I2 axes (Fig. 2A) and tertiary structures resembling the profile of a left-handed mitten (Fig. 2B), with the apical “finger” ends situated nearest the I5 vertices (Fig. 2B and C). Extending outward from the I5 vertices, the narrow “fingers” transition into the thicker “palm” and “thumb”, before tapering to the thinner “cuff” end (Fig. 2B). This creates the appearance of raised plateaus extending outwards along the I5 axes and sloping steeply towards the I2 and I3 axes approximately 70 Å from the I5 vertices (Fig. 2A). The decameric assembly of the monomers seems to grasp the flanking CSP between its “thumb” and “fingers” with separation between monomers creating the appearance of a grooved surface (Fig. 2A).

Following the naming conventions established for the inner capsid layer of blue tongue virus (BTV) [33], we likewise separated the TVV CSP into three domains. These domains encompass 24 α-helices (α1-α24) and 21 β-strands (β1-β21) and are denoted as the; apical "fingers", nearest the I5 vertices; carapace "palm", near the CSP centers; and dimerization "cuff", nearest the sites of I2 symmetry (Fig. 2B and C). While the boundary between the carapace and dimerization domains is clear, as the tertiary structure of the latter is appreciably thinner and proportionally richer in β-strands (Fig. 2B), the separation between the carapace and apical domains is not obvious. However, the amino acid sequence reveals an insertion whose tertiary structure is replete with short β-strands and loops seated atop the carapace without interweaving itself into other domains (Fig S2).

The apical domain (residues 121-408) appears as an N-terminal insertion composed of 11 α-helices (α3-α13) and 8 short β-strands (β5-β12) (Fig. 2B and C, Yellow). This domain exhibits a β-sheet sandwiched between 6 α-helices (α5-13) nearer the carapace domain and 2 α-helices (α3-α4) which approach and surround the I5 vertices. Two loops between α3-α4 and β7-β8 extend polar residues (K157, R159 and N255) outward to line the I5 vertex generating the channels observed at the I5 axis (see Figs. 4C and S2). Between α11 and α12, what appears to be an anchor loop, tipped with hydrophobic residues, extends over the core helices of the carapace towards the dimerization domain where it is encircled by residues of the latter. The apical domains involvement at both the I5 vertex and interface between carapace and dimerization domains suggests it may participate in varied functional roles.

The carapace domain (residues 37-120, 403-567, and 678-701) appears to act as the stable molecular core, composed of 8 α-helices and 7 β-strands concentrated around the molecule center (Fig. 2B and C, Green). This domain contains a great number of the residues conserved across the TVV strains, along with many features implicated in the decameric subunit interface. Atop the structured core sit flexible loops that interact with and potentially stabilize adjacent domains and protrude from the CSP exterior forming the “thumb” of the conformers that insert into and stabilize the adjacent monomer (Fig. 2A). Several longer strands (β2-β4) extend in a perpendicular manner from the helices of the carapace, forming a large anti-parallel β-sheet which separates the dimerization domain from the I3 axes.

The dimerization domain (residues 568-677) is a C-terminal insertion composed of 5 helices (α19-α23) and 6 strands (β15-β20) which borders the Carapace domain near the I2 and I3 positions and exhibits the thinnest section of the capsid shell (Fig. 2B and C, Orange). The transition between the thick carapace and thin dimerization domains appears as a steep descent on the surface of the capsid creating a cleft lined with flexible coils originating from all three domains. The helices at the core of this domain are shorter than those observed in the others and appear sandwiched between β-sheets from both the dimerization and carapace domains. The 2 antiparallel β-sheets of this domain are conjoined perpendicularly to one another with a flexible loop protruding from the apex between β-15 and β-16.

CSP-Bs occupy a non-equivalent environment, further displaced from the I5 vertex compared to CSP-As, and introduce unique inter-subunit interactions. CSP-Bs exhibits similar morphological structure to CSP-As, as illustrated by the limited differences when they are superposed atop one another (Fig. 2D). Analysis by the Dali server indicated a root-mean-square deviation (rmsd) of 1.4 Å calculated between all 664 aligned Cα atoms (Z-score = 47.7) of the pair [34]. Key differences between the conformer structures are localized to sites of inter-subunit contact (Fig. 2D), as the involved residues apparently fulfil different roles between conformers. Notably, we observe an unwinding of α4 and α15 from CSP-A to CSP-B to accommodate the interface with the adjacent CSP-A subunits (Fig. 2E).

2.4.3. CSP-CSP interactions stabilize the TVV2 capsid.

The observed T=2* organization of the TVV2 capsid is shared amongst many dsRNA viruses which vary in their size and complexity, and incorporate differing numbers of capsid layers and stabilization strategies [35,36]. Regardless of the number of capsid layers or complexity, most dsRNA viruses employ interlocking domains such as the extensive N-terminal anchor or the auxiliary, "clamping" proteins of the single shelled cytoplasmic polyhedrosis virus (CPV) or domain swapping of partiti-, cryso-, and quadriviruses [20-22,37,38]. These interactions play an important role in maintaining the structural integrity of the viral capsid, concealing their dsRNA genomes from their hosts' innate immune systems, and protecting those with extracellular phases from environmental stress [20-23,37,38]. TVVs differ in these regards, as they have no extracellular phase and are not burdened by the robust anti-viral machinery of higher-order eukaryotes during replication in protists. Instead TVVs must, at most, avoid detection by the hypothesized Tv RNAi system which passively detects exposed dsRNA genome [24,25], begging the question as to whether TVVs must adhere to the same strategy for viral capsid assembly as their relatives in higher-order hosts.

Upon initial inspection, it was apparent TVV capsid shells lacked any of the aforementioned interlocking domains of their distant cousins. In lieu of distinguishable CSP-CSP interlocking interactions, we surmised the order of TVV2 capsid assembly based on the buried surface area between CSPs (Fig. 3A & Movie S1). The CSP-pair, A_1B_1 (Fig. 1A and 3A), represents a reasonable initial species formed within the cytosol as it buries the greatest surface area between subunits (~1540 Å^2). The alternative A_1B_2 interface is ~25% smaller (~1223 Å^2) and both A_1B_1 and A_1B_2 have greater buried surface area than any other CSP-CSP interface (Fig. 3A). The apical and carapace domains appear solely responsible for the interactions governing intra-decamer assembly, as the dimerization domains appear involved in anchoring the decamers together near the I2 positions. Interactions along the B_1A_1 interface (Fig. 3C) involve the unwound α4 from CSP-B_1 which accommodates a protruding loop from the adjacent CSP-A_1 subunit (Fig. 3D). Similarly, the unwound α15 of CSP-B_1 inserts between the long β-strands (β2-4 and β21) of CSP-A_1 and the dimerization domain of the CSP-A_3 from the tetrameric IAU (Fig. 3E). These interactions are absent along the A_1B_2 interface wherein α4 is wound and involved in A_1A_2 interfaces, which likely coordinates assembly of IAUs into decamers but not monomers into IAUs. Likewise, the proximity of CSP-As to the I5 vertex renders α15 unable to insert beneath the long β-strands of CSP-Bs. Additionally, the interface between the possible tetrameric intermediate, A_1B_1-A_3B_3 (~2525 Å^2), is ~56% larger than the alternative A_1B_1-A_2B_2 tetramer (1619 Å^2) and ~80% larger than the A_1B_2-A_4B_3 tetramer (~1405 Å^2), suggesting the first may be more stable and therefore more abundant in the cytosol (Fig. 3A). Interfaces at the I3 axes exhibit middling buried surface area between each monomer interface (~711 Å^2), but interactions appear localized to a protrusion (residues 671-674) beneath β21 of the adjacent CSP-B, indicating electrostatic interactions between R674 and D568 of neighboring monomers (Fig. 3F). The dimerization domains of the A and B subunits (Fig. 3A, CSP-A_3 and CSP-B_1) exhibit parallel β-sheets (both β16, β17, and β19) with the slight twist within β-17's creating the appearance of a near β-sheet, if not for the approximately 6 Å separating them (Fig. 2G). The protruding loop from the floor of the dimerization domain (residues 610-620) appears rich in positively charged residues (R612, R617, K616, R619, R620) and inserts into a negatively charged pocket between the carapace and dimerization domains of tetramer subunits (Fig. 3H). This insertion may act as an anchor to stabilize the association of the capsid decamers through electrostatic interactions, protecting the viral genome in place of more robust interlocking strategies. Based on our observations, it appears TVV2 employs a distinct strategy to assemble and stabilize the capsid in a manner optimized to conceal the dsRNA genome in the homeostatic cytosolic environment.

2.4.4. TVV2 capsids organize and coordinate the viral genome.

In contrast to the grooved exterior of TVV2 capsids, the interior appears smooth, as demonstrated in a cross section of the capsid at an arbitrarily high contour level (Fig. 4A). This planar view also reveals three concentric ring-like densities, likely corresponding to the dsRNA genome, featuring an approximate inter-duplex distance of 33 Å (Fig. 4A), slightly greater than those of other dsRNA viral genomes which measure around 27-30 Å [39,40]. The outermost ring of the TVV2 genome appears traced along the interior surface of the capsid in a quasi-hexagonal arrangement, separated from the interior wall by as little as ~10 Å below the I2 positions and up to ~30 Å at the I5 vertices

(Fig. 4A). In addition to increased separation, the outer ring of the dsRNA genome exhibits perturbations beneath the I5 vertices, suggesting that a structure poorly conserved during icosahedral averaging occupies this space. The location of the C-terminal tail of CSP, thought fused to the RdRp complex, sits adjacent to the negative space beneath the I5 portal (Fig. 4A, white arrows).

The capsids of dsRNA viruses frequently exhibit electronegative interiors, hypothesized to improve transcriptional efficiency amongst other dsRNA viruses of low genomic densities [36,41]. Electrostatic potential maps of the atomic model revealed the varied charge properties of the TVV2 capsid (Fig. 4B). The interior capsid surface appears predominantly electronegative (red) with discrete electropositive regions (blue) including the protruding loop of the dimerization domain and residues along α1 nearer the I5 vertices (Fig. 4B). Additionally, secondary structure predictions of the 36 unmodeled N-terminal CSP residues indicate a flexible domain with a calculated isoelectric point of 3.8. The position of α1 along the CSP interior suggests this flexible domain may extend inward and provide greater negative character along the capsid's inner surface. These charge properties are reminiscent of those observed along the interior of fungal dsRNA viruses in addition to picobirnaviruses wherein these interactions are hypothesized to prevent the association of the genome with the capsid surface via electrostatic repulsion of the negatively charged phosphate backbone [36,41]. This phenomenon, combined with the lower genomic density of TVV relative to other dsRNA viruses (Table 1), likely explains the consistent separation observed between the capsid and template, and contributes to increased template motion as reflected by the poorly resolved nature of the dsRNA genome (Fig. 4A).

As dsRNA viruses must export transcription intermediates into the cytosol, we sought to locate a site capable of such activity. As noted in the planar view, channels spanning the width of the capsid shell are located at the I5 vertices (Fig. 4A). These I5 channels are lined with residues extending from loops between α3-α4, and β7- β8 within the apical domain. Closer inspection of the decamer model from the outside reveals the residues lysine 157, arginine 159, and asparagine 255 from all 5 CSP-As lining this channel (Fig. 4C). A cutaway view through the electrostatic potential map of the I5 vertex illustrates the positively charged nature of this channel (Fig. 4D). The channel measures approximately 9.2 Å in diameter nearest the capsid interior, to 12.8 Å near the exterior, and this contrasts with the 16-20 Å I5 channel of TVV1 derived from a lower resolution cryoEM map in a previous investigation (Fig. 4D) [26]. The internal diameter of the TVV2 I5 channel may be expanded to approximately 10.8 Å via conformational rearrangement of K157 residues, providing a reasonable egression site for viral mRNA. Sequence alignment of the major TVV strains (TVV1-4) indicates K157, R159 and N255 residues are conserved despite divergent sequences, suggesting these residues may be necessary for channel functionality (Fig. S2). Structural comparison of the TVV2 channel to that of CPV from *Reoviridae* reveals shared features including similar channel diameters of 11.8 to 13.1 Å and linings of positively charged residues (Fig. 4D and E), despite significantly different capsid size and complexity [42].

2.4.5. Putative guanylyltransferase sites on the TVV2 capsid exterior.

The global architecture of TVV2 capsids is reminiscent of the well characterized L-A virus of *Saccharomyces cerevisiae* (ScV-L-A) [28]. Both viruses encode a single capsid protein, 120 copies of which encapsidate unsegmented dsRNA genomes [15,43]. Although they share little sequence identity (<19%) (Fig. S2), both TVV2 and ScV-L-A virions exhibit T=2* icosahedral symmetry and similar α-helix rich folds [Dali server reported an rmsd of 4.7 Å across 365 aligned Cα residues (Z score = 7.8) of the two proteins]. In particular, superposition of the two proteins reveal that 5 α-helices and 7 β-strands are shared between the molecule cores (Fig. 5A and B). Moreover, comparison of the well characterized guanylyltransferase (GTase) site of Gag to the corresponding position along CSP-A, revealed a cleft with similar amino acid composition including those integral to ScV-L-A GTase activity (Fig. 5C and E) [28,44]. In the TVV2 CSP-A we observe positively charged residues (R74, K523, R525, K539, R646, and R653) arranged on the cleft exterior generating a positively charged region similar to that observed in ScV-L-A (Fig. 5D and F). Likewise, Histidine residues (H537, H648, and H658) line the cleft interior whilst the aromatics (Y368 and Y653) appear buried deeper within the cavity (Fig. 5C and E). In a similar manner to the ScV-L-A GTase site, the local residues in the CSP cleft are directed towards a central point, indicating where they may interact with host mRNA and coordinate the enzymatic capture of the 7-methyl-guanosine-5'-monophospahte (m^7GMP) cap (Fig. 5C and E). However, the precise positioning of these amino acids differ between the two viruses, including the position of histidine residues which are integral for cap cleavage in ScV-L-A, wherein they form the cap-His intermediate that precedes cap transfer to viral mRNA (Fig. 4C and E) [28,44,45].

2.5. Discussion

Among totiviruses, the single capsid layer is responsible for protecting and sequestering the viral genome while facilitating RdRp activity and the egression of viral mRNA into the cytosol [26,28-30]. The details revealed by our 3.6 Å structure of the TVV2 capsid represent a marked improvement over other protozoan totivirus capsid structures, which had previously been limited to ~6.0 Å resolution [26,30], insufficient to build atomic models. Importantly, the TVV2 atomic model we derived enables us to characterize the interactions among its capsid subunits and draw meaningful insights from the features governing capsid assembly and maintenance of the viral genome.

The proposed assembly pathways of icosahedral dsRNA viruses featuring T=2* capsids have historically belonged to two camps, both of which begin with the formation of the basic AU, 60 of which are used to generate the capsid [46]. The first pathway, often observed amongst *Reoviridae*, relies on the formation of compact decamers composed of 5 AUs surrounding the I5 axes, 12 of which combine to form the complete capsid [33,46-48]. The second pathway involves the formation of tetramers, composed of AU dimers, related about the I2-symmetry axis which have been suggested for a many single-layered dsRNA viruses [20,49,50]. In the first pathway, absent nucleating factors, the formation of decameric intermediates is kinetically disfavored as the incomplete decamers would have low concentration, and subsequent combinations of these complexes would be similarly impeded by slow diffusion and low concentration [50]. By contrast smaller, I2-related tetramers represent far more kinetically favorable intermediates, likely existing in much higher concentrations in the cytosol and functioning to conjoin neighboring decameric complexes. In the case of TVV2, we observe what may be the strongest complimentary electrostatic interactions between the I2 related CSPs, with positively charged protrusions inserting into negatively charged pockets of the I2 related CSPs (Fig. 3H). This, in addition to the significantly greater buried surface area between the A_1B_1-A_3B_3 tetramer than the alternative tetramers, suggests that A_1B_1-A_3B_3 represents the most stable and therefore most abundant capsid intermediate.

Unlike many dsRNA viruses, most totiviruses, including TVV2, do not transmit extracellularly, and are instead transmitted vertically from mother to daughter cells similar to plasmid inheritance [26,51-54]. However, the dsRNA genomes of these viruses present distinct challenges to replication which differentiate them from other extranuclear genetic material. As dsRNA is a potent inducer of eukaryotic antiviral response [23,55], the viral genome must be enclosed within a protective capsid, necessitating dsRNA specific replication strategies including transcription via capsid-associated RdRp complexes and coordinated maintenance of these transcription intermediates [26,28-30]. The TVV2 capsid appears to closely conform to the general morphology of previously characterized totiviruses and is similarly devoid of any features indicating extracellular transmission. As these capsids do not encounter the robust immune system found in higher-order eukaryotes or the harsh conditions of the extracellular environment, it is conceivable that the strategies governing capsid assembly and stability may vary greatly from those that do. Other dsRNA families such as *Chrysoviridae*, *Partitiviridae*, *Quadriviridae* and the well-characterized *Reoviridae* include multi- and single-layered viruses transmitted both extracellularly and intracellularly across hosts of varying complexity. Regardless of layers or mode of transmission all of these viruses incorporate specific domains that buttress the association between the subunits of their icosahedral T=2* capsid shells [20-22,37,38,56,57]. The TVV2 capsid displays similar T=2* symmetry, but does not exhibit comparable subunit interactions, and instead seems to rely on the abutting interface between the

CSP conformers. The TVV CSPs are thicker (~45 Å) than those of the reoviruses BTV (~35 Å) and CPV (~30 Å) [33,38], and this seemingly provides ample lateral contact area between TVV subunits (Movie S1). Interestingly, those dsRNA viruses featuring capsid stabilizing domains infect hosts whose innate immune systems utilize pattern-recognition receptors (PRRs) whereas totiviruses infecting hosts lacking PRRs, including ScV-L-A and now TVV2, appear to lack such domains [24,25,58,59]. Additionally, TVV capsids have been shown to be more prone to degradation under conditions of increased temperature or pH than the extracellularly transmitted totivirus infecting *Giardia lamblia*, a host which may employ similar anti-viral RNAi defenses to Tv [25,30]. Along with our observations of the associations between TVV2 subunits, this indicates these less robust interactions are likely sufficient to conceal the viral genome from the passive RNAi pathways and maintain capsid integrity within but not outside of the cytosol. This also suggests that the robust stabilizing strategies common amongst dsRNA viruses of higher-order eukaryotes, are not necessary for efficient replication in their protozoan bound relatives. As the dsRNA genome remains enclosed within the TVV capsid throughout its replication cycle, RdRp complexes are necessary to transcribe viral mRNA for translation at the host ribosomes. Further, it has been demonstrated that RdRp activity in other dsRNA virus families such as *Reoviridae* is dependent on encapsidation for transcriptional activity [60-62]. It is possible therefore that TVV RdRp activity may be likewise dependent on the capsid, with the shell perhaps facilitating RdRp-genome interaction and transcription. In the case of ScV-L-A, it is thought the translational efficiency of the -1 ribosomal frame shift leads to the incorporation of just 1-2 RdRps per capsid [63]. The RdRp of TVV2 is likewise thought to be expressed as a minor, C-terminal fusion protein to the CSP via -1 ribosomal frameshifting [15,57,64], however, the RdRp stoichiometry has not been firmly established. The limited number of RdRp complexes means that even when data sufficient for a high-resolution reconstruction of the viral capsid is available, there are often too few RdRp complexes to resolve a structure. Indeed, despite extensive efforts following previous strategies to resolve asymmetrically attached structures [60,65], we were unable to resolve any density corresponding to the RdRp complex. Further effort with a substantially larger number of images and an optimized image processing method is needed to locate and resolve the RdRp within TVV2. Notably, we observe the TVV CSP C-terminus, to which the RdRp is fused [15], beneath the I5 vertices and adjacent to the only visibly unoccupied space of reasonable size to accommodate the 83 kDa RdRp domain (Fig. 4A).

Consistent with other dsRNA viruses, like those in the *Reoviridae* family [60,65], our data suggest the placement of the RdRp complexes beneath the I5 vertices, and this may provide nascent viral mRNA direct access to the I5 vertex. As previously mentioned, the I5 channel is the largest capsid spanning feature of TVV2, measuring 9.2-12.8 Å across, making it the most reasonable site for mRNA egress. This diameter is significantly smaller than the 16-20 Å suggested in previous findings of TVV1 [26], however the positively charged residues lining the I5 channel in TVV2 appear conserved across TVV1-4, suggesting this feature may be present throughout TVVs (Fig. S2). Interestingly, the channel dimensions are consistent with those of distantly related *Reoviridae* members like CPV and are significantly smaller than the 18 Å openings at the I5 vertices of ScV-L-A [28,42]. During transcription of viral mRNA in TVV2, we suspect that the positively charged residues at the I5 channels act as attractants for the negatively charged phosphate backbone of the mRNA to coordinate its transport into the cytosol. It is conceivable that the proximity of these channels to the newly synthesized mRNA benefits export by limiting the opportunity for viral mRNA to adopt secondary structures within the capsid, further supporting our assignment of RdRp beneath the I5 vertices.

Though the capsid and RdRp complex are capable of synthesizing and transporting the viral mRNA to the cytosol, TVVs inability to synthesize a 5' m^7GMP cap leaves the mRNA susceptible to detection and degradation by the exoribonucleases of the host cell [66,67]. As such, an enzymatic GTase domain capable of scavenging the 5' caps from host mRNA and transferring them onto viral mRNA must be incorporated into the limited TVV2 proteome to ensure efficient translation. As the CSP is the only TVV2 protein with apparent access to cytosolic mRNA, the cap-snatching domain is very likely incorporated here. Structural comparison between the TVV2 CSP and Gag of ScV-L-A revealed similar α-helix rich folds and analogous histidine rich sites along their exterior surfaces (Fig. 5). These similarities suggest TVV2 may employ a comparable m^7GMP cap-snatching strategy with species specific variations. In TVV, the putative enzymatic cleft is less pronounced than the trench of ScV-L-A, but similar positively charged residues lining the exterior may nonetheless assist in attracting and coordinating the negatively charged phosphate backbone of host mRNAs. Aromatic residues within the putative enzymatic cleft (Y368 and Y655) may stabilize the guanosine ring of the host mRNA's 5' cap via π-π stacking interactions similar to those of other GTase sites [28,42]. The presence of multiple histidine residues suggests the putative TVV2 site may carry out the nucleophilic attack of the caps phosphate backbone to generate a cap-His intermediate in a similar manner as ScV-L-A [44]. Interestingly, despite low sequence identity, Clustal alignment revealed many of the residues in this region, including one histidine (H648) are conserved across all TVV strains (Fig. S2), suggesting these residues may be important to TVV replication. Having 120 highly similar CSPs containing a GTase domain may enable the decapitation of many times more host mRNA than are needed to provide caps for the nascent viral mRNA. As has been suggested for ScV-L-A [68], these decapitated host mRNAs may serve as decoys to distract the hosts mRNA degradative machinery preceding translocation of the vulnerable, uncapped viral mRNA. Likewise, the many m^{7}GMP caps decorating the capsid surface may give the viral mRNA more opportunities to acquire a cap before traveling to the ribosome. It is conceivable that such cap-snatching activity may also induce conformational changes within CSPs which initiate RdRp activity and/or expand the I5 channels providing transcriptional regulation.

If the cap transfer reaction catalyzed by the putative GTase of TVVs is reversible and somewhat promiscuous, TVV superinfection may prove mutually beneficial to the viruses. Here the capacity of the putative TVV GTase domain to transfer caps non-specifically between viral mRNAs would enable caps procured by one TVV strain to be transferred to the mRNA of another helping to establish superinfection. Though we cannot confidently assert the identity or strategy of the TVV cap-snatching domain at this time, the suggested scenario represents an exciting opportunity through which future biochemical and structural investigations may uncover a cap-snatching strategy shared amongst totiviruses.

2.6. Acknowledgement

This project is supported partly by grants from the US NIH (DE028583/DE025567/AI094386 to Z.H.Z. and R01AI103182/R33AI119721 to P.J.J). K.A.M. received support from NIH Ruth L. Kirschstein National Research Service Award AI007323. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by UCLA and grants from the NIH (1S10OD018111 and 1U24GM116792) and the National Science Foundation (DBI-1338135 and DMR-1548924).

2.7. Author Contributions:

Z.H.Z. and P.J. conceived the project; K.M. prepared samples; Y.C. recorded cryoEM images and reconstructed the 3D maps; A.S. analyzed the map and built atomic models; Z.H.Z. and A.S. interpreted results and wrote the paper; all authors edited and approved the paper.

2.8. Materials And Methods

2.8.1. Virus Preparation

T. vaginalis strain G3 was cultured in Diamond's modified Trypticase-yeast extract-maltose (TYM) medium supplemented with 10% horse serum (Sigma-Aldrich), 10 U/mL penicillin,10 µg/mL streptomycin (Gibco), 180 µM ferrous ammonium sulphate, and 28 µM sulfosalicylic acid [69]. Parasites, grown at 37°C and passaged daily, were harvested by centrifugation and washed twice with phosphate-buffered saline. The cells were resuspended in 50mL Buffer 1 (2% IPEGAL CA-630, 2% Triton X-100,10% Glycerol, 10mM Tris, 2mM EDTA, 150mM KCl, 2mM $MgSO_4$, 1mM DTT, 1X HALT protease inhibitors, pH 7.4) and lysed in a Stansted cell disrupter with 30 psi front pressure and 12 psi back pressure. The cell lysate was centrifuged at 1000g for 10 minutes, and the resulting supernatant was further centrifuged at 5000g for 30 minutes to obtain the pellet enriched with TVV particles. The pellet was resuspended in Buffer 2 (150mM NaCl, 50mM Tris, 2mM MgCl2, 1mM DTT, 1X HALT protease inhibitors) and was found to contain TVV particles based on cryoEM examination.

For cryoEM sample preparation, 2.0 µL of the sample was applied to a glow-discharged lacey grid (Ted Pella) and incubated for 1 min. The grid was then washed once with Buffer 2 to remove large debris on the grid. ~1.5 µl of buffer remained on the grid before being transferred into a manual plunger apparatus. The grid was then blotted manually and flash-frozen in liquid ethane. The frozen grids were stored in liquid nitrogen prior to cryoEM imaging.

2.8.2. CryoEM Imaging.

Movies of dose-fractionated image frames were acquired in a Titan Krios microscope (Thermo Fisher Scientific) equipped with a Gatan imaging filter (GIF) Quantum LS and a post-GIF Gatan K2 Summit direct electron detector operated in super-resolution mode at a nominal magnification of 105,000 x (yielding a calibrated pixel size of 0.68 Å on the sample level) with SerialEM [70]. The GIF slit width was set to 20 eV. The dose rate on the camera was set to ~8 electrons/pixel/s, and the total exposure time of each movie was 6 s, which fractionated into 30 frames of images with 0.2 s exposure time for each frame. Dose-fractionated frames were 2x binned (pixel size 1.36 Å) and aligned for beam-induced drift correction to generate both dose-weighted (used for final reconstruction) and dose-unweighted (used for manual screening, CTF determination and particle picking) averaged micrographs using MotionCorr2 [71].

2.8.3. Data Processing.

The defocus values of the micrographs were determined with CTFFIND4 [72]. Micrographs with ice contamination or defocus value outside the range −0.8 to −3 μm were discarded. From a total of 2177 micrographs, 2034 were selected, and 5076 TVV manually picked particles were boxed out in 480 square pixels and 2x binned to 240 square pixels (pixel size of 2.72 Å) to speed up further data processing with RELION 3.1 [73]. Orientation and center parameters of each particle image were first determined and refined with icosahedral symmetry, yielding a 5.6 Å reconstruction. After one round of 3D classification, 2493 high-quality particle images (49% of all particles) were selected and re-extracted by centering the coordinates for further refinement. Then these high-quality particles were subjected to another round of 3D refinement with icosahedral symmetry, yielding a 4.0 Å reconstruction (Fig. S1) finally with the benefit of CTF refinement.

To obtain a higher resolution reconstruction for model building, we used a sub-particle reconstruction strategy similar to those detailed in previously published papers [60,65,74]. Briefly, the RELION command "relion_particle_symmetry_expand" was used with the icosahedral symmetry option (I3) to expand the icosahedral orientation entry in the data file generated from icosahedral reconstruction step into 60 icosahedrally oriented entries in a new data STAR file. In these 60 entries, every five C5 related entries corresponded to one penton vertex, so only one entry of those five was kept, yielding a duplicated-entries-removed data STAR file. The icosahedral reconstruction (I3) orients in such a way that one of its penton vertices is in the positive z-axis so we could easily estimate its coordinates (for example, x = 0, y = 0 and z = 110 pixel in the bin1 map) in the map. The coordinates for this penton vertex and the duplicated-entries-removed data STAR file were then used as the input to re-extract 12 penton vertex sub-particles from every TVV virus particle image (29,916 sub-particles in total). By imposing C5 symmetry, RELION further refined the local orientation and center parameters of these vertex sub-particles, and we obtained a reconstruction of penton vertex capsid with an average resolution of 3.6 Å (Fig. S1).

The resolution of the cryoEM maps were estimated on the basis of the gold-standard FSC = 0.143 criterion (Supplementary Fig. 1) [27]. The cryoEM maps were sharpened with a B-factor (-80 $Å^2$) in RELION. The local resolution evaluations were determined by ResMap [75].

2.8.4. Atomic Modeling, Model Refinement and Graphics Visualization.

Based on the positions of discernable aromatic residues and guided by secondary structural predictions from Phyre2 TVV2 was identified as the resolved structure during *de novo* atomic model building in the macromolecular modeling software Coot [31,32]. The resulting model was improved with PHENIX following the procedure detailed previously [76]. The models were then refined by the PHENIX real-space refine function and validated by the wwPDB validation server [77]. Visualization of the atomic model, including figures and movies, are made using UCSF Chimera [78]. Sequence alignments were made using Clustal Omega and alignment visualizations were rendered using ESPript 3.0 [79,80]. Structural comparisons between capsid shell protein conformers were quantified using the Dali server [34]. Buried surface area between capsid subunits was calculated using ChimeraX “measure buriedArea” function [81].

2.9. Data Availability:

The cryoEM density maps and atomic models are deposited in the Electron Microscopy Data Bank and the Protein Data Bank under the accession code of EMD-23560 (7LWY) and EMD-23561 (7M12) for focused and complete assemblies respectively.

2.10. Figures

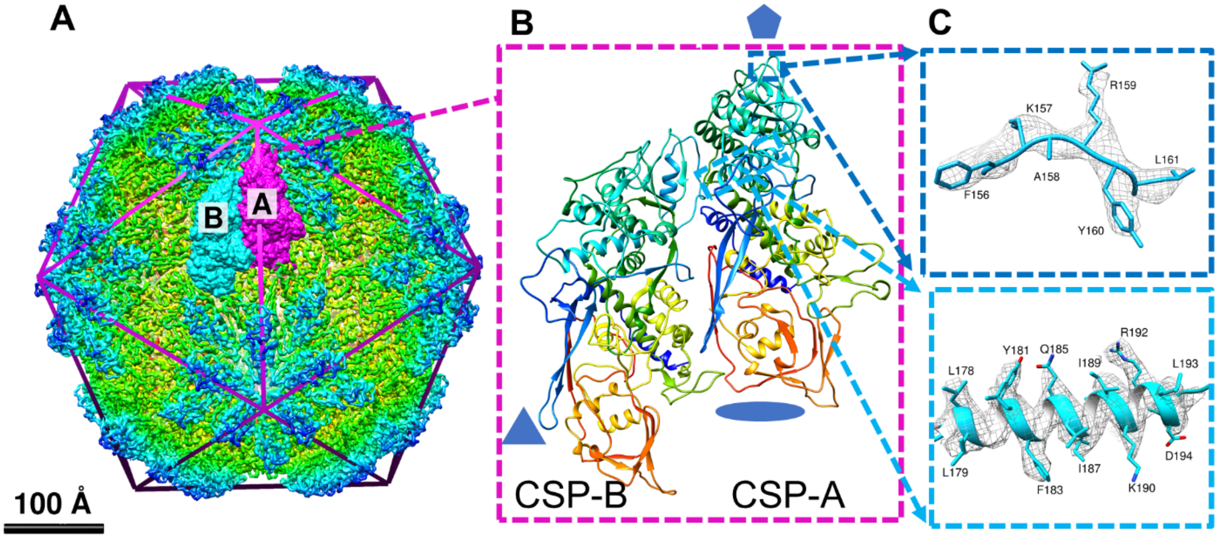


Figure 2- 1. CryoEM reconstruction and atomic model of TVV2 capsid

(A) 4.0 Å cryoEM map of TVV2 icosahedral viral capsid, radially colored (rainbow) to enhance detail. Icosahedron edges are overlaid to indicate symmetry sites. Two conformers, CSP-A (magenta) and CSP-B (cyan), are displayed as surfaces fit into the cryoEM map. (B) Ribbon diagram of the conformer pair as shown in the cryoEM map, rainbow-colored from blue (N-terminus) to red (C-terminus). Residues preceding S37 and following P700 (CSP-A) and K 701 (CSP-B) are not modeled. (C) Representative sample of the 3.6 Å sub-particle map (grey mesh) with side chains of residues 156-161 (top, blue), and atomic structure of residues 177-196 (bottom, cyan) shown as fit.

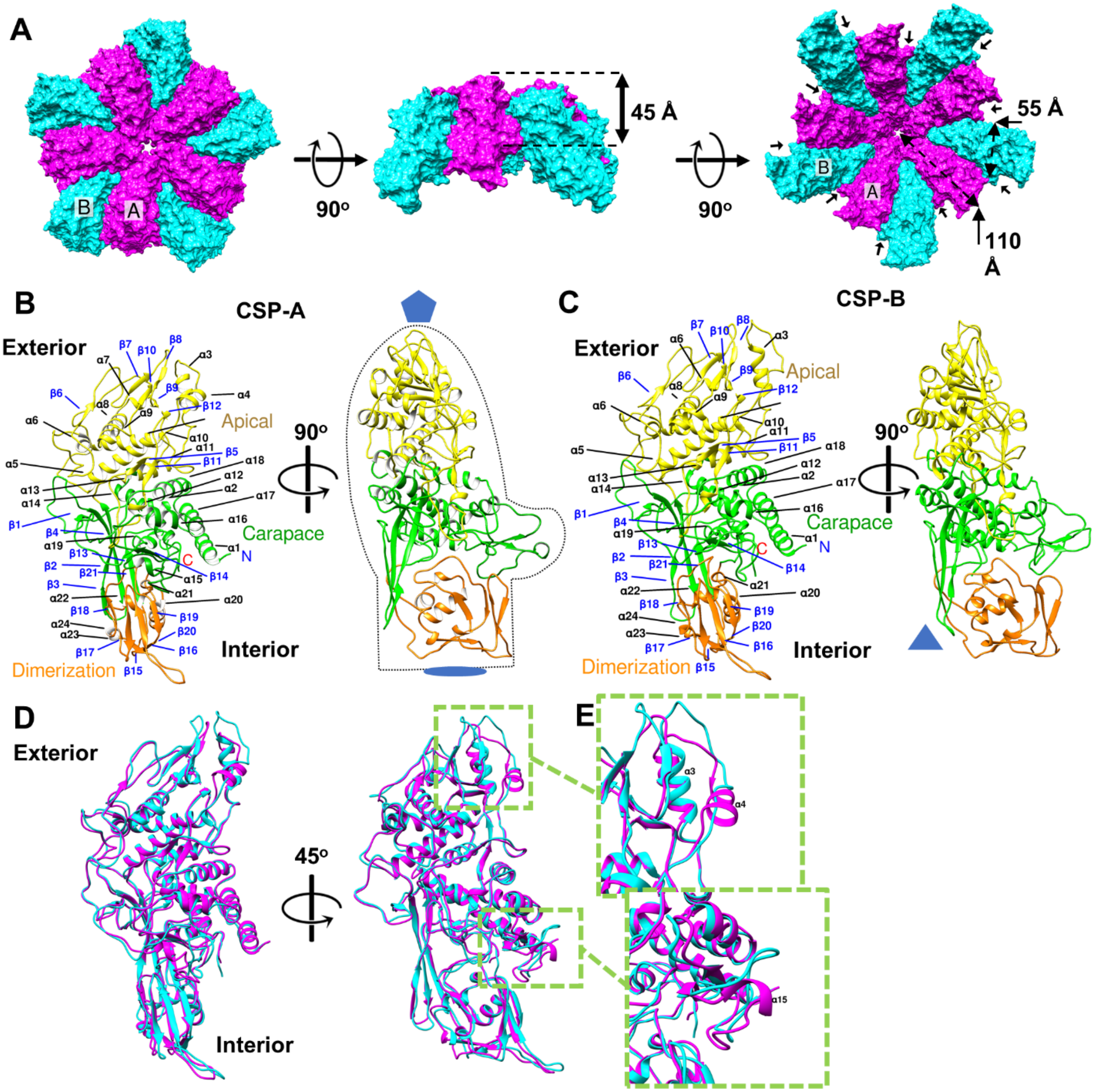


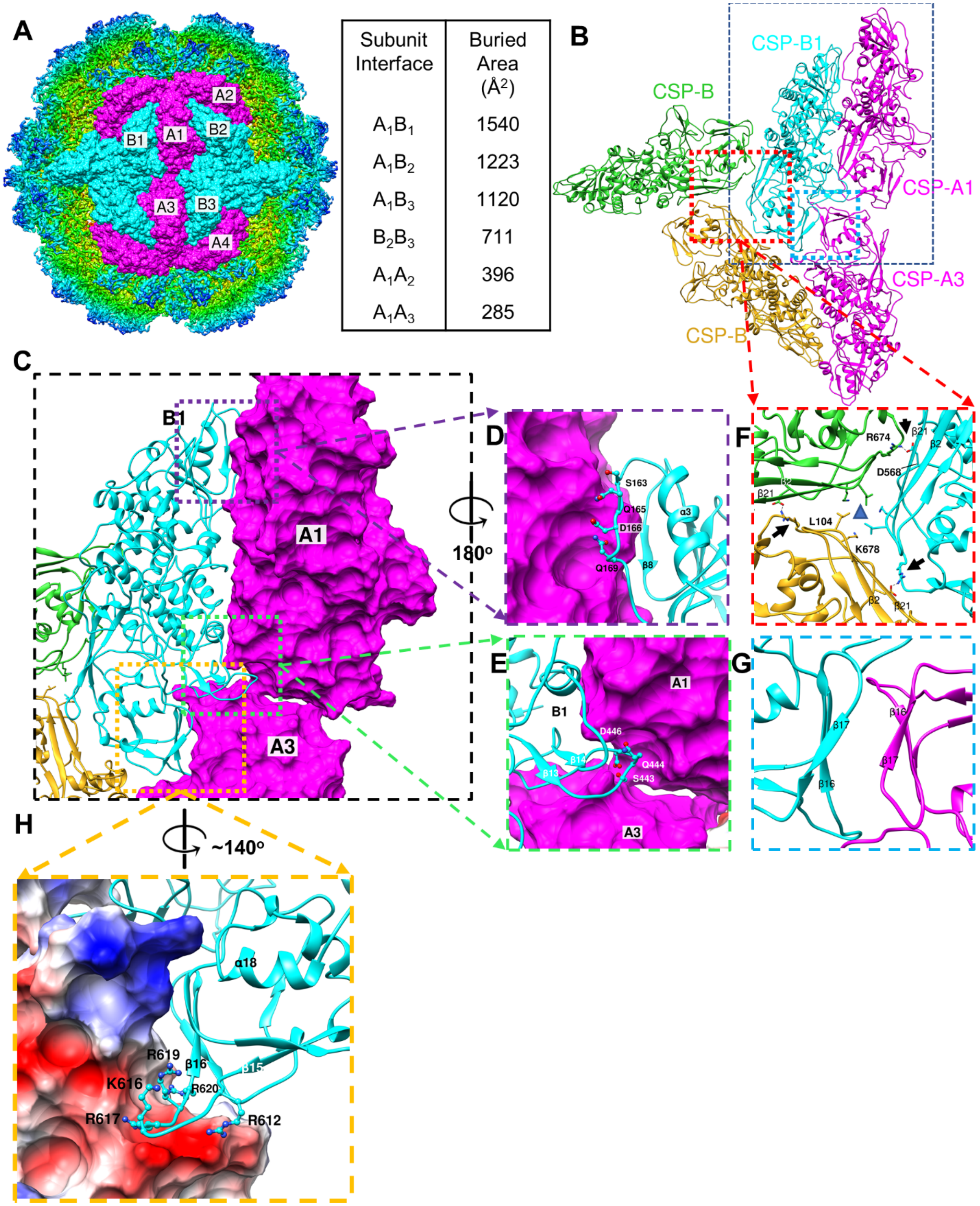
Figure 2- 2. Characterization of the capsid shell protein.

(A) Surface diagrams of decamers subunits with CSP-As (magenta) nearest the I5 vertex and surrounded by CSP-Bs (cyan) viewed from the exterior (left), side (middle), and interior (right). Approximate measurements of the conformer dimensions are overlaid in Å. Stabilizing protrusions are labeled with black pointers. (B and C) Orthogonal views of A and B conformers (CSP-A and CSP-B) respectively with secondary structural elements labeled and symbols indicating sites of icosahedral symmetry and mitten shape provided to enhance visualization. Colored; Carapace, Green; Apical, yellow; dimerization domain, Orange. (D) Ribbon diagram of CSP-A (magenta) and CSP-B (cyan) superimposed upon one another. (E) Close-up view of sites of structural discrepancies between conformers display on winding of α4 (top) and α15 (bottom) from CSP-A to CSP-B.

Table 2- 1. Genome packaging densities of double-stranded ribonucleic acid (dsRNA) viruses.

Virus	**Genome Features**			**Capsid features**		
Family, member	No. Genome Segments	Size (kbp)[b]	Length (μm)[c]	CSP (aa's)	Φ[d]/ir[e] (nm)	Genome density (bp/100nm^3)[f]
HCMV[a]	1	200	68	1370	130/45	38
Reoviridae						
Cypovirus	10	31.4	8.79	1333	58/24	54
Aquareovirus, GCRV	11	23.6	6.63	1027	60/23	46
Orbivirus, BTV	10	19.2	5.4	901	52/22	43
Rotavirus	11	18.5	5.18	880	52/23.5	34
Cystoviridae, phage Φ6	3	13.4	3.77	769	50/20	40
Totiviridae, TVV2	1	4.65	1.31	709	18	19
Totiviridae, ScV-L-A	1	4.6	1.29	680	43/17	21
Picobirnaviridae	2	4.2	1.18	590	35/14	36
***Partitiviridae**, PsV-S*	1(2)[g]	1.7/3.3	0.477	420	35/12	23
***Chrysoviridae**, PcV*	1(4)[g]	3.2/12.6	0.899	109	40/16	19
***Quadriviridae**, RnQV1*	1-2(4)	4.3/17.1	1.21	1356 + 1061	4716	25 (50)[h]

[a] Human Cytomegalovirus, a dsDNA virus, a lineage lacking genome segmentation, and characterized with B-from dsDNA.
[b] Approximate genome length
[c] Genome length calculated on rise/bp of 2.81 Å for dsRNa and 3.4 Å for dsDNA
[d] Approximate outer diameter
[e] Approximate inner radius
[f] Genome density when capsid cavity is assumed to be a perfect sphere containing only nucleic acids.
[g] PsV-S and PcV have genomes formed of 2 and 4 segments respectively.
[h] Density value of 25 if 1 dsRNA molecule per particle or 50 if there are two dsRNA molecules per particle.
Table adapted from Luque, D., Mata, C. P., Suzuki, N., Ghabrial, S. A. & Castón, J. R. Capsid Structure of dsRNA Fungal Viruses. *Viruses* **10**, 481 (2018).



Subunit Interface	Buried Area (Å^2)
A_1B_1	1540
A_1B_2	1223
A_1B_3	1120
B_2B_3	711
A_1A_2	396
A_1A_3	285

Figure 2- 3. CSP interactions coordinate capsid structure.

(A) Six copies of both CSP-A and CSP-B conformers fit into the TVV2 cryoEM map as viewed down the axis of I2 symmetry. Buried surface area (Å^2) between subunit interfaces are listed at right. (B) Ribbon diagram of TVV capsid proteins illustrating contact surfaces between two CSP-As (Magenta) and 3 CSP-Bs belonging to different decameric units (cyan, green, golden rod). (C) closer view of B_1A_1 interface with A_1 and A_3 surfaces displayed to illustrate contacts. (D) Close-up view of the unwound α4 of CSP-B with involved CSP-B residues shown. (E) View of the unwound α15 of CSP-B inserting beneath the long helices of the viral carapace domain. (F) View down three-fold axis with residues involved in threefold symmetry position shown residues displayed (G) Opposing β-strands with β17 from both CSPs demonstrating slight twist towards β-sheet like character. (H) electrostatic potential surface map displaying the positively charged residues of the β15-β16 loop of CSP-B, fit into the negatively charged (red) pocket between carapace and dimerization domains of CSP-A.

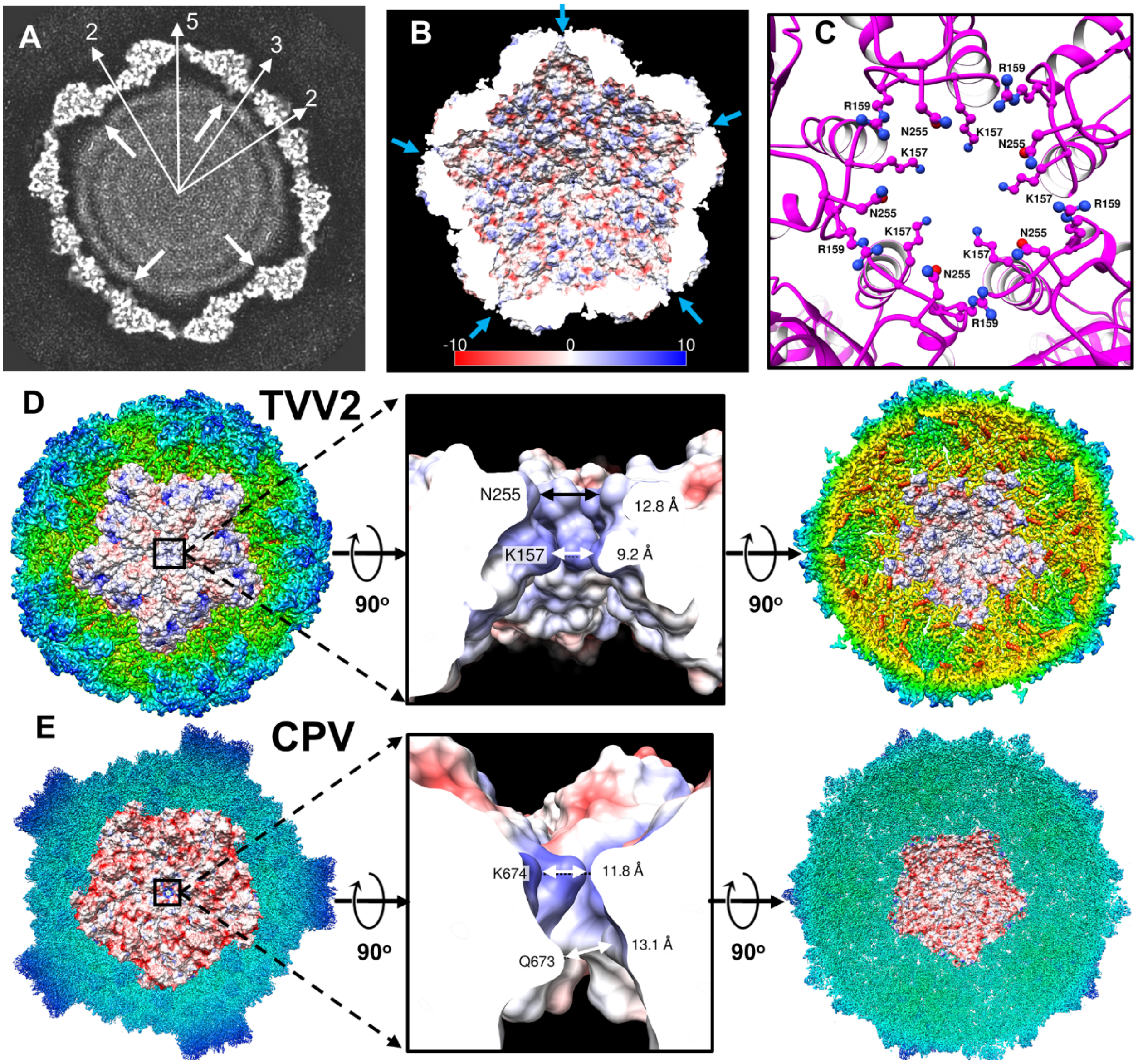


Figure 2- 4. TVV2 capsid coordinates genome arrangement and egression of transcription intermediates.

(A) 10 Å thick slab view of the Icosahedral viral capsid of arbitrary contour colored in greyscale with map densities in white and 0-values in black. Sites of icosahedral symmetry through the slab are labeled with (thin, white arrows), and arrows (thick, white) indicate the location of the C-terminus based on the atomic model. Distance between dsRNA strands measures ~30 Å. (B) Cutaway view of capsid interior with electrostatic potential diagrams illustrating positive (blue) and negative (red) surfaces. I5 vertices are marked with arrows (blue) (C) I5 vertex as viewed from the capsid exterior, K157, R159 and N255 residues are shown as ball and stick diagrams. (D) Exterior view of the TVV2 capsid protein decamer fit into the icosahedral map (radially colored) with surface colored by electrostatic potential (Left). 90^o rotation of A and cutaway view of the 5-fold channel, with local residues from an CSP-A subunit (K157, R159, and N255) and channel diameters labeled (middle). Interior view of TVV2 decamer surface fit into map (right). (E) Exterior view of the dsRNA virus CPV CSP in the quiescent state (RCSB: 3JAZ) decamer fit into the icosahedral map with surface colored based on electrostatic potential (Left). A cutaway view of the 5-fold channel, with involved residues from an CSP A conformer (K674 and Q673) and channel diameters labeled (middle). Interior view of CPV CSP decamer fit into map (right).

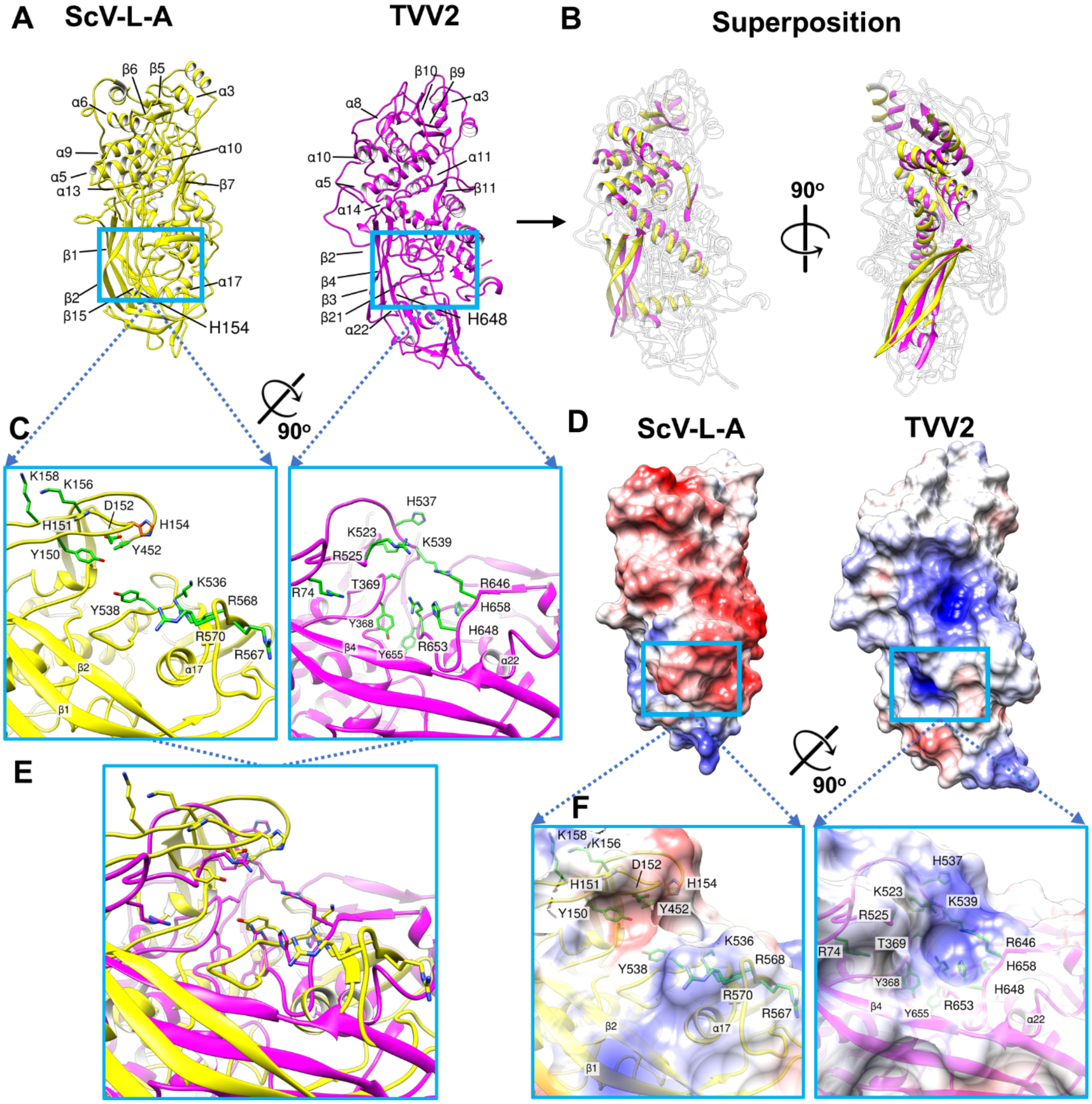
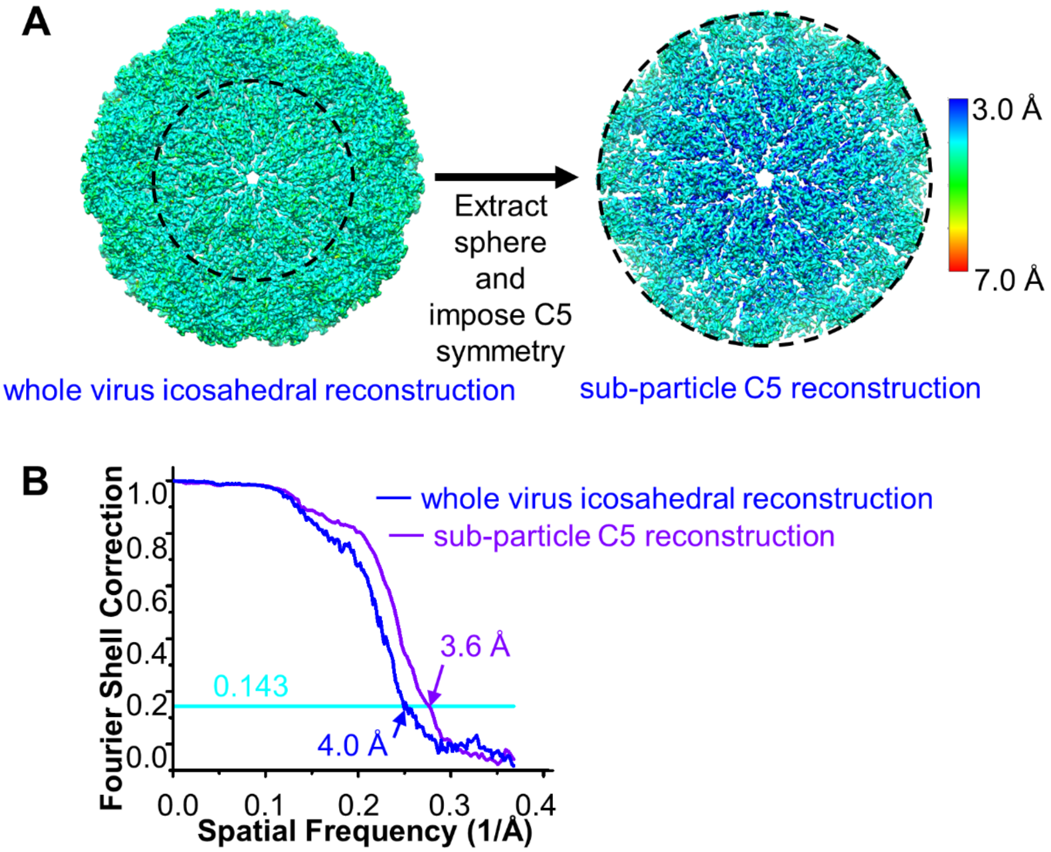


Figure 2- 5. Structural Comparison of TVV2 CSP to ScV-L-A Gag.

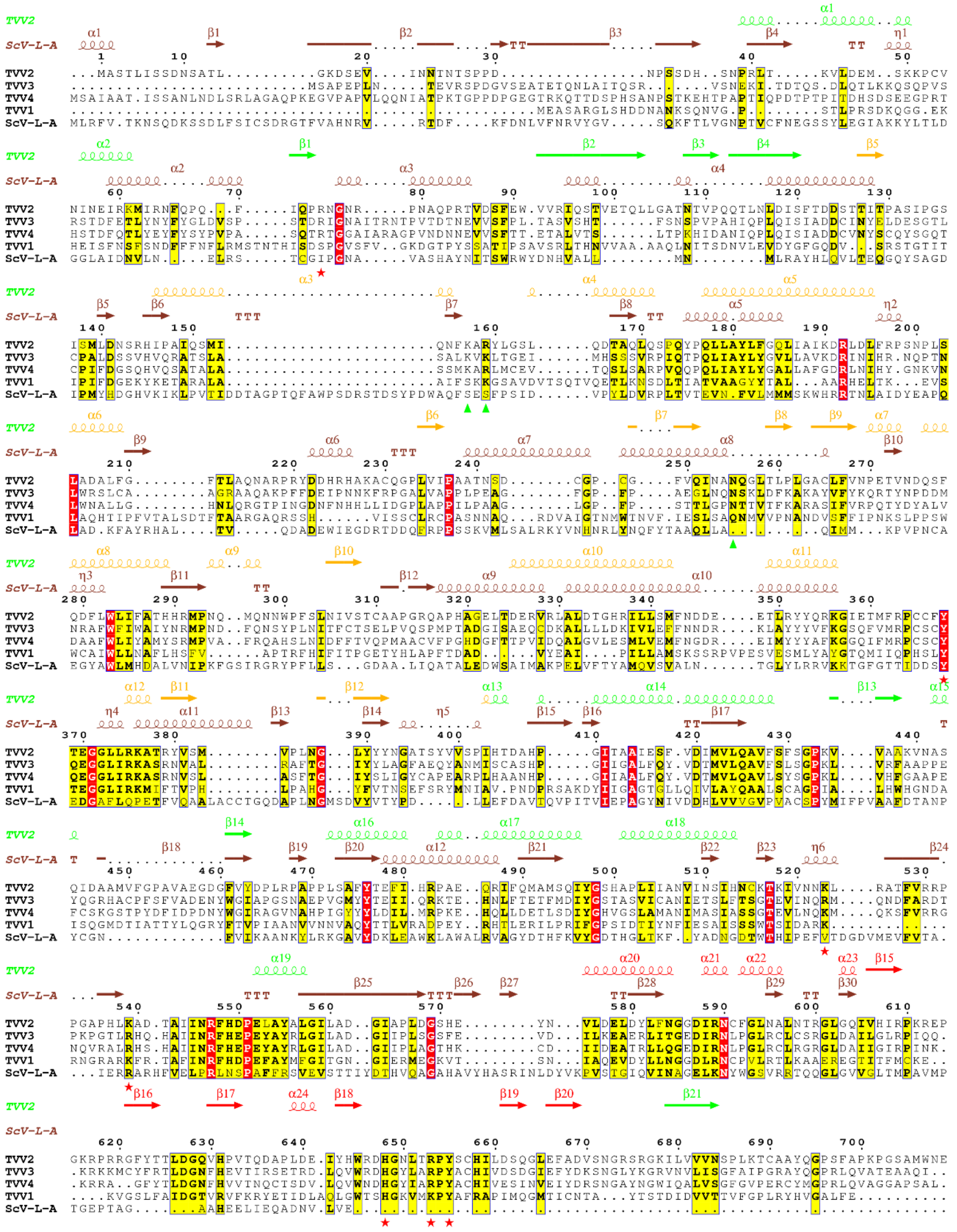
(A) Exterior view of ScV-L-A (yellow) and TVV2 (magenta) capsid shell proteins displayed with secondary structural elements of similar position labeled. (B) Orthogonal view of similar secondary structural elements between TVV2 and ScV-L-A capsid shell proteins superposed with outlines of the non-conserved segments. (C) Zoomed in view of ScV-L-A GTase site colored in yellow (left) with residues affecting in enzymatic activity

(Y150, D152, K156, K158 Y452, D540, Y538, K536, R567, R568, and R570) colored green and the essential His154 labeled in orange (left). The labeled and putative GTase site of TVV2 colored in magenta (right) with local residues colored green and labeled (R74, T369, Y368, K523, R525, H537, K539, R646, H648, R653, Y655, and H658) (right). (D) Exterior view of electrostatic potential maps of ScV-L-A (left) and TVV2 (right) CSP monomers, with negative (red) and positive (blue) charge distributions. Opacity changed to illustrate residue locations (E) Enzymatic sites from (C) superposed. (F) Close up views of the sites from (C) with surfaces colored based on electrostatic potential.



Supplementary Figure 2- S1

Resolution assessment. (A) Local resolution evaluation on the whole virus icosahedral reconstruction (left) and the sub-particle C5 reconstruction (right). (B) Fourier shell correction (FSC) coefficient as a function of spatial frequency (*i.e.*, 1/resolution) for the whole virus icosahedral reconstruction and sub-particle C5 reconstruction.



Supplementary Figure 2- S2

Sequence alignment of 4 recognized TVV strains and ScV-L-A illustrating disparate sequence identities. Secondary structural elements of TVV2 belong to the CSP-A conformer and are colored based on color scheme in Fig. 3B. Residues with 100% identity are boxed in red, similar residues are bolded and boxed in yellow. Conserved residues from the I5 channel of TVV2 are marked with arrows (green) and those from the putative GTase site are marked with stars (red).

2.11. References

1. WHO. Global incidence and prevalence of selected curable sexually transmitted infections. (2012).
2. Laga, M., Nzila, N. & Goeman, J. The interrelationship of sexually transmitted diseases and HIV infection: implications for the control of both epidemics in Africa. in *AIDS* Vol. 5 Suppl 1 S55-63 (1991).
3. Bowden, F.J. & Garnett, G.P. Trichomonas vaginalis epidemiology: parameterising and analysing a model of treatment interventions. in *Sexually transmitted infections* Vol. 76 248-256 (2000).
4. McClelland, R.S. et al. Infection with Trichomonas vaginalis increases the risk of HIV-1 acquisition. *The Journal of Infectious Diseases* **195**, 698-702 (2007).
5. Wendel, K.A. & Workowski, K.A. Trichomoniasis: Challenges to Appropriate Management. in *Clinical Infectious Diseases* Vol. 44 S123-S129 (2007).
6. Johnston, V.J. & Mabey, D.C. Global epidemiology and control of Trichomonas vaginalis. in *Current opinion in infectious diseases* Vol. 21 56-64 (2008).
7. Stark, J.R. et al. Prospective study of Trichomonas vaginalis infection and prostate cancer incidence and mortality: Physicians' Health Study. in *Journal of the National Cancer Institute* Vol. 101 1406-1411 (2009).

8. Allsworth, J.E., Ratner, J.A. & Peipert, J.F. Trichomoniasis and Other Sexually Transmitted Infections: Results From the 2001–2004 National Health and Nutrition Examination Surveys. in *Sexually Transmitted Diseases* Vol. 36 738-744 (2009).

9. Kissinger, P. & Adamski, A. Trichomoniasis and HIV interactions: a review. in *Sexually Transmitted Infections* Vol. 89 426-433 (2013).

10. Silver, B.J., Guy, R.J., Kaldor, J.M., Jamil, M.S. & Rumbold, A.R. Trichomonas vaginalis as a Cause of Perinatal Morbidity: A Systematic Review and Meta-Analysis. in *Sexually Transmitted Diseases* Vol. 41 369-76 (2014).

11. Goodman, R.P. et al. Clinical isolates of Trichomonas vaginalis concurrently infected by strains of up to four Trichomonasvirus species (Family Totiviridae). in *Journal of Virology* Vol. 85 4258-4270 (2011).

12. Fraga, J., Rojas, L., Sariego, I., Fernández-Calienes, A. & Nuñez, F.A. Species typing of Cuban Trichomonas vaginalis virus by RT-PCR, and association of TVV-2 with high parasite adhesion levels and high pathogenicity in patients. in *Archives of Virology* Vol. 157 1789-95 (2012).

13. Khoshnan, A. & Alderete, J.F. Multiple Double-Stranded RNA Segments Are Associated with Virus Particles Infecting Trichomonas vaginalis. in *Journal of Virology* Vol. 67 6950-6955 (1993).

14. Tai, J.-H. & Ip, C.-F. The cDNA sequence of Trichomonas vaginalis virus-T1 double-stranded RNA. in *Virology* Vol. 206 773-776 (1995).

15. Bessarab, I.N., Liu, H.-W., Ip, C.-F. & Tai, J.-H. The Complete cDNA Sequence of a Type II Trichomonas vaginalis Virus. in *Virology* Vol. 267 350-359 (2000).

16. Benchimol, M., Chang, T.H. & Alderete, J.F. Trichomonas vaginalis: Observation of coexistence of multiple viruses in the same isolate. in *FEMS Microbiology Letters* Vol. 215 197-201 (2002).

17. Bessarab, I.N., Nakajima, R., Liu, H.W. & Tai, J.H. Identification and characterization of a type III Trichomonas vaginalis virus in the protozoan pathogen Trichomonas vaginalis. in *Archives of Virology* Vol. 156 285-294 (2011).

18. Khoshnan, A. & Alderete, J.F. Trichomonas vaginalis with a double-stranded RNA virus has upregulated levels of phenotypically variable immunogen mRNA. in *Journal of Virology* Vol. 68 4035-4038 (1994).

19. Provenzano, D., Khoshnan, A. & Alderete, J.F. Involvement of dsRNA virus in the protein compositionand growth kinetics of host Trichomonas vaginalis. in *Archives of Virology* Vol. 142 939-952 (1997).

20. Pan, J. et al. Atomic structure reveals the unique capsid organization of a dsRNA virus. in *Proceedings of the National Academy of Sciences* Vol. 106 4225-4230 (2009).

21. Luque, D. et al. Cryo-EM near-atomic structure of a dsRNA fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. in *Proceedings of the National Academy of Sciences* Vol. 111 7641-7646 (2014).

22. Luque, D. et al. Heterodimers as the Structural Unit of the T=1 Capsid of the Fungal Double-Stranded RNA Rosellinia necatrix Quadrivirus 1. in *Journal of Virology* Vol. 90 (ed. Lyles, D.S.) 11220-11230 (2016).

23. Ding, S.-W. & Voinnet, O. Antiviral Immunity Directed by Small RNAs. *Cell* **130**, 413-426 (2007).

24. Carlton, J.M. et al. Draft Genome Sequence of the Sexually Transmitted Pathogen Trichomonas vaginalis. in *Science* Vol. 315 207-212 (2007).

25. Chen, X.S., Collins, L.J., Biggs, P.J. & Penny, D. High Throughput Genome-Wide Survey of Small RNAs from the Parasitic Protists Giardia intestinalis and Trichomonas vaginalis. in *Genome Biology and Evolution* Vol. 1 165-175 (2009).

26. Parent, K.N. et al. Structure of a Protozoan Virus from the Human Genitourinary Parasite Trichomonas vaginalis. in *mBio* Vol. 4 (ed. Racaniello, V.) 1-11 (2013).

27. Rosenthal, P.B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. in *Journal of Molecular Biology* Vol. 333 721-745 (2003).

28. Naitow, H., Tang, J., Canady, M., Wickner, R.B. & Johnson, J.E. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. in *Nature Structural Biology* Vol. 9 725-728 (2002).

29. Dunn, S.E. et al. Three-dimensional structure of victorivirus HvV190S suggests coat proteins in most totiviruses share a conserved core. in *PLoS pathogens* Vol. 9 e1003225 (2013).

30. Janssen, M.E.W. et al. Three-Dimensional Structure of a Protozoal Double-Stranded RNA Virus That Infects the Enteric Pathogen Giardia lamblia. in *Journal of Virology* Vol. 89 (ed. López, S.) 1182-1194 (2015).

31. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**, 486-501 (2010).

32. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. in *Nature Protocols* Vol. 10 845-858 (Nature Publishing Group, 2015).

33. Grimes, J.M. et al. The atomic structure of the bluetongue virus core. in *Nature* Vol. 395 470-478 (1998).

34. Holm, L. Using Dali for Protein Structure Comparison. in *Structural Bioinformatics* (eds Gaspari, Z. & Walker, J.M.) 29-42 (Humana, New York, 2020).

35. Reddy, V.S. & Johnson, J.E. Structure-Derived Insights into Virus Assembly. in *Advances in Virus Research* Vol. 64 45-68 (2005).

36. Luque, D., Mata, C.P., Suzuki, N., Ghabrial, S.A. & Castón, J.R. Capsid Structure of dsRNA Fungal Viruses. in *Viruses* Vol. 10 481 (2018).

37. Zhou, Z.H., Zhang, H., Jakana, J., Lu, X.-Y. & Zhang, J.-Q. Cytoplasmic Polyhedrosis Virus Structure at 8 A by Electron Cryomicroscopy: Structural Basis of Capsid Stability and mRNA Processing Regulation. in *Structure* Vol. 11 651-663 (2003).

38. Yu, X., Ge, P., Jiang, J., Atanasov, I. & Zhou, Z.H. Atomic Model of CPV Reveals the Mechanism Used by This Single-Shelled Virus to Economically Carry Out Functions Conserved in Multishelled Reoviruses. in *Structure* Vol. 19 652-661 (Elsevier Ltd, 2011).

39. Gouet, P. et al. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. in *Cell* Vol. 97 481-490 (1999).

40. Xia, Q., Jakana, J., Zhang, J.-Q. & Zhou, Z.H. Structural Comparisons of Empty and Full Cytoplasmic Polyhedrosis Virus. in *Journal of Biological Chemistry* Vol. 278 1094-1100 (2003).

41. Ortega-Esteban, Á. et al. Cryo-electron Microscopy Structure, Assembly, and Mechanics Show Morphogenesis and Evolution of Human Picobirnavirus. in *Journal of Virology* Vol. 94 (ed. López, S.) 1-17 (2020).

42. Yu, X., Jiang, J., Sun, J. & Zhou, Z.H. A putative ATPase mediates RNA transcription and capping in a dsRNA virus. in *eLife* Vol. 4 e07901 (2015).

43. Icho, T. & Wickner, R.B. The Double-stranded RNA Genome of Yeast Virus L-A Encodes Its Own Putative RNA Polymerase by Fusing Two Open Reading Frames. in *The Journal of biological chemistry* Vol. 264 6716-6723 (1989).

44. Tang, J. et al. The structural basis of recognition and removal of cellular mRNA 7-methyl G 'caps' by a viral capsid protein: A unique viral response to host defense. in *Journal of Molecular Recognition* Vol. 18 158-168 (2005).

45. Fujimura, T. & Esteban, R. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 108 17667-17671 (2011).

46. Poranen, M.M. & Bamford, D.H. Assembly of Large Icosahedral Double-Stranded RNA Viruses. in *Advances in Experimental Medicine and Biology* Vol. 726 379-402 (2012).

47. Reinisch, K.M., Nibert, M.L. & Harrison, S.C. Structure of the reovirus core at 3.6 Å resolution. in *Nature* Vol. 404 960-967 (2000).

48. McClain, B., Settembre, E., Temple, B.R.S., Bellamy, A.R. & Harrison, S.C. X-ray Crystal Structure of the Rotavirus Inner Capsid Particle at 3.8 Å Resolution. in *Journal of Molecular Biology* Vol. 397 587-599 (2010).
49. Poranen, M.M., Paatero, A.O., Tuma, R. & Bamford, D.H. Self-Assembly of a Viral Molecular Machine from Purified Protein and RNA Constituents. in *Molecular Cell* Vol. 7 845-854 (2001).
50. Kainov, D.E., Butcher, S.J., Bamford, D.H. & Tuma, R. Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. in *Journal of Molecular Biology* Vol. 328 791-804 (2003).
51. Wickner, R.B. Yeast virology. in *The FASEB Journal* Vol. 3 2257-2265 (1989).
52. Wang, A.L. & Wang, C.C. Viruses of parasitic protozoa. in *Parasitology today (Personal ed.)* Vol. 7 76-80 (1991).
53. Ghabrial, S.A. & Suzuki, N. Viruses of Plant Pathogenic Fungi. in *Annual Review of Phytopathology* Vol. 47 353-384 (2009).
54. Goodman, R.P., Ghabrial, S.A., Fichorova, R.N. & Nibert, M.L. Trichomonasvirus: a new genus of protozoan viruses in the family Totiviridae. in *Archives of Virology* Vol. 156 171-179 (2011).
55. Fichorova, R.N. et al. Endobiont Viruses Sensed by the Human Host - Beyond Conventional Antiparasitic Therapy. in *PLoS ONE* Vol. 7 e48418 (2012).
56. Zhang, X. et al. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. in *Nature structural & molecular biology* Vol. 23 74-80 (2016).

57. Lefkowitz, E.J. et al. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). in *Nucleic Acids Research* Vol. 46 D708-D717 (Oxford University Press, 2018).

58. Dyrka, W. et al. Diversity and Variability of NOD-Like Receptors in Fungi. in *Genome Biology and Evolution* Vol. 6 3137-3158 (2014).

59. Kawasaki, T. & Kawai, T. Toll-Like Receptor Signaling Pathways. in *Frontiers in Immunology* Vol. 5 461 (2014).

60. Zhang, X. et al. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**, 531-534 (2015).

61. Ding, K., Nguyen, L. & Zhou, Z.H. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. *Journal of Virology* **92**, e00774-18 (2018).

62. Ding, K. et al. In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. *Nature Communications* **10**, 2216 (2019).

63. Dinman, J.D. & Wickner, R.B. Ribosomal Frameshifting Efficiency and gag/gag-pol Ratio Are Critical for Yeast M1 Double-Stranded RNA Virus Propagation. in *Journal of virology* Vol. 66 3669-3676 (1992).

64. Liu, H.-W., Chu, Y.D. & Tai, J.-H. Characterization of Trichomonas vaginalis virus proteins in the pathogenic protozoan T. vaginalis. in *Archives of Virology* Vol. 143 963-970 (1998).

65. Cui, Y., Zhang, Y., Zhou, K., Sun, J. & Zhou, Z.H. Conservative transcription in three steps visualized in a double-stranded RNA virus. *Nature Structural & Molecular Biology* **26**, 1023-1034 (2019).

66. Koonin, E.V. & Moss, B. Viruses know more than one way to don a cap. in *Proceedings of the National Academy of Sciences* Vol. 107 3283-3284 (2010).

67. Simoes-Barbosa, A., Hirt, R.P. & Johnson, P.J. A Metazoan/Plant-like Capping Enzyme and Cap Modified Nucleotides in the Unicellular Eukaryote Trichomonas vaginalis. in *PLoS Pathogens* Vol. 6 (ed. Tschudi, C.) e1000999 (2010).

68. Masison, D.C. et al. Decoying the cap- mRNA degradation system by a double-stranded RNA virus and poly(A)- mRNA surveillance by a yeast antiviral system. in *Molecular and cellular biology* Vol. 15 2763-2771 (1995).

69. Clark, C.G. & Diamond, L.S. Methods for Cultivation of Luminal Parasitic Protists of Clinical Importance. in *Clinical Microbiology Reviews* Vol. 15 329-341 (2002).

70. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. in *Journal of Structural Biology* Vol. 152 36-51 (2005).

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

72. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. in *Journal of Structural Biology* Vol. 192 216-221 (Elsevier Inc., 2015).

73. Scheres, S.H.W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. in *Journal of Structural Biology* Vol. 180 519-530 (Elsevier Inc., 2012).

74. Zhang, Y. et al. Atomic structure of the human herpesvirus 6B capsid and capsid-associated tegument complexes. in *Nature Communications* Vol. 10 5346 (Springer US, 2019).

75. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. in *Nature Methods* Vol. 11 63-65 (2014).

76. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. in *Acta Crystallographica Section D Structural Biology* Vol. 75 861-877 (2019).

77. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **10**, 980-980 (2003).

78. Pettersen, E.F. et al. UCSF Chimera-A visualization system for exploratory research and analysis. in *Journal of Computational Chemistry* Vol. 25 1605-1612 (2004).

79. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. in *Molecular Systems Biology* Vol. 7 539 (2011).

80. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. in *Nucleic Acids Research* Vol. 42 W320-W324 (2014).

81. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14-25 (2018).

3. Chapter 3: Asymmetric Reconstruction of the Aquareovirus Core at Near-Atomic Resolution and Mechanism of Transcription Initiation

Alexander Stevens[1,2,3,#], Yanxiang Cui[2,#], Sakar Shivakoti[1,#], Z. Hong Zhou[1,2,*]

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

[2]California NanoSystems Institute, UCLA, Los Angeles, California 90095, USA

[3]Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90095, USA

[#] Contributed equally

*To whom correspondence should be addressed

3.1. Abstract

The *Reoviridae* family of dsRNA viruses is characterized by its members' capacity for endogenous transcription of their multipartite genomes within proteinaceous capsids of 1 to 3 layers. These viruses share inner core particles (ICPs) that conform to icosahedral, T=2*, symmetry, but differ in two major respects: first, the presence or absence of RNA-capping turrets at each icosahedral vertex; second, the number of additional host-specific capsid layers that are often lost upon cell entry. While the role of these additional layers in host infection is generally understood, the absence of asymmetric ICP structures from turreted, multilayered reoviruses has obfuscated our understanding of how successive removal of these external layers impact the structural organization of the ICP and transcription initiation. Here, we present the 3.3 Å resolution

structure of the aquareovirus (ARV) ICP, and atomic models of the capsid proteins VP3 and VP6, transcriptional enzymatic complex (TEC) subunits VP2 and VP4, and RNA-capping turret protein VP1. These structures reveal significant differences when compared to those of the coated ARV, as well as their counterparts in single-layered cytoplasmic polyhedrosis virus (CPV). Compared to the double-layered ARV virion and infectious subvirion particle structures, the ARV ICP undergoes significant capsid expansion and widening of the nucleotide processing channels in its TEC and turret. Thus, the loss of outer capsid layers may regulate transcription initiation in ARV, unlike CPV which relies solely on allosteric regulation by binding transcriptional cofactors. These results shed new light on the mechanism of transcription initiation amongst turreted, multilayered members of *Reoviridae*.

3.2. Introduction

Aquareovirus (ARV) is one of 15 genera from the *Reoviridae* family of double-stranded RNA (dsRNA) viruses (collectively called reoviruses), which causes hemorrhagic disease in economically important aquaculture of North America and Asia. *Reoviridae* members infect a wide range hosts and include the human borne rotavirus which causes approximately 128,000 deaths in young children each year [1]. As reoviruses are non-enveloped, host-cell susceptibility is determined by their proteinaceous capsids which enclose multisegmented genomes and are most readily distinguished by the number of icosahedral capsid layers, and the presence (*Spinareovirinae* subfamily, 9 genera) or absence (*Sedoreovirinae* subfamily, 6 genera) of messenger RNA (mRNA) capping turrets along the innermost layer [2]. Member viruses of both *Sedo*- and *Spinareovirinae* enclose their segmented dsRNA genomes within capsids of one to three concentric layers but, regardless of the number of genome segments, capsid layers, or turrets, their innermost layer is an icosahedral, T=2*, inner capsid particle (ICP). During cell entry, reoviruses often shed their outer layers leaving the ICP intact to evade the host's viral genome detection machinery, and act as nanomachines to endogenously transcribe positive-sense viral mRNA at transcriptional enzymatic complexes (TECs) situated underneath the icosahedral vertices within the ICP [3-5].

Among turreted reoviruses, CPV is the simplest, structurally, possessing a single-layered capsid equivalent to the ICP of double- or triple-layered reoviruses but which is also the intact virion [6-8]. Owing to its simple structural organization and more thorough characterization, CPV has long served as a model to study the lifecycles of other turreted reoviruses. When presented a susceptible cell, a quiescent CPV virion enters via endocytosis [9]. Upon cell entry, the CPV turret proteins (TPs) bind S-adenosyl methionine (SAM) and adenosine triphosphate (ATP), inducing subtle capsid expansion along with architectural changes to the TECs that facilitate transcription [10-12]. In the transcribing state, viral RNA-dependent RNA polymerases (RdRps), along with nucleotide triphosphatases (NTPases), transcribe viral mRNA, which receives its 5' cap from the turrets en route to the cytosol where they utilize the host's translational machinery.

However, the single-layered architecture that makes CPV an attractive model system also renders it an inadequate model to describe the numerous, multi-layered, *Spinareovirinae* members such as the double-layered ARVs whose outer layer(s) may prevent the same conformational changes to the inner capsid [10-12]. Endemic aquareoviruses in North America and China cause hemorrhagic diseases in golden shiner and grass carp, respectively [13,14]; and the cryo-electron microscopy (cryoEM) structures of the American and Chinese aquareovirus virion isolates have been resolved to 22 and 3 Å resolution, respectively [15-17]. These double-layered structures of ARV are highly similar to those of the mammalian orthoreovirus (MRV) virion [18,19], but are distinct to those of CPV. The second layer of ARV capsids give them an additional life stage between the quiescent and transcribing stages known as the primed state, which involves a more complicated array of structural changes and outer layer removal to prepare the virus for transcription [16].

The outer ARV capsid layer is comprised of small VP7 protection proteins and large VP5 penetration proteins. During infection, VP7 is cleaved extracellularly [20], exposing VP5 and creating a maximally infectious form of ARV known as the infectious subvirion particle (ISVP) (Nason et al., 2000). The ISVP uses the newly exposed penetration proteins to escape the endocytic pathway and deposit the transcriptionally competent ICP to the host cytosol (Fig. S1) [3,14,21,22]. Within the cytosol, the ICP can simultaneously synthesize and export viral mRNA from all genome segments and append 5' methylated guanosine caps to the nascent mRNA. This T=2* icosahedral core is composed of 120 copies each of the capsid shell protein (CSP) VP3 and the clamp proteins VP6. Each ICP encloses the 11 dsRNA segments and their accompanying TECs, each which is a heterodimer of RdRp VP2 and its cofactor, NTPase VP4. Asymmetric sub-particle reconstruction methods coupled with iterative symmetric relaxation workflows have enabled visualization of the asymmetrically attached TECs within the core of several dsRNA viruses [4,16,23]. To date, only the asymmetric structures of complete ARV virions and ISVPs have been resolved to high-resolution, thus it is unclear what, if any, structural changes occur within the ICP during uncoating [16,17]. Consequently, whether the external layers of the multi-layered turreted dsRNA viruses also play a structural and/or functional role in transcription regulation in addition to their involvement in cell entry is not known in the absence of a structure for comparison of ICPs to virions.

In this study, we report the first asymmetric reconstruction of the ARV ICP by cryoEM. Comparison with existing ARV virion and ISVP structures reveals subtle architectural changes between the coated and uncoated core particles which may act to facilitate endogenous transcription. These changes yield a particle with significantly increased internal volume and widened mRNA exit channels. Our structure also enabled modeling of previously unresolved TEC and CSP segments, which illustrate the interactions between the CSP N-termini, the TEC, and the adjacent genomic RNA absent the protective VP7-VP5 layer. As the first high-resolution asymmetric structure of the innermost core of a multi-layered member of the *Spinareovirinae* subfamily of *Reoviridae*, this structure fills in a knowledge gap in the ever-growing repository of asymmetrically reconstructed dsRNA viruses, demonstrating the conformational difference between coated and uncoated reoviruses, and informing on how those changes may help to regulate transcription.

3.3. Results and Discussion

3.3.1. Asymmetric reconstruction of ARV core resolving 11 TEC-occupying and 1 RNA-filled vertices

We determined the asymmetric structure of the ARV ICP by following a sequential symmetry relaxation approach as described previously [16,23]. Virus structures were initially refined with icosahedral symmetry imposed and finer details of the nonicosahedral related regions were resolved via symmetry expansion and focused classification as detailed in previous works [16,18] (Materials and Methods). Following this procedure, four asymmetric structures were obtained: three from the sorted I5 vertices, which include the polar vertices at 3.3 Å and the tropical vertices with and without the TECs to 3.3 Å and 3.7 Å respectively, along with a reconstruction of the complete capsid featuring 11 TECs at 4.2 Å. The resolution of TEC containing sub-particles following averaging of equivalent copies reached 3.4 Å. Our asymmetric reconstruction of the ARV core contains 60 VP1 turret proteins, 120 VP3 capsid shell proteins (CSPs), 120 VP6 clamp proteins, 11 TECs under 12 vertices, with only one southern tropical vertex lacking a TEC complex (Fig. 1F).

These 11 TECs within the ARV core are internally organized following pseudo-D_{3d} symmetry, with 6 TECs occupying the icosahedron's poles while 5 are situated along the equatorial plane at the northern and southern tropics (Fig. 1D). 11 dsRNA segments between 0.8 and 3.9 kbps compose the ARV genome and are tightly packaged inside the capsid (Fig. 1C); the outermost dsRNA layer shows large, persistent, parallel lengths for each RNA double-strand, with an average of 27 Å between neighboring strands, consistent with the previous ARV structures, despite the increased internal capsid volume (Fig. 1B-C) [16].

3.3.2. Loss of outer coat proteins leads to expansion of ICP capsid

Lacking the outer capsid proteins (OCPs) VP5 and VP7, the ARV ICP retains the icosahedral, T=2*, inner capsid shell composed of 60 asymmetric dimers of the 1,214-residue, wedge-shaped CSPs and 120 symmetrically arranged copies of the clamp protein (VP6) which form the ICP frame and provide support respectively (Fig. 1A).

Consistent with the previously resolved ARV capsid structures, VP3 dimers ($VP3_A$ and $VP3_B$) encircle each 5-fold vertex; with $VP3_A$ conformers seated around the 5-fold vertex center, creating pores adjacent to each TEC's template exit channel for direct transcript capping and release via VP1 turret proteins (TPs) (Fig. 1C-D). $VP3_B$ conformers partially intercalate between $VP3_A$ monomers and form 3-fold vertices with neighboring decameric assemblies (Fig. 1D). Absent conspicuous CSP rearrangements, the internal volume of the ICP increased by about 9.3% relative to the previously resolved grass carp reovirus (GCRV) virion and ISVP with which golden shiner reovirus (GSRV) shares 96-100% a.a. sequence identity (Fig. 1B) [16,17,24]. The motion undergone by individual decamers includes a 10 Å rise away from the virion origin and subtle expansion of the CSPs (Movie S2) [25]. The observed expansion can be attributed to a non-uniform elongation of the CSP monomers approximately 6 Å radially from the I5 vertices, relative to their coated counterparts, and this creates an expanding and rising motion around the decameric subunit (Movie S1,3). Motion isn't uniform across the CSP monomers however and differs between the VP3 conformers.

Per historic naming conventions, each ARV CSP is divided into three distinct domains. Extending radially from the I5 vertices, these domains include the α-helix rich apical "tip" nearest the I5 vertices (a.a 486-830), the large carapace domain (a.a. 190-485, 831-976, and 1144-1214), and the small β-sheet rich dimerization domains (a.a. 977-1143) (Fig 1D). The local motion of CSP dimers (Movie S2) reveal a non-uniform elongation of the apical and carapace domains, with the alpha helices migrating away from the I5 vertices and towards the icosahedral two-fold and three-fold (I2 and I3) vertices for $VP3_A$ and $VP3_B$ respectively (Movie S2). In the apical domain helices 13 and

14 and their conjoining loop (a.a. 490-518), which line the I5 transcript exit channel at the luminal side of the capsid, are largely unperturbed by the capsid shifts and help to maintain a consistent pore diameter. By contrast, the other I5 adjacent elements of the apical domains move away from the I5 channel, and the helices 13 and 14 of $VP3_{\mathbf{B}}$ migrate in a similar manner as the other secondary structural elements. Closer inspection reveals striking conservation of $VP3_A$ residues adjacent to the I5 pore (a.a. 490-518, RMSD .464 Å) when compared to the quiescent virion and ISVP structures [16,17]. Despite exhibiting outward motion, the secondary structural elements of CSPs which interface with the clamp and turret proteins interfaces appear to move as rigid bodies, presumably constrained by interactions with the essential clamp and turret proteins (Movie. S2). The expansion of the CSPs leads to a significant increase of internal capsid volume, from 5.51 $\times 10^7$ Å^3 to 6.02 $\times 10^7$ Å^3, or about 9% from the ARV virion to the ICP (Fig. 1B). By reducing the packaging density and therefore the viscosity of the genome, the enlarged ICP provides the ridged dsRNA segments greater freedom of motion and presumably lowers the energy required to initiate transcription [26].

The inner-outer capsid protein (VP3-VP5) interactions of ARV are mediated through the clamp protein VP6, making it important not only for stabilization against the genome but in secondary layer association. Previous work has shown the closely related MRV ICPs can be recoated to form ISVPs [27], suggesting the ARV ICP clamps remain in a VP5 receptive state following uncoating. Indeed, superimposition of the ICP and virion clamp proteins reveal significant conformational similarity (RMSD 0.667 Å) despite the migration of the clamps away from the I5 vertices in the ICP particles. The architectural changes of VP6 can, therefore, be summarized as shifts of 4-5 Å away from the I5 vertices along the capsid wall, with no conspicuous, irreversible changes that would inhibit OCP VP5 association. So, in addition to its role stabilizing the VP3 ICP against the force exerted by the genome, VP6 provides a platform onto which the VP5 layer can bind and possibly compress core particles to halt transcription. While this dynamic may be inconsequential for parental ARV particles, whose VP5 proteins are irreversibly cleaved during entry, it may be a key mechanism to halt transcription in newly assembled ARV particles following complimentary strand synthesis, as observed in nascent rotavirus particles [28].

3.3.3. Capsid expansion leads to subtle TEC rearrangements

Several dsRNA segments interact with each TEC and these interactions stabilize the genome, enabling improved visualization of their major and minor grooves within the ARV ICP as observed in other RNA viruses (Fig. 2A) [18,29]. 5 major dsRNA segments are observed adjacent to each TEC and are labeled “Front Bottom,” “Bound,” “Front Top,” “Rear Top,” and “Rear Bottom” RNA segments based on their positions relative to the TEC (Fig. 2A). At the polar TECs a 6th RNA molecule is observed adjacent to the TEC template entrance, labeled “Terminal” (Fig. 2B). Consistent with previous ARV structures, VP2 is a 1274-residue RNA-dependent RNA polymerase (RdRp) which, like CPV RdRp VP2, is organized into three major domains—N-terminal domain (1-387), C-terminal bracelet (902-1274), which sandwich the RdRp core (388-901), and are further differentiated into the thumb (793-901), fingers (388-557 and 595-690), and palm (558-594 and 691-792) subdomains (Fig. 2B-C) [16,17]. The fingers house the NTP entry channel and, with the thumb, assist in elongation and proofreading while the palm catalyzes phosphodiester bond formation between new NTPs and the growing strand via D591, D740, and D741, all three of which appear in beta sheets and are highly conserved in reovirus RdRps (Fig. 2C) The N- and C-terminal domains are versatile: In addition to shielding the RdRp core from any internal environmental changes, they bind dsRNA and other viral core proteins (Fig. 3B,H). Consistent with our previous findings, the C-terminal bracelet of this non-transcribing ICP does not occlude the template exit channel as it does in quiescent CPV (Fig. 2G) [16]. As previously observed, the polymerase possesses several channels to funnel RNA templates and transcription products (Movie S3) (Parker et al., 2002).

Our RdRp structure reveals several local architectural changes within these RdRp channels when compared to structures form the ARV ISVP (Fig 2D-E, and Movie S3). When viewed down the template entry channel we observe a positively charged finger domain loop (a.a.s 520-535) moves further into the template entry channel from the ISVP to ICP, widening the channel for template entry (Fig. 2E, Movie S3). From within the channel, we observe the priming loop (a.a. 560-572), thought to separate template and transcript strands, moves away from the transcript exit channel and orients the catalytically important S567 residue towards the would-be incoming template (Fig. 2 F, Movie S3). This migration away from the bracelet domain also widens the mRNA exit channel, likely promoting exit through the adjacent I5 pores and turrets. From the external TEC view, we also observe an RdRp expansion along the capsid lumen appears linked to the radial expansion of the capsid beneath it (Movie S3). This contrasts with the VP4 NTPase which undergoes a unidirectional shift, consistent with the radial expansion of the capsid monomers with which it associated (Movie S3). These movements may be linked to the asymmetric association of the TEC along the expanding decameric subunit. As the NTPase is situated primarily atop the $VP3_1$ dimer it moves along with the $VP3_1$ dimer elongation, whereas the RdRp, which is seated across $VP3_{1-4}$, is seemingly drawn in several directions based on the expansion of the capsid and the proportion of RdRp associated with each conformer pair (Fig. 3A). As uncoating is necessary to synthesize complete viral transcripts [27], the subtle conformational changes in the TECs observed here may be essential to carrying out efficient viral transcription.

3.3.4. CSP VP3 and NTPase VP4 N-termini interact with replicative machinery

In the asymmetrically reconstructed sub-particles, the previously unmodeled RNA-binding domain in TEC NTPase are resolved alongside the Zn-finger containing N-terminal domains of VP3 CSPs each of which exhibit interactions with the viral genome (Fig. 3). We will first describe the newly resolved N-terminal region of the ARV NTPase.

Viral TECs are situated beneath 11 of the 12 I5 vertices in ARV, containing one RdRp and accompanying NTPase molecule per TEC (Fig. 1E). We previously reported the structure of the ARV NTPase, revealing three distinct domains: the N-terminal RNA-binding domain (RBD, a.a. 1-285), the NTPase domain (a.a. 286-602), and the C-terminal domain (CTD a.a. 603-718) [16]. The NTPase RBD is composed of two subdomains, the 'tip' (a.a. 80-123) and 'main' (a.a. 1-79 and 124-285) which possess distinct electrostatic potentials (Fig. S3). In the previous structure, the *tip* domain was entirely unresolved along with a large portion of the *main* domain (a.a. 73-189) (Fig S3). In our structures we observe these previously missing NTPase segments extend away from the globular body of the TEC complex. This flexibility may help accommodate dynamic RNA interactions throughout transcription. The newly modeled N-terminal residues reveal a flexible region homologous to that of ARV's MRV cousin but of a distinct fold [18], exhibiting limited RdRp contacts but extensive genome interactions (Fig 2I-J).

Previously, the N-terminal residues of $VP3_A$ conformers were shown to associate with and lie along the exterior of the TEC, on both the RdRp and NTPase [16], and were suggested to anchor the TEC into place. The N-terminal region of ARV and MRV CSPs are also thought responsible for binding dsRNA during virus assembly [30]. The newly modeled $VP3_A$ N-terminal residues include the previously unresolved residues 108-152 ($VP3_{A1\text{-}4}$) which contain previously identified Zn-finger domains (a.a. 116-141) (Fig 3C-G), and conform to a traditional Cys_2His_2 motif, known to bind nucleotides (Fig 3E). Previous reporting suggests that the N-terminal Zn-fingers of MRV CSPs are unnecessary for dsRNA binding, and instead mediate stability of the capsid, despite demonstrated nucleotide binding affinity [30-32]. In keeping with this we observe the four newly resolved Zn-fingers ($VP3_{A1\text{-}4}$) lie along the TEC, with $VP3_{A2\text{-}4}$ situated along an RdRp cleft opposite the NTPase, and $VP3_{A1}$ seated along the VP4 NTPase domain forming a four-pronged setting that anchors the TEC complex to the capsid shell (Fig 3D). The $VP3_{A5}$ N-terminus also extends towards the RdRp, however we were unable to resolve the Zinc finger in either sub-particle reconstruction. We also observe that at both the tropical and polar vertices the $VP3_{A3}$ Zn-finger contacts the rear bottom genome segment (Fig 3E-F). This observation suggests these N-terminal Zn-fingers may help to regulate transcription initiation as previously observed in rotavirus [33]. It may be the case therefore that ARV VP3s are multifunctional, promoting TEC assembly and stability while maintaining genome organization in the quiescent particles. Despite the significant conformational changes of the capsid shell, the Zn-fingers appear in positions consistent with the low resolution ISVP structures, suggesting that the Zn-fingers may function independent of the capsid shell. This may be enabled by the flexible linker within the of the CSP N-terminal domains (a.a. 140-190), which extend slightly to accommodate the capsid expansion, while maintaining their TEC positioning and without altering the genome organization or activation state of the TEC (Movie S3).

3.3.5. Structure of the ICP turret protein VP1, a multi-enzyme complex

Atop each I5 vertex of ARV sits a pentameric turret composed of 5 copies of VP1 TPs (Fig 1A & 4A). These pentameric turrets form axial channels atop the pores through which the genome is translocated and provides mRNA capping functionality [34]. CPV and MRV virions both have spike proteins which occupy the turret channel and serve dual roles, mediating cell entry, and occluding the mRNA exit channel to prevent premature transcript escape (Figure 4A) [11,18]. ARVs lack a spike homologue therefore it is unclear how they might prevent premature transcript escape.

Like MRV, ARV TPs are composed of five domains: the GTase (a.a. 2-390), a bridge domain (a.a. 391-438, 695-806), MTase I (439-694) MTase II (807-1029), and Flap (1030-1298) which contains 3 immunoglobulin-like (IG) folds, two of which extend to cover the axial channel (Fig 4C). When ICP and virion turrets are compared we observe a significant axial extension of the ARV turrets and widening of the turret channels (Fig. 4B-C). Local alignment of the 5 TP domains reveals that the enzymatic GTase and MTase domains remain stable (RMSD 0.842 Å and 0.907 respectively) while domain separation occurs at the flexible linker regions (a.a. 391-438, 695-806) between the GTase, MTase, and flap domains. This separation is enabled by the loss of the OCP

VP5 that would otherwise clash with the new TP architecture (Fig 4A). Despite MRV possessing a similar double layered architecture, when the crystal structure of the MRV ICP is compared to the recently resolved asymmetric reconstruction of the complete virions the turret does not undergo analogous changes.

As ARV lacks a spike protein homologue, we investigated the flap domains that line the turret exit channel. When local alignment is performed it is clear there are significant conformational changes within the domain (Fig. 4 D). This difference is caused by the movement of the two C-terminal IG domains moving at the flexible linker region (a.a. 1129), which undergoes an outward shift of 14 Å and counterclockwise twist when viewed from within the I5 channel (Fig. 4D). This finding, coupled with the changes to capsid and TEC architecture, suggest a regulatory mechanism of transcription wherein the uncoating of ARV yields a particle whose capsid better accommodates genome transcription and whose capping enzymes are widened, or primed, for transcript export.

3.4. Conclusions

This work provides the first near-atomic resolution asymmetric reconstruction of an ICP from a turreted, multilayered reovirus particle which exhibits subtle but functionally important conformational changes compared to coated ARV structures. These changes in the internal capsid volume, along-side a widening of TEC nucleotide channels and extension of the 5'-cap synthesizing turret proteins create the architectural environment conducive to endogenous transcription. As the ICP genome remains in the quiescent state, it is conceivable that transcription initiation requires factors such as S-adenosyl methionine and NTPs in addition the conformational changes observed here. In this manner, the highly specialized outer capsid layers serve as a useful transcriptional regulator, ensuring that transcription requires not only the presence of cofactors, but also loss of outer capsid layers which occurs upon cell entry.

3.5. Acknowledgements

This work was supported in part by grants from the National Institutes of Health (AI094386 to Z.H.Z.). A.S. received support from NIH Ruth L. Kirschtein National Research Service Award AI007323. We acknowledge the use of resource at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057, 1S10OD018111 and U24GM116792) and NSF (DBI-1338135 and DMR-1548924) and at the UCLA AIDS Institute supported in part by UCLA-CDU CFAR (NIH AI152501), the James B. Pendleton Charitable Trust and the McCarthy Family Foundation.

3.5.1. Author contributions

Z.H.Z. designed and supervised the project; S.S. prepared sample and made cryoEM grids; Y.C. performed cryoEM imaging and 3D reconstruction; A.S. built the atomic models, interpreted the structures, made the figures, and wrote the paper with input from Z.H.Z.; All authors reviewed and approved the paper.

Conflict of interest. The authors declare no competing interests.

3.6. Methods and Data Availability

3.6.1. Viral Culture and Isolation

Golden shiner reovirus stock was purchased from ATCC (ATCC® VR957™) and was propagated in Fat Head Minnow cells (FHM; ATCC® CCL-42™) with a slight modification to previously described method [15]. FHM cells were cultured in MEM, hanks' balanced salts (Gibco™) supplemented with 10% fetal bovine serum (Atlanta biologicals), Penicillin (10,000 IU) and Streptomycin (10,000 µg/ml) (Corning ®, USA), and maintained at 28°C. Confluent monolayers of FHM were infected with ~5 plaque forming units of GSRV per cell for ~48 hours. Once cell syncytia were visible, media was collected from the infected cell culture and centrifuged at 11,000 g for 15 min to get rid of cells and cellular debris. Clean supernatant was centrifuged at 100,000g for 3 hours (SW28 rotors, Beckman Coulter) to pellet the viral particles. Supernatant was discarded and 100 µl of ice-chilled PBS buffer was added to each pellet to resuspend the viral particles overnight. Pellets were pooled and loaded on a 15%-50% sucrose gradient and banded by centrifuging at 100,000 g for 1 hour (SW41 rotor) at 4°C. Band containing the GSRV cores was harvested and diluted in PBS, pelleted at 100,000 g for 2 hours. The cores in the pellet were resuspended in PBS and quality confirmed by negative-staining (2% uranyl acetate) stain transmission electron microscopy.

3.6.2. CryoEM imaging

For cryoEM sample preparation, 2.5-µL aliquots of the GSRV core sample were applied to 200-mesh Quantifoil grid with 2µm-diameter hole size (2/2 or 2/1 Quantifoil) and plunge frozen in a Vitrobot Mark IV (Thermo Fisher Scientific) (Thermo Fisher Scientific) after blotting (blot time 15 seconds, blot force -5, 100% humidity and 4°C). Movies of dose fractionated image frames were recorded using a Titan Krios microscope (Thermo Fisher Scientific) equipped with a Gatan imaging filter (GIF) quantum LS and a post-GIF Gatan K2 Summit direct electron detector operated in super resolution mode with a calibrated pixel size of 0.68 Å/pixel with SerialEM [35]. The GIF slit width was set to 20 eV at the zero-loss peak, and dose rate as recorded by the camera was set to with a total exposure dosage of ~45 e/Å^2 at the sample level which was fractionated equally into 40 frames. A defocus range of 1.5µm-2.5 µm was targeted for recordings. Dose-fractionated frames were binned 2× (effective pixel size 1.36 Å) and aligned for beam induced motion correction to produce dose-weighted and unweighted micrographs, used for final reconstruction and initial screening and CTF estimation respectively, using MotionCor2 [36]. In total 12,258 movies were recorded.

3.6.3. Data processing

Defocus values of the micrographs were initially estimated using CTFFIN4 [37], and those micrographs with significant ice contamination or defocus values beyond -0.8 and -3.0 µm were removed leaving 10,935 micrographs. 38,486 particles were initially selected by Ethan [38]. The particles were extracted and initially subjected to an icosahedral refinement in RELION which yielded a 3.4 Å cryoEM map of the complete viral particle (Scheres, 2012). In our icosahedral reconstruction, strong densities were observable beneath the positions of I5 vertices. We then followed a previously established stepwise symmetry relaxation workflow within RELION to generate asymmetric sub-particle reconstructions of both polar and tropical vertices [16,23,39]. Briefly, particles from the 3.2 Å icosahedral reconstruction were subjected to symmetry expansion (option I3) using the *relion_symmetry_expand* command, to generate a RELION STAR file containing 60 orientation entries for each ARV particle, corresponding to 5 copies of each I5 vertex differentiated by the rotation about the I5 vertex, and listed as _rlnAngleRot. One entry corresponding to each vertex was kept, yielding a duplicate-entries-removed STAR file with ~12 entries, or one for each vertex, for each ICP. The icosahedral reconstruction orients in such a way that the location of one vertex consistently lies along the *z*-axis, so we performed classification on ICP particles with the capsid density subtracted to identify tropical and polar vertices and estimated their location for extraction of the penton sub-particles using the RELION “Particle extraction” tool [40]. Polar and Tropical vertices were separately refined into the final sub-particle maps provided here. Unfortunately, this method failed to resolve the true asymmetric structure featuring 11 TECs, as the pseudo-D3 symmetrical particle featured 12 TECs, hindering efforts to resolve the unoccupied vertex. This was overcome following the D3-symmetry expansion of the ICP particles, generating 6 duplicates within a new STAR file. The particles were then subjected to exhaustive classification with a reference mask beneath a tropical vertex to identify a class without a TEC. Using the orientations determined during this classification we were able to reconstruct the complete asymmetric ICP.

3.6.4. Atomic Model Building and Refinement

Atomic models of the GSRV RdRp, CSP, NTPase, turret and clamp proteins were initially modeled into the cryoEM density in Coot using the GCRV ISVP (PDB 6M99) as a homology following previously established protocols [16,41,42]. Following initial fitting, the polypeptide chains were mutated to match the GSRV sequence using Coot’s Mutate Residue Range tool prior to local and global refinement in PHENIX [43]. Residues were built into previously unresolved residues, de novo using Coot, except for the first 200 residues of VP4, whose structure was predicted in AlphaFold2 prior to fitting into the cryoEM map [44]. Upon initial fitting the model was refined within the sub-particle map using the molecular dynamic flexible fitting (MDFF) feature in ISOLDE from the UCSF

ChimeraX user interface [25,45]. Models were then subjected to a final round of PHENIX real space refine and the resulting structure was validated by the worldwide protein data bank validation server [43,46]. Root mean squared deviations (RMSDs) between protein chains were calculated using the matchmaker tool in UCSF ChimeraX. Internal capsid volume was calculated using the ChimeraX recipe “Measure volume enclosed by a virus capsid” on our ARV ICP and the previously published virion structure (EMD-6969) [17]. TEC distribution maps were generated using UCSF Chimera [47].

Modeling of dsRNA.

We inserted poly-AU, A-form dsRNA segments from a previously published MRV model (PDB 7ELH) into our cryoEM maps [18]. Segments were roughly placed into the cryoEM map, and segments were further fit in ISOLDE [45]. Briefly, adaptive distance restraints were applied to the nucleic acid segments with reduced weight applied to the non-crystallographic map were fit into the cryoEM maps using ISOLDEs MDFF feature [45].

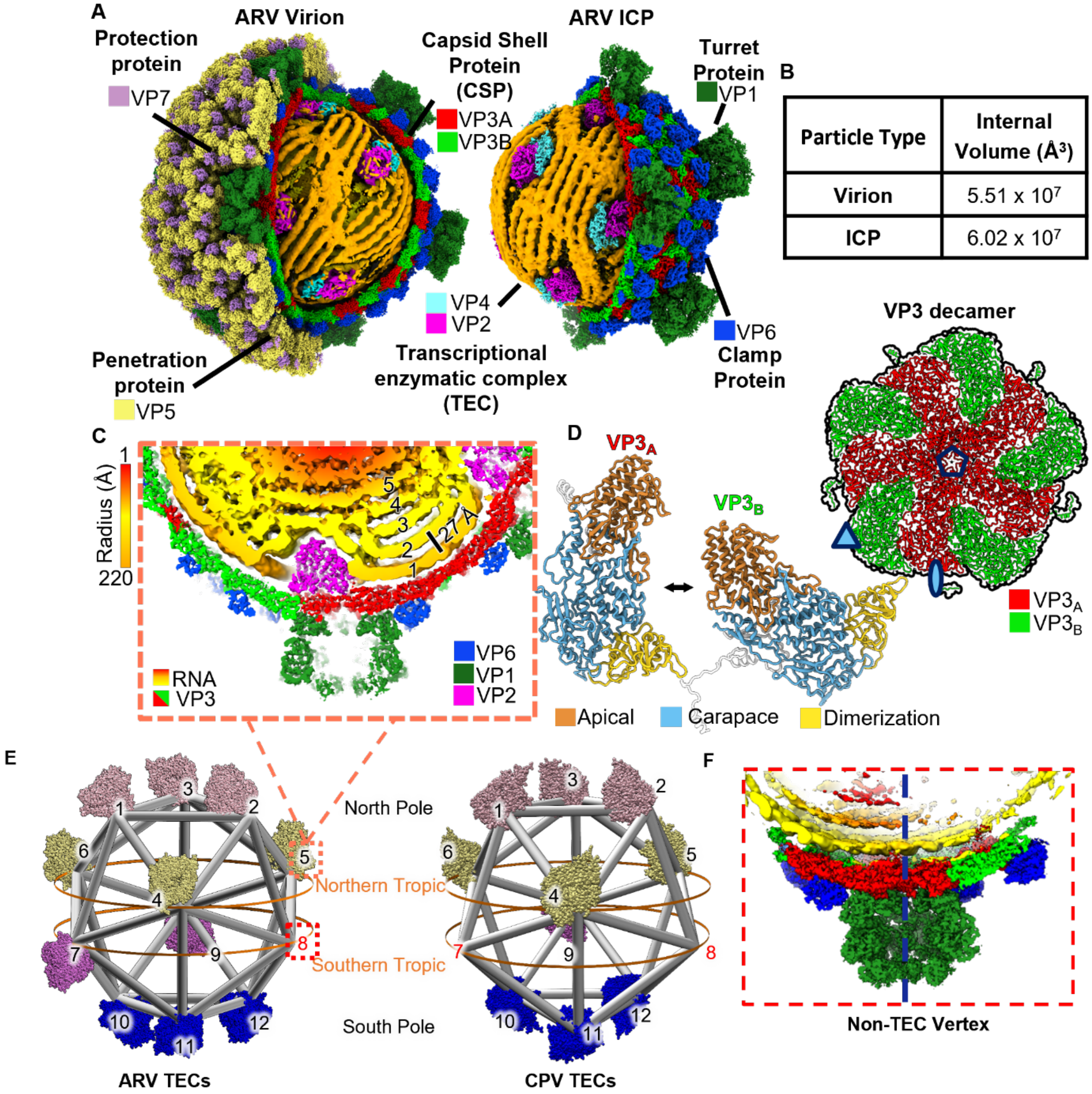
3.6.5. Data Availability

CryoEM maps of the asymmetric ARV ICP and sub-particles have been submitted to the Electron Microscopy Data Bank and can be found under accession numbers EMD-29337, EMD-29243, EMD-29244, and EMD-29336 for the cryoEM structures of the complete capsid, polar vertex, tropical vertices with and without TECs respectively. The coordinates of ICP polar and tropical vertex sub-particles were deposited in the Protein Data Bank under accession number 8FJK and 8FJL.

Table 2- S1: Cryo-EM data collection, refinement, and validation statistics

	ICP Asymmetric reconstruction (EMD-29337)	Polar vertex with TEC (EMD-29243) (PDB 8FJK)	Tropical Vertex with TEC (EMD- 29244) (PDB 8FJL)	Tropical Vertex without TEC (EMD- 29336)
Data collection and processing				
Magnification	81,000	81,000	81,000	81,000
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	45	45	45	45
Defocus range (μm)	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5
Pixel size (Å)	1.36	1.36	1.36	1.36
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	36,764	36,764	36,764	36,764
Final particle images (no.)	15,433	119,182	99,323	21,844
Map resolution (Å)	4.2	3.3	3.3	3.7
FSC threshold	0.143	0.143	0.143	0.143
Refinement				
Initial model used (PDB code)		PDB 6M99	PDB 6M99	
Model resolution (Å)		3.0	2.9	
FSC threshold		0.143	0.143	
Model resolution range (Å)				
Map sharpening *B* factor (Å^2)		-100	-100	
Model composition				
Non-hydrogen atoms		183,951	183,948	
Protein residues		23,916	23916	
RNA Nucleotides				
Ligands		4	4	
R.m.s. deviations				
Bond lengths (Å)		0.006	0.006	
Bond angles (°)		1.000	1.000	
Validation				
MolProbity score		1.39	1.39	
Clashscore		3.67	3.67	
Poor rotamers (%)		0.53	0.53	
Ramachandran plot				
Favored (%)		96.47	96.47	
Allowed (%)		3.52	3.52	
Disallowed (%)		0.00	0.00	

3.7. Figures



Particle Type	Internal Volume (Å³)
Virion	5.51×10^7
ICP	6.02×10^7

Figure 3- 1. Asymmetric reconstruction of ARV ICP.

(A) CryoEM structures of ARV virion (left) (5VST) and ICP (right) with portion of their external layers removed along hemispheres to expose genome and TECs. (B) Table comparing volumetric difference between virion and ICP lumen. (C) Cross sectional view of Tropical Vertex featuring TEC and a turret. Layers of genome segments are numbered from 1-5 based on proximity to capsid wall. (D) Ribbon Diagram of ICP capsid decamer (top right) and VP3$_A$ (left) and VP3$_B$ (right) conformers with domains indicated and colored differently. (E) Position and orientation of the 11 ICP TECs relative to an icosahedron (grey) with Northern and Southern “poles” and “tropics” indicated. For clarity, RdRp distribution model not drawn to scale. (F) Cross sectional view of ICP vertex without TEC showing that the TEC-free vertex is occupied by RNA density (yellow).

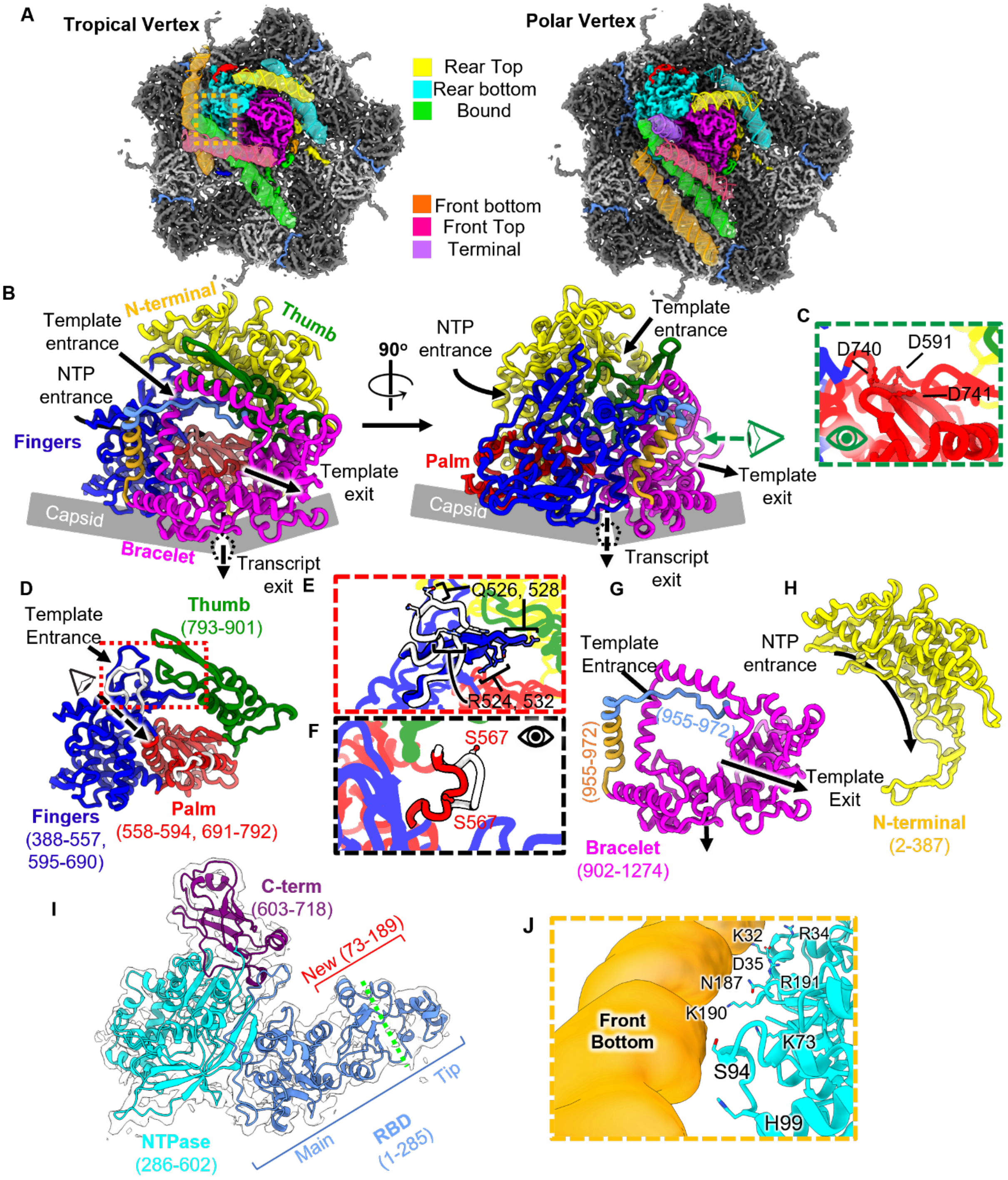


Figure 3- 2. CryoEM maps show distinct RNA arrangement around different ICP vertices with identical TEC structures.

(A) Comparison of the CryoEM density maps from polar and tropical vertices, highlighting the TEC (magenta and cyan) and surrounding RNA, shown semitransparently and superposed with corresponding dsRNA models (colored by segment). (B-H) Atomic model of ICP RdRp VP2 colored by domain, either shown together (B), or separately (D-H) for detailed features. (C) provides magnified view of catalytic palm domain from B, with catalytically important residues indicated. In E and F views from D are shown to demonstrate the difference between the ICP and ISVP of the template channel finger loop (E) and catalytically important priming loop (F). (I) Atomic model of RdRp cofactor VP4 colored by domains with newly resolved residues indicated (red) and division of RBD into Main and Tip subdomains (green separator). (J) Interactions of the polar and charged residues from the newly resolved NTPase residues with the adjacent RNA segment from boxed region in A (orange).

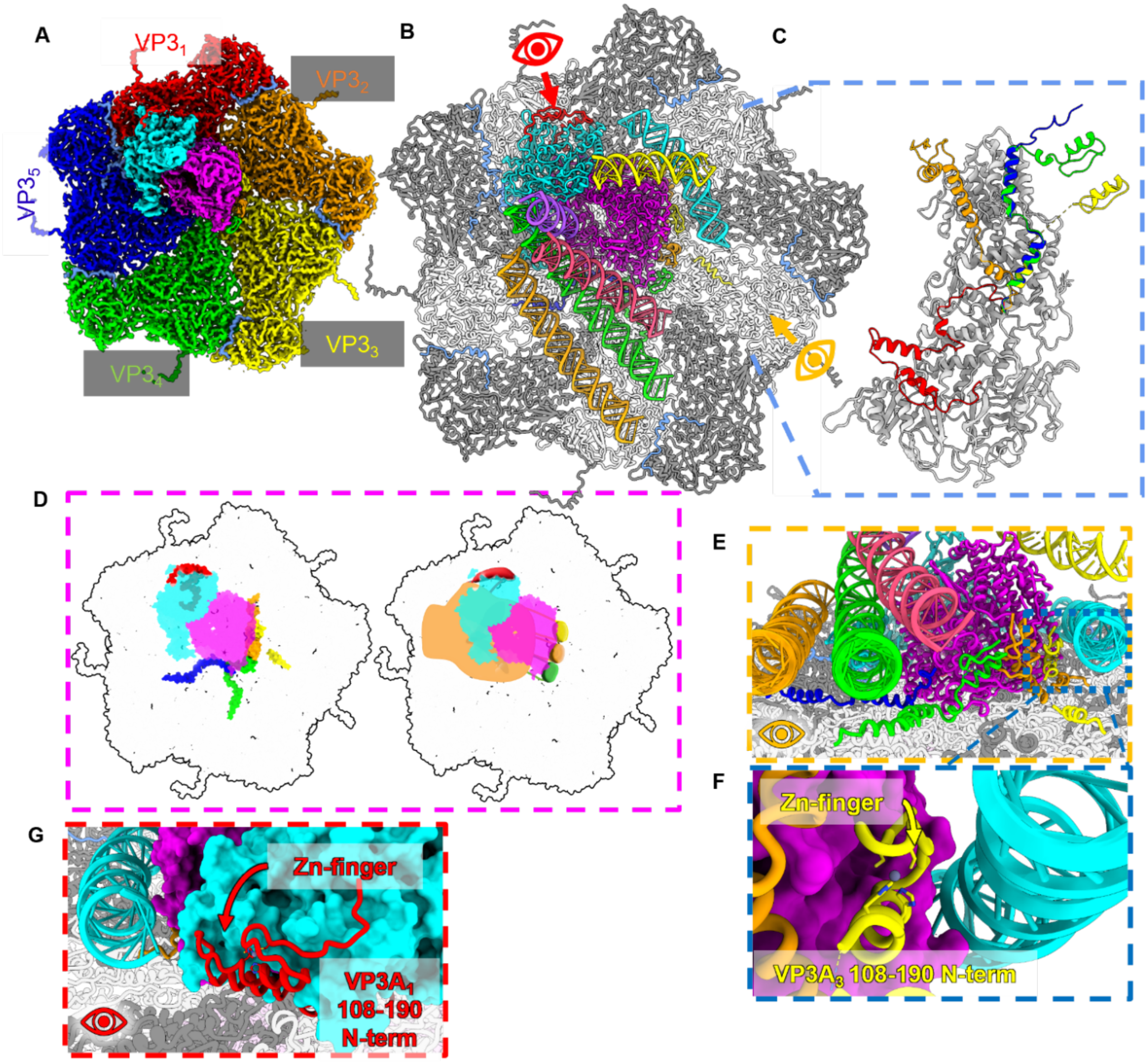


Figure 3- 3. Asymmetric interactions of $VP3_A$ N-termini with TEC and nucleic acids. (A and B) CryoEM map of polar decamer and attached TEC (left) with dimer pairs colored based on TEC association (A), and corresponding atomic model (B). Note the RNA is only shown in the atomic model (B) and not shown in the cryoEM density for clarity. (C) Superposition of $VP3_A$ proteins from each dimer pair, with N-terminal residues colored to match dimer pairs from A. (D) Depiction of TEC along decamer lumen with N-terminal surfaces of $VP3_A$ CSPs colored according to dimer pair in A (left),

and hand shape added to improve visualization of Zn-fingers grasping TEC. (E-G) Side views of atomic model as indicated by eye symbols in B, with $VP3_A$ N-termini colored as in C (E and G), and magnified view of $VP3_{A3}$ Zn-finger domain interacting with both RdRp (magenta) and rear bottom dsRNA (cyan).

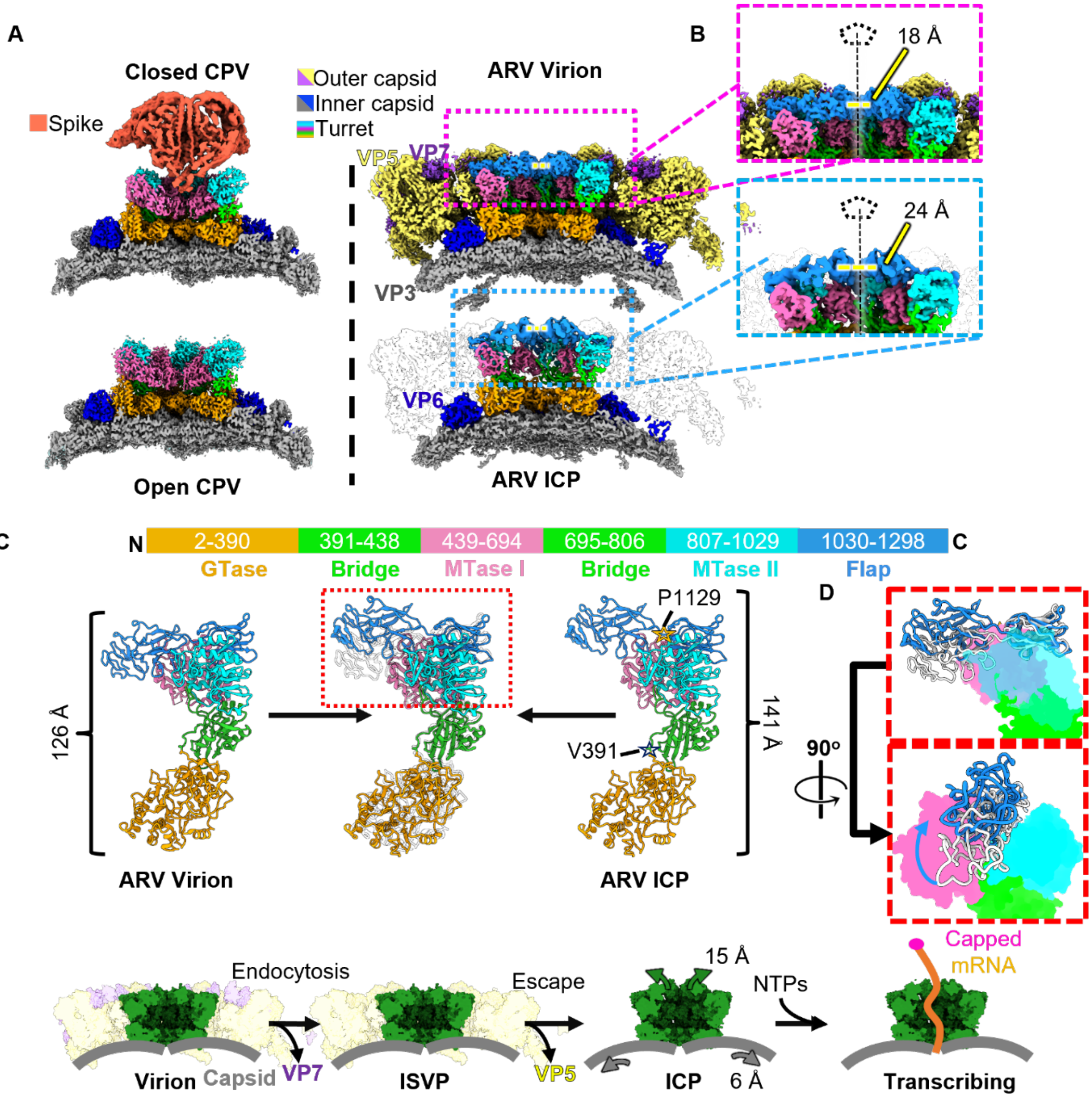
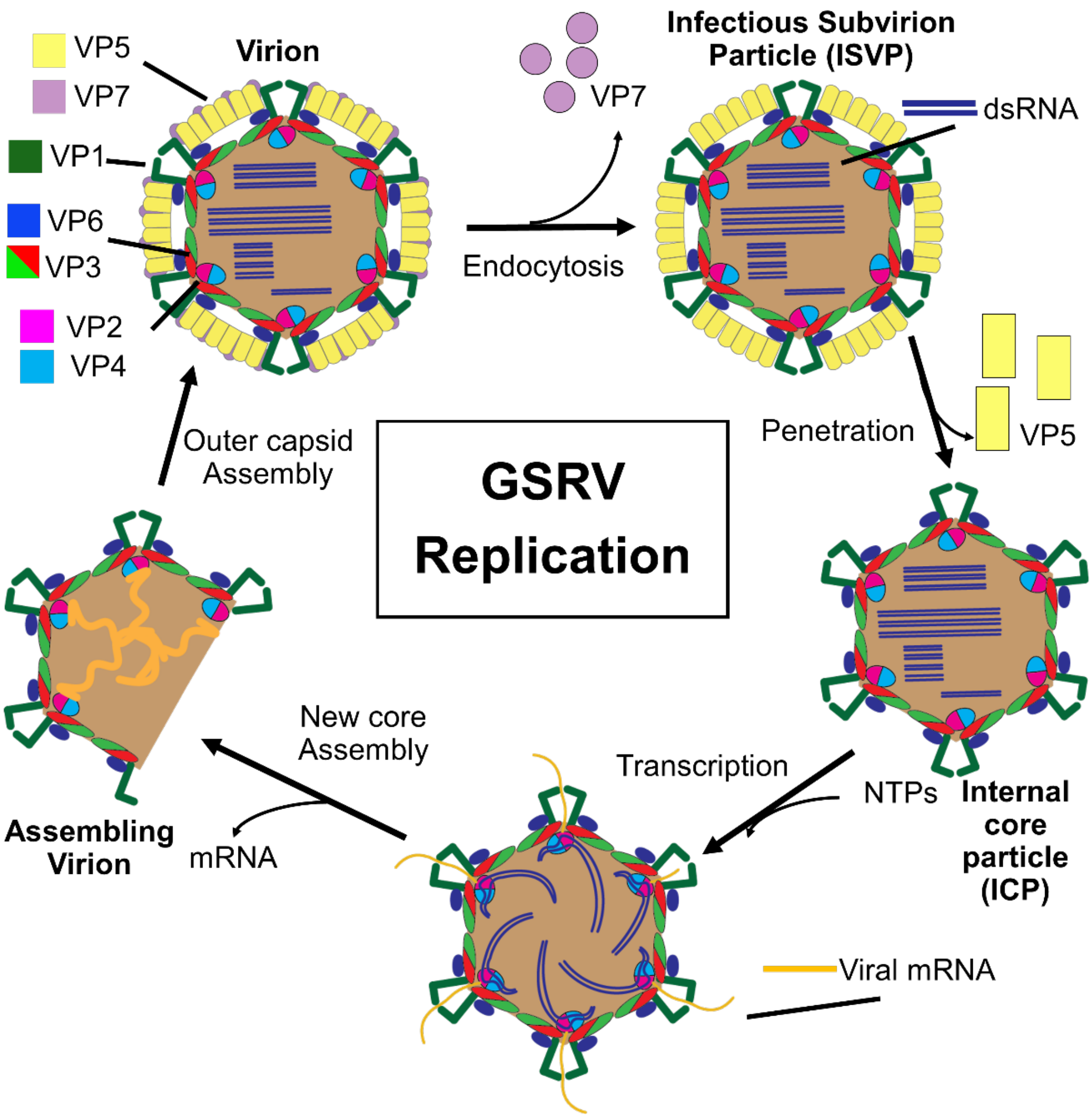
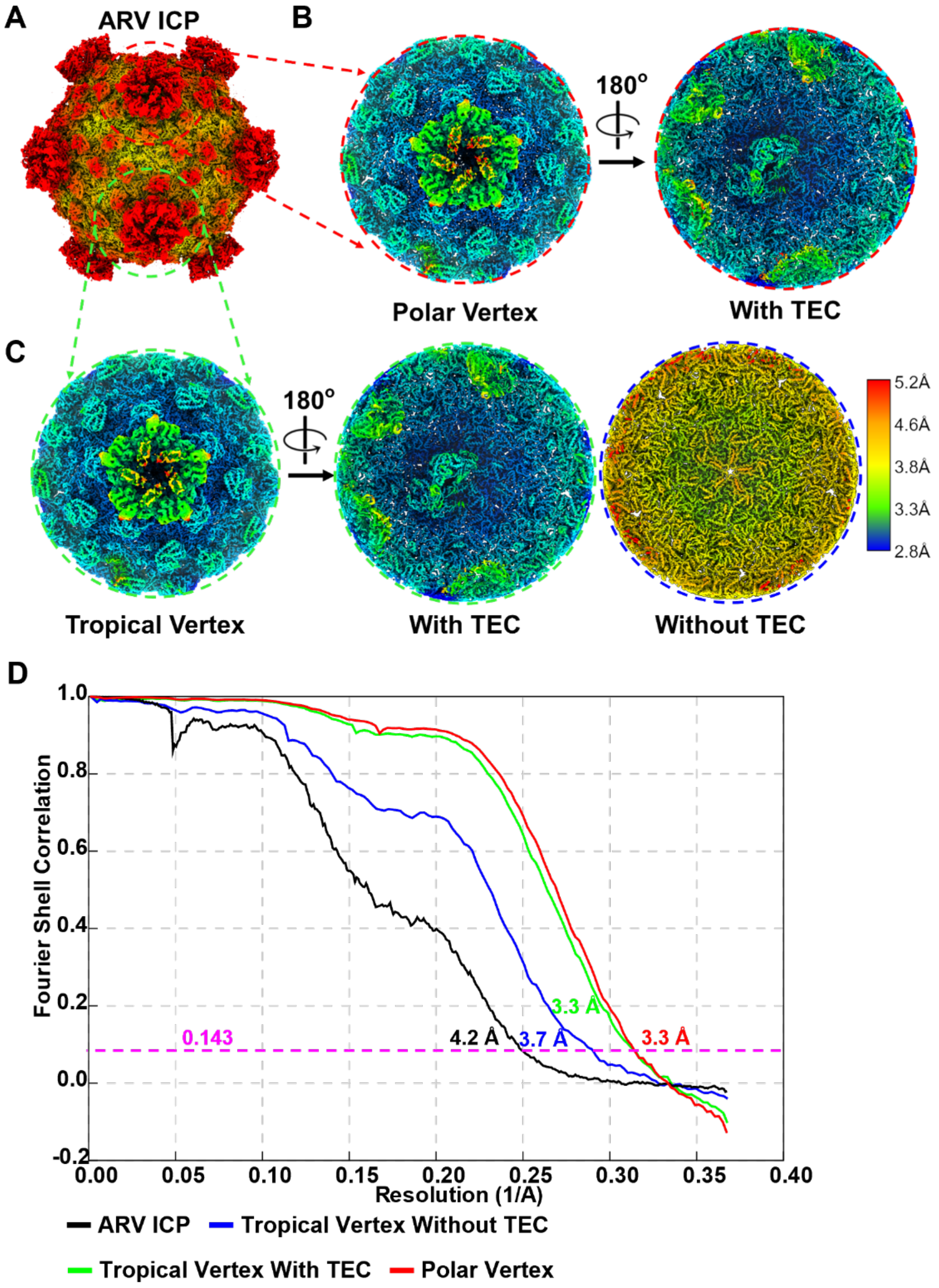


Figure 3- 4. The novel conformational changes of the capping enzyme turret in uncoated virus.

(A) comparison of CPV before and after detachment, where the Receptor binding spike remains attached to the turret and corresponding region of ARV virion and ICP. (B) Magnified view of turret pores from ARV virion (top) and ICP (bottom) with I5 symmetrical axis indicated with pentagon and line segment. (C) Atomic structure of the ARV turret protein VP1 from virion (left) and ICP (right) with superimposition of MTase domains (center) to show difference, with domains depicted (top) with the primary sequence number and corresponding color scheme. (D) magnified view of boxed region from C, with solid color of non-flap domains used to highlight virion to ICP differences. (E) Schematic demonstrating the confirmational changes differences undergone by ARV turrets (green) from virion to ISVP and ICP, which coincide with the loss of outer capsid proteins VP5 and VP7 (ghost yellow and purple).

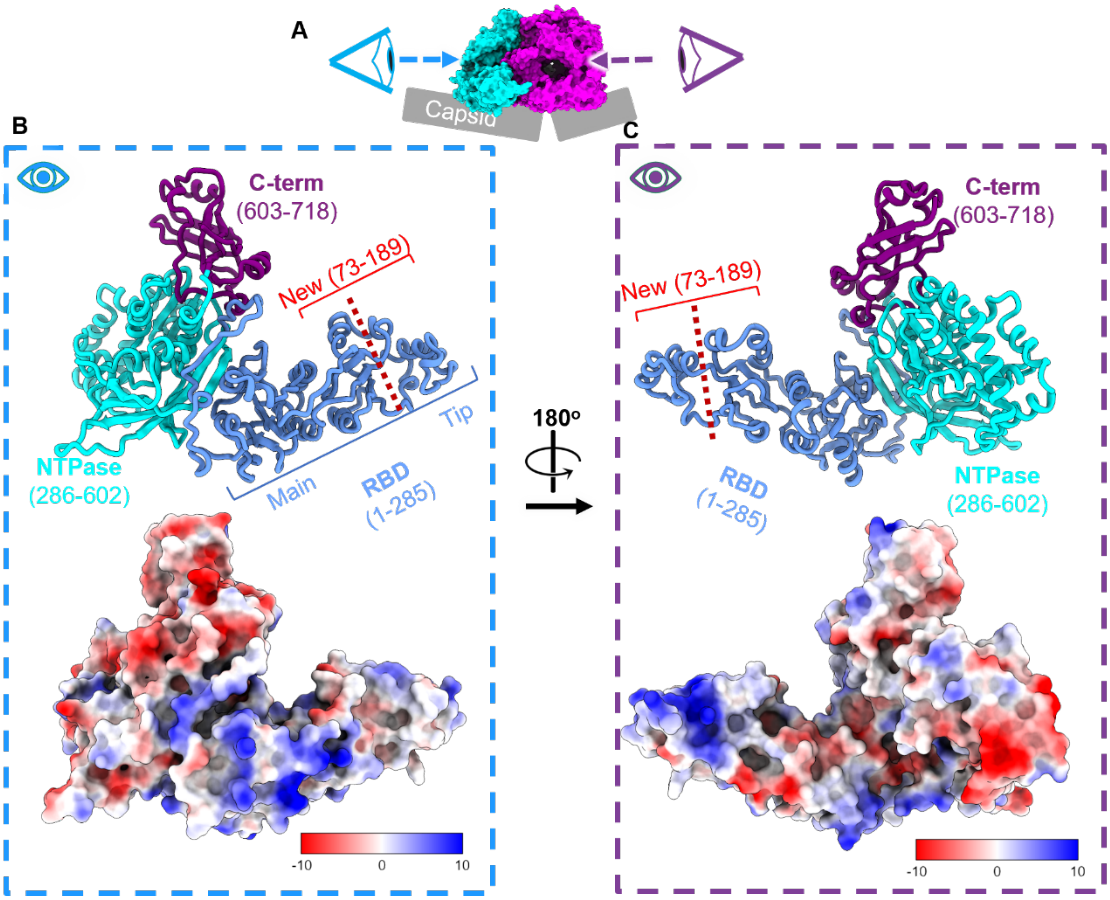


Supplementary Figure 3- S1. Typical changes to Aquareovirus capsid during replication. Diagrammatic representation of an Aquareovirus in cross section and the proposed replication process. The identities of viral structural and enzymatic proteins are indicated using the nomenclature for *Aquareoviruses* and assigned colors. Double-stranded DNA is represented by parallel blue line segments and single-stranded RNA is represented by orange lines.



Supplementary Figure 3- S2. Asymmetric reconstructions from ARV ICP.
(A-C) CryoEM maps of GSRV ICP (A) and sub particles from Polar (B) and Tropical (C) vertices colored by local resolution. Internal views of Polar and Tropical vertices in B and C (right and middle panels respectively) showing the asymmetrically associated

TEC in both species and the lack thereof in a subset of Tropical vertices (C, right). (D) Gold-standard Fourier shell correlation (FSC) curves calculated from the independently refined half-maps of each cryoEM reconstruction in A-C, with resolution values listed for each species at FSC = 0.143.



Supplementary Figure 3- S3. Columbic potential maps of NTPase VP4.

(A-C) different views of NTPase VP4 protein depicted by ribbon diagrams and colored by domains (B-C, top). (B) View of NTPase as viewed along the face opposite the associated RdRp with surface colored by columbic potential (bottom). (C) View of the RdRp associated face of NTPase depicted in the same manner as B.

3.8. References

1. Troeger, C. et al. Rotavirus Vaccination and the Global Burden of Rotavirus Diarrhea Among Children Younger Than 5 Years. in *JAMA pediatrics* Vol. 172 958-965 (2018).
2. King, A.M.Q., Adams, M.J., Carstens, E.B. & Lefkowitz, E.J. Family - Reoviridae. in *Virus Taxonomy* (eds King, A.M.Q., Adams, M.J., Carstens, E.B. & Lefkowitz, E.J.) 541-637 (Elsevier, 2012).
3. Skehel, J.J. & Joklik, W.K. Studies on the in vitro transcription of reovirus RNA catalyzed by reovirus cores. in *Virology* Vol. 39 822-831 (1969).
4. Cui, Y., Zhang, Y., Zhou, K., Sun, J. & Zhou, Z.H. Conservative transcription in three steps visualized in a double-stranded RNA virus. in *Nature Structural & Molecular Biology* Vol. 26 1023-1034 (Springer US, 2019).
5. Ebert, D.H., Deussing, J., Peters, C. & Dermody, T.S. Cathepsin L and Cathepsin B Mediate Reovirus Disassembly in Murine Fibroblast Cells. in *Journal of Biological Chemistry* Vol. 277 24609-24617 (2002 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 2002).
6. Zhang, H. et al. Visualization of Protein-RNA Interactions in Cytoplasmic Polyhedrosis Virus. in *Journal of Virology* Vol. 73 1624-1629 (1999).
7. Hill, C.L. et al. The structure of a cypovirus and the functional organization of dsRNA viruses. in *Nature Structural Biology* Vol. 6 565-568 (1999).

8. Zhou, Z.H., Zhang, H., Jakana, J., Lu, X.-Y. & Zhang, J.-Q. Cytoplasmic Polyhedrosis Virus Structure at 8 A by Electron Cryomicroscopy: Structural Basis of Capsid Stability and mRNA Processing Regulation. in *Structure* Vol. 11 651-663 (2003).

9. Chen, F. et al. Clathrin-mediated endocytosis is a candidate entry sorting mechanism for Bombyx mori cypovirus. in *Scientific Reports* Vol. 8 7268 (2018).

10. Yu, X., Jiang, J., Sun, J. & Zhou, Z.H. A putative ATPase mediates RNA transcription and capping in a dsRNA virus. in *eLife* Vol. 4 e07901 (2015).

11. Zhang, Y., Cui, Y., Sun, J. & Zhou, Z.H. Multiple conformations of trimeric spikes visualized on a non-enveloped virus. in *Nature Communications* Vol. 13 550 (Springer US, 2022).

12. Yang, C. et al. Cryo-EM structure of a transcribing cypovirus. in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 109 6118-6123 (2012).

13. Fang, Q., Shah, S., Liang, Y. & Zhou, Z.H. 3D reconstruction and capsid protein characterization of grass carp reovirus. in *Science in China, Series C: Life Sciences* Vol. 48 593-600 (2005).

14. Nason, E.L., Samal, S.K. & Venkataram Prasad, B.V. Trypsin-Induced Structural Transformation in Aquareovirus. in *Journal of Virology* Vol. 74 6546-6555 (2000).

15. Shaw, A.L., Samal, S.K., Subramanian, K. & Prasad, B.V. The structure of aquareovirus shows how the different geometries of the two layers of the capsid are reconciled to provide symmetrical interactions and stabilization. in *Structure* Vol. 4 957-967 (1996).

16. Ding, K., Nguyen, L. & Zhou, Z.H. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. in *Journal of Virology* Vol. 92 (ed. López, S.) e00774-18 (2018).

17. Wang, X. et al. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. in *Proceedings of the National Academy of Sciences of the United States of America* Vol. 115 7344-7349 (2018).

18. Pan, M. et al. Asymmetric reconstruction of mammalian reovirus reveals interactions among RNA, transcriptional factor µ2 and capsid proteins. in *Nature Communications* Vol. 12 1-16 (Springer US, 2021).

19. Dryden, K. et al. Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. in *Journal of Cell Biology* Vol. 122 1023-1041 (1993).

20. Zhang, F., Sun, D. & Fang, Q. Molecular Characterization of Outer Capsid Proteins VP5 and VP7 of Grass Carp Reovirus. in *Viruses* Vol. 14 1032 (2022).

21. Kim, J., Tao, Y., Reinisch, K.M., Harrison, S.C. & Nibert, M.L. Orthoreovirus and Aquareovirus core proteins: conserved enzymatic surfaces, but not protein–protein interfaces. in *Virus Research* Vol. 101 15-28 (2004).

22. Zhang, X., Jin, L., Fang, Q., Hui, W.H. & Zhou, Z.H. 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. in *Cell* Vol. 141 472-482 (2010).

23. Zhang, X. et al. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. in *Nature* Vol. 527 531-534 (Nature Publishing Group, 2015).

24. McEntire, M.E., Iwanowicz, L.R. & Goodwin, A.E. Molecular, physical, and clinical evidence that golden shiner virus and grass carp reovirus are variants of the same virus. in *Journal of Aquatic Animal Health* Vol. 15 257-263 (2003).

25. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. in *Protein Science* Vol. 27 14-25 (2018).

26. Demidenko, A.A. & Nibert, M.L. Probing the transcription mechanisms of reovirus cores with molecules that alter RNA duplex stability. in *Journal of virology* Vol. 83 5659-70 (2009).

27. Farsetta, D.L., Chandran, K. & Nibert, M.L. Transcriptional Activities of Reovirus RNA Polymerase in Recoated Cores. in *Journal of Biological Chemistry* Vol. 275 39693-39701 (Â© 2000 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 2000).

28. Trask, S.D., McDonald, S.M. & Patton, J.T. Structural insights into the coupling of virion assembly and rotavirus replication. in *Nature Reviews Microbiology* Vol. 10 165-177 (2012).

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

30. Lemay, G. & Danis, C. Reovirus 1 Protein: Affinity for Double-stranded Nucleic Acids by a Small Amino-terminal Region of the Protein Independent From the Zinc Finger Motif. in *Journal of General Virology* Vol. 75 3261-3266 (1994).

31. Yu, X., Ge, P., Jiang, J., Atanasov, I. & Zhou, Z.H. Atomic Model of CPV Reveals the Mechanism Used by This Single-Shelled Virus to Economically Carry Out Functions Conserved in Multishelled Reoviruses. in *Structure* Vol. 19 652-661 (Elsevier Ltd, 2011).

32. McDonald, S.M. & Patton, J.T. Rotavirus VP2 Core Shell Regions Critical for Viral Polymerase Activation. in *Journal of Virology* Vol. 85 3095-3105 (2011).

33. Ding, K. et al. In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. in *Nature Communications* Vol. 10 2216 (Springer US, 2019).

34. Reinisch, K.M., Nibert, M.L. & Harrison, S.C. Structure of the reovirus core at 3.6 Å resolution. in *Nature* Vol. 404 960-967 (2000).

35. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. in *Journal of Structural Biology* Vol. 152 36-51 (2005).

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

37. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. in *Journal of Structural Biology* Vol. 192 216-221 (Elsevier Inc., 2015).

38. Kivioja, T., Ravantti, J., Verkhovsky, A., Ukkonen, E. & Bamford, D. Local average intensity-based method for identifying spherical particles in electron micrographs. in *Journal of structural biology* Vol. 131 126-34 (2000).

39. Scheres, S.H.W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. in *Journal of Structural Biology* Vol. 180 519-530 (Elsevier Inc., 2012).

40. Scheres, S.H.W. Single-particle processing in RELION. in *Manuals* 1-21 (2013).

41. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. in *Acta Crystallographica Section D Biological Crystallography* Vol. 66 486-501 (International Union of Crystallography, 2010).

42. Yu, I. et al. Building atomic models based on near atomic resolution cryoEM maps with existing tools. in *Journal of Structural Biology* Vol. 204 313-318 (Elsevier, 2018).

43. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. in *Acta Crystallographica Section D: Structural Biology* Vol. 74 531-544 (International Union of Crystallography, 2018).

44. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. in *Nature* Vol. 596 583-589 (Springer US, 2021).

45. Croll, T.I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. in *Acta Crystallographica Section D: Structural Biology* Vol. 74 519-530 (International Union of Crystallography, 2018).

46. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. in *Nature Structural & Molecular Biology* Vol. 10 980-980 (2003).

47. Pettersen, E.F. et al. UCSF Chimera-A visualization system for exploratory research and analysis. in *Journal of Computational Chemistry* Vol. 25 1605-1612 (2004).

4. Chapter 4: Structure-Guided Mutagenesis Targeting Interactions between pp150 Tegument Protein and Small Capsid Protein Identify Five Lethal and Two Live-Attenuated HCMV Mutants

Alexander Stevens[a,b,c], Ruth Cruz-Cosme[d], Najealicka Armstrong[d], Qiyi Tang[d], Z. Hong Zhou[a,b,c,*]

[a] California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA

[b] Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

[c] Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

Department of Microbiology, Howard University College of Medicine, Washington, DC 20059, USA

*Corresponding author.

4.1. Abstract

Human cytomegalovirus (HCMV) replication relies on a nucleocapsid coat of the 150kDa, subfamily-specific tegument phosphoprotein (pp150) to regulate cytoplasmic virion maturation. While recent structural studies revealed pp150-capsid interactions, the role of specific amino-acids involved in these interactions have not been established experimentally. In this study, pp150 and the small capsid protein (SCP), one of pp150's binding partners found atop the major capsid protein (MCP), were subjected to mutational and structural analyses. Mutations to clusters of polar or hydrophobic residues along the pp150-SCP interface abolished viral replication, with no replication detected in mutant virus-infected cells. Notably, a single amino acid mutation (pp150

K255E) at the pp150-MCP interface significantly attenuated viral replication, unlike in pp150-deletion mutants where capsids degraded outside host nuclei. These functionally significant mutations targeting pp150-capsid interactions, particularly the pp150 K255E replication-attenuated mutant, can be explored to overcome the historical challenges of developing effective antivirals and vaccines against HCMV infection.

4.2. Introduction

Human cytomegalovirus (HCMV) is a member of the β-herpesvirus subfamily of the *Herpesviridae,* which establishes lifelong, latent infection with periodic reactivation and exhibits a worldwide seroprevalence near 83% [1]. Though innocuous in healthy adults, HCMV poses a serious threat to the immunocompromised, such as solid organ transplant recipients, and immune naïve. HCMV is the leading viral cause of birth defects world-wide, infecting 1 in every 200 newborns, and frequently causing neurological disabilities such as blindness, microcephaly, and deafness [2,3]. Indeed, HCMV is implicated in nearly a quarter of all cases of hearing loss in the United States [2,3]. Despite considerable vaccine development efforts [4-7], there are no HCMV prophylactics available. Additionally, though anti-HCMV drugs exist, their nephrotoxicity and the emergence of resistant HCMV strains necessitates the development of new therapies [8,9].

Live-attenuated viruses represent attractive herpesvirus vaccine candidates as they can induce the same complex cellular and humoral immune responses as wild-type herpesviruses that lead to long-term immunity as exhibited in varicella zoster virus [10,11]. Previous live-attenuated HCMV vaccine candidates have been derived from laboratory strains and lacked multiple genes compared to WT [7,12,13]; and have so far shown limited efficacy in preventing infection to date [4-6]. The deleted genes often include major immunomodulators like viral protein pUL144, but which specific deletion impacts vaccine efficacy is unclear [12-14]. More recently, structural biologists elucidated structures of the HCMV virion [15-17], revealing near-atomic level details and enabling rational design of live-attenuated mutants.

As a herpesvirus, HCMV shares general morphological characteristics of viruses from the α-herpesvirus (e.g., herpes simplex virus or HSV) and γ-herpesvirus (e.g., Kaposi's sarcoma-associated herpesvirus or KSHV) subfamilies [18]. The virion has four layers: the large double-stranded DNA genome is packed tightly into the icosahedral capsid, which is coated in amorphous tegument proteins and enclosed within a lipid envelope. The HCMV capsid is composed of 955 copies of the major capsid protein (MCP) arranged into 150 hexons with eleven pentons at the icosahedral 5-fold (I5) vertices. The 12th I5 vertex is occupied by a unique portal vertex, through which the viral genome is translocated. 360 triplex protein complexes (Tri), each composed of one Tri1 and two Tri2 (Tri2A & Tri2B) copies, decorate the exterior of the MCP shell, riveting neighboring MCPs together, while small capsid protein (SCP) molecules tip each MCP tower (Figure 1A). SCP, which varies from 8kDa in HCMV and 15 kDa in KSHV, has the greatest structural and sequence variability amongst herpesviruses [19,20], being dispensable among both α- and γ-herpesviruses but not β-herpesviruses [15,21].

As a β-herpesvirus, HCMV has several features which differentiate it from members of the other two *Herpesviridae* subfamilies. For instance, with greater genomic density than its α-herpesvirus cousins [22,23], the HCMV capsid is subject to internal pressures beyond the 18 atmospheres experienced by herpes simplex virus (HSV) [24].

Interestingly, in HCMV the stabilizing capsid vertex-specific components (CVSCs) present in HSV are replaced with the 150 kDa β-herpesvirus specific tegument phosphoprotein (pp150) [16,25,26], despite HCMV CVSCs remaining at the portal vertex [17,27,28]. Indeed, structural studies by our lab and others have found pp150 coats the capsid exterior of mature HCMV virions, bridging the gap between triplex heterotrimers and the small capsid protein (SCP) seated atop the neighboring major capsid proteins (MCP) of pentons and hexons (Fig. 1A) [15,17,28].

Pp150 is thought to bind to nascent HCMV nucleocapsids in the nucleus and co-egress to the cytoplasm where it enhances capsid stability and regulates maturation events at the viral assembly complex (vAC) [29-31]. The capsid bound, N-terminal region of pp150 consists of an upper and lower helix bundle with the later standing off Tri and the former contacting a neighboring SCP (Figure 1A) [15,16,32]. Near the SCP and Tri binding sites of the helix bundles are two highly conserved regions (designated as CR1 & CR2) [16,32]. Peptide mimetics of CR2 were shown to inhibit capsid binding of pp150 though the precise mechanism of inhibition remains unclear [33]. While other investigations including those from our lab have observed SCP to mediate pp150-capsid association the interface between SCP and pp150 has yet to be thoroughly characterized and it remains unclear which, if any, residues or domains are implicated in pp150-capsid binding [15,16].

In this study, we utilized bacterial artificial chromosomes (BACs) to introduce structure guided mutations at the pp150-SCP interface of HCMV and observed the impact on viral protein interactions and replication. Biochemical analysis and fluorescence imaging suggest mutations along the pp150-SCP interface invariably produce lethal phenotypes in HCMV which can likely be attributed to disruption of pp150-SCP association. Meanwhile, we identified a mutation at the pp150-MCP interface which significantly attenuates but does not entirely abolish viral replication. Through ultra-structural analysis, we observed no qualitative effect on capsid assembly, genome retention, or cytoplasmic maturation events, suggesting that this mutation to pp150-MCP interface may impact replication following egress from the cell and that these attenuated virions may represent live-attenuated vaccine candidates. We also provide the first structural evidence to support previous assertions that pp150 associates with capsids within the nucleus of infected cells and preferentially binds genome containing capsids.

4.3. Materials and Methods

4.3.1. Bacterial artificial chromosome mutagenesis of HCMV AD169WT

To mutate the pp150 and SCP, we employed a seamless BAC technique using galK as the selection marker [34]. Briefly, the BACmid of wild type AD169, vDW215-BADrUL131 [35] was transformed into E. coli SW102. A galK DNA fragment that was made from pgalK by PCR and contains ends homologous to the open reading frame (ORF) of pp150 or SCP was electroporated into SW102 (harboring vDW215-BADrUL131) to replace pp150 or SCP by homologous recombination, resulting in the following BACmid: vDW215-BADrUL131-pp150galK or vDW215-BADrUL131-SCPgalK. Then the galK DNA was replaced with a PCR fragment composed of pp150 or SCP DNA, from which fragment the indicated mutations were made, resulting in vDW215-BADrUL131-pp150 E273K D274K D276K, vDW215-BADrUL131-pp150 F245A Y249A, vDW215-BADrUL131-pp150 R240E R251E K255E, vDW215-BADrUL131-pp150- N248E N252E, vDW215-BADrUL131-pp150 K255E, vDW215-BADrUL131-pp150 R240E R251E, vDW215-

BADrUL131-pp150 N260D S263D, vDW215-BADrUL131-pp150 D262K D265K, vDW215-BADrUL131-SCP K46E R49E, and vDW215-BADrUL131-SCP M67W. The resultant BACmids were transfected into ARPE-19 cells (ATCC® CRL-2302™) to make respective viruses. The BACmids and viruses were verified by restriction enzyme digestion, DNA sequencing, and PCR. The complete pp150 and SCP gene was sequenced and confirmed to be correct.

4.3.2. Molecular cloning

To express PP150 and SCP proteins, we performed cloning by inserting the pp150 or SCP gene DNA into the linearized pcDNA3 vector using BamHI and HindIII enzymes. DNA fragments were produced through PCR, utilizing BAC DNAs (wild type or mutant) as templates. During PCR, we incorporated an HA tag at the C-terminus of pp150 and a FLAG tag at the N-terminus of SCP. Subsequently, these DNA fragments were ligated with the linearized pcDNA3 vector using DNA ligase. Following transformation into E. coli, colonies were selected and verified via electrophoresis in an agarose gel. The wild type (WT) and mutant plasmids were confirmed through Sanger sequencing.

4.3.3. Co-Immunoprecipitation (Co-IP) assay and immunoblot analysis

Initially, PP150- and SCP-expressing plasmids were co-transfected into HEK293T cells using Lipofectamine 3000 Reagent (ThermoFisher Scientific, Cat# L3000015) and allowed to incubate for 24 hours. At 24 hours post transfection (hpt), cells were lysed. Lysis was performed using an ice-cold lysis buffer comprising 25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% NP-40, 1 mM EDTA, and 5% glycerol. Additionally, this lysis buffer was supplemented with a protease inhibitor cocktail (Sigma, Cat# P8340), and the cell lysis process was performed on ice for a duration of 10 minutes. Following this, the lysates underwent centrifugation at 3,000 x g for 5 minutes, and the resulting supernatants were meticulously transferred to new tubes.

For the subsequent immunoprecipitation step, these supernatants were subjected to overnight incubation at 4°C, in the presence of specific antibodies such as mouse anti-HA, and mouse anti-FLAG. These incubation mixtures were then combined with protein G-Sepharose beads (Amersham Pharmacia Biotech AB, Sweden), as per the manufacturer's instructions, and allowed to further incubate for 3 hours.

Following immunoprecipitation, the beads were washed three times with PBS containing 0.1% bovine serum albumin, along with a protease inhibitor cocktail. The immune-precipitated complexes were then resuspended in a mixture of PBS and 2× Laemmli buffer (20 μl each). After a brief heating to 95°C for 5 minutes, the beads were separated by centrifugation, and the resultant supernatants were subjected to SDS-PAGE and subsequent immunoblotting.

For the immunoblot analysis, protein samples (ranging from 10 to 20 μg loaded in each lane) were separated utilizing a 7.5% polyacrylamide gel, followed by their transfer onto nitrocellulose membranes (Amersham Inc., Piscataway, NJ). Subsequently, these membranes were blocked with nonfat milk (5% in PBS pH 7.4) for a period of 60 minutes at room temperature. Thereafter, the membranes were incubated overnight at 4°C with primary antibodies. Post-primary antibody incubation, the membranes were treated with a horseradish peroxidase-coupled secondary antibody (Amersham Inc.). Protein detection was achieved through an enhanced chemiluminescence method (Pierce, Rockford, Ill.), following standard protocols. To facilitate the detection of additional proteins of interest, the membranes were stripped using a stripping buffer (comprising 100 mM β-mercaptoethanol, 2% SDS, and 62.5 mM Tris-HCl, pH 6.8), washed with PBS containing 0.1% Tween 20, and subsequently reused for subsequent different protein detection.

4.3.4. Immunofluorescence assay (IFA)

For IFA, ARPE-19 cells were cultured on coverslips and subsequently fixed with 1% paraformaldehyde for 10 minutes at room temperature. Following fixation, the cells were permeabilized by treatment with 0.2% Triton-X100 for 20 minutes on ice. Subsequently, a sequential series of incubations was performed. The cells were first incubated with primary antibodies, followed by incubation with Texas Red (TR)- or FITC-labeled secondary antibodies, all of which were dissolved in phosphate-buffered saline (PBS). Each incubation step lasted for 30 minutes. Finally, the cells were equilibrated in PBS, stained to visualize DNA using Hoechst 33258 at a concentration of 0.5 µg/ml, and then securely mounted using Fluoromount G (obtained from Fisher Scientific, Newark, Del.).

4.3.5. Confocal microscopy

We utilized a Leica TCS SPII confocal laser scanning system to examine the cells. During imaging, we recorded two or three channels either simultaneously or sequentially. We took special care to control for potential signal interference between the fluorescein isothiocyanate and Texas Red channels, as well as between the blue and red channels.

4.3.6. CryoEM of BAC derived nucleocapsids

HCMV nucleocapsids were prepared as described previously [36]. Briefly, ARPE-19 cells infected with HCMV-AD169+GFP BAC derived virus were seeded onto 18 x T175 flasks. When ~80% of cells exhibited cytopathic effects of infection, or after ~10 days in pp150-C4 mutants, cells were scrapped and pelleted. Pellets were washed and resuspended in PBS (pH 7.4) containing 1% NP40 solution and protease inhibitors (PI) two times to disrupt the cell membrane. Nuclei were pelleted and resuspended in 8 mL hypertonic lysis buffer (PBS pH 7.4 (387mM NaCl, 1 mM EDTA, 1X protease inhibitor). Nuclei were lysed via repeated passes through a 23 Gauge needle. The Nuclear lysate was centrifuged at 10,000 x g to remove membrane debris and clarified supernatant was placed atop a 30-50% sucrose cushion and centrifuged at 100,000 x g for 1hr at 4°C. Nucleocapsids were drawn from the 30-50% interface and resuspended in 10 times draw volumes of hypertonic lysis buffer and pelleted via ultracentrifugation at 100,000 x g for 1 hr. The pellet was finally resuspended in 10-20 μL of hypertonic lysis buffer prior to TEM grid preparation and imaging.

To prepare cryoEM samples, 2.5 μL of hypertonic lysis solution containing either purified WT or mutant HCMV nucleocapsids was applied to 200 mesh holey Quantifoil carbon or gold grids, which had been glow discharged for 30 seconds. The sample was flash-frozen in a 50:50 mixture of liquid ethane and propane using a Gatan Vitrobot IV. CryoEM movies were recorded on a K3 direct electron detector in a Titan Krios electron microscope (FEI) equipped with a Gatan Imaging Filter (GIF). The microscope was operated at 300 kV at a nominal magnification of 81,000X and calibrated pixel size of 0.55 Å at the specimen level with the camera in super-resolution mode. Using SerialEM, 2693 and 1261 movies with viral particles were recorded for the WT and mutant capsid samples, respectively. The cumulative electron dose on the sample was ~45 $e^-/Å^2$.

4.3.7. Ultra-thin sectioning of infected cells and transmission electron microscopy

WT AD169-BAC infected cells were prepared for TEM as described previously [37]. Briefly, cells were seeded on tissue culture plates for 48 hours and fixed in 2% glutaraldehyde in 1X PBS (pH 7.4). Cells were scraped and thereafter pelleted by low-speed centrifugation. Pellets were subject to 3% uranyl acetate staining following treatment with 1% osmium tetroxide. Samples were dehydrated in ascending ethanol series and embedded in Agar Low Viscosity Resin (Agar Scientific). Samples were sectioned using an UCT ultra microtome (LEICA) with a 35° diamond knife (Diatome, Switzerland). Sections were 55-65 nm thick and collected on 150 mesh hexagonal or slot grids with Formvar-carbon coated support films and stained with saturated uranyl acetate followed by Sato lead citrate on both sides.

To enrich AD169-BAC infected cells for sectioning, fluorescent cells were concentrated via fluorescent activated cell sorting (FACS). Briefly, tissue culture flasks were seeded with mutant AD169-BAC infected cells and grown for 5 days. To prepare for sorting, cells were thoroughly rinsed with PBS (pH 7.4) before trypsinization and incubation at 37° C for several minutes, or until cells detached. Trypsinized cells were resuspended in PBS (pH 7.4) containing 1% bovine serum albumin (BSA) and soybean trypsin inhibitor (ATCC). Cells were pelleted and resuspended in 1% BSA in PBS solution and passed through a 40 um cell strainer (pluriSelect) into FACS culture tubes. Cells were sorted using a FACSAria III High-Speed Cell Sorter (BD Biosciences) gated on GFP. Sorted cells were seeded on two wells of a 6-well plate and supplemented with DMEM growth medium and grown for 2 days prior to fixation.

For nucleocapsid quantitative analysis, TEM imaging was performed using a FEI Titan 80-300 electron microscope operated at 300 kV. Images were recorded on a DE-64 camera (Direct Electron) at a nominal magnification of 10,000 x. Montage micrographs of cell sections were collected using SerialEM [38], by collecting polygons in "view" mode. Defocus was maintained at -40 um during the data acquisition. A, B, and C capsids were annotated on montages using 3dmod in the IMOD package [39]. Non-montaged TEM micrographs were recorded on a charged coupled detector (CCD) camera in an FEI Tecnai G2 T20 electron microscope operated at 200 kV to enhance contrast.

4.3.8. CryoEM image pre-processing and 3D reconstruction

We performed frame alignment and motion correction to generate a cryoEM micrograph from each movie with cryoSPARC [40]. Patch-aligned and dose-weighted micrographs were transferred to Relion 4.0 for further processing and Topaz automated particle picking [41,42]. In total, micrographs for the WT sample contained approximately 2966 A-capsids, 1527 B-capsids, and 42 C-capsids, whereas those for the mutant sample contained 1215 A-capsids, 595 B-capsids, and 22 C-capsids. To improve processing speed, particles were extracted and binned eight times prior to 2D image classifications. A 3D Gaussian ball created by back projection of particles without orientation assignments was used as an initial model for 3D classification while imposing icosahedral symmetry. The icosahedral reconstructions of both A- and B-capsids reached a resolution of 17.6 Å.

To improve reconstruction resolution, we generated sub-particle reconstructions of the capsid vertices as described previously [28]. Briefly, we used Relion's symmetry expansion method, with I3 symmetry selected, to generate sixty equivalent orientations for each capsid particle with unique Euler angles for rotation, tilt, and psi. Given that there are five redundant orientations of each particle about each 5-fold symmetrical vertex, which have identical tilt and psi angles but different in-plane angle, we kept only one of each set of five, to represent each 5-fold vertex sub-particle. These orientations were then used to identify the position of each 5-fold vertex in Cartesian coordinates within the 2D projection images for extraction as new particles totaling 31,661; 16,255; and 482 WT A-capsid, B-capsid, and C-capsid and 14,354 and 6,958 for pp150-C4 A-capsid and B-capsids, respectively. Further processing was then carried out imposing C5 symmetry onto the newly extracted particles to reach 3.1, 3.3, and 13.6Å for WT A-capsid, B-capsid, and C-capsid and 3.7, and 4.1 Å for pp150-C4 mutant A-capsid and B-capsid, respectively.

4.4. Results

4.4.1. Amino acid mutations at the SCP-pp150 interface produce lethal and attenuating phenotypes

Atomic models of HCMV present a relatively small (~446 Å^2) pp150-SCP interface, composed of clusters of complementary hydrophobic or polar amino acids (Fig. 1A-C) [15,16,32]. From the pp150 face, negatively charged residues (E273, D274, and D276) accommodate a positively charged protrusion from SCP (K46, R49), while aromatic residues from pp150 (F245, Y249) form an Aro-Met-Aro bridge with SCP (M67), and polar uncharged residues from pp150 (N248, N252) contact the C-terminal residues of SCP. We introduced mutations to substitute each complementary cluster with a non-complementary one (Fig. 1C-D). The impact of each set of mutations on pp150-SCP interaction was first predicted *in silico* using the software package mmCSM-PPI [43]

(Table S1). Mutations to the F245-M67-Y249 bridge and charged faces were predicted to significantly alter the free energy of association between pp150 and SCP, while those pp150 mutations near the SCP C-terminus (pp150-C4) and negative controls far from the interface (pp150-C5 & C6) were not predicted to significantly impact binding (Table S1).

To validate these structure-based predictions, we transfected ARPE19 cells with AD169-BACmids carrying cluster mutations to either pp150 or SCP and GFP as a positive marker (Fig. 1). As positive control, cells were transfected with wild-type (WT) AD169-BACmids which produced WT virus that spread throughout cell culture as characterized by spread of GFP signal (Fig. 1E-F). In keeping with computational predictions, cluster mutations outside the pp150-SCP interface had no observable impact on viral replication (Fig. 1E), however mutations at any polar or hydrophobic regions of the interface, including near N248 and N252 of pp150, proved lethal to the virus as indicated by lack of GFP spread beyond initially transfected cells 7 days post transfection (Fig. 1E-F). Interestingly, the R240E, R251E, K255E mutant (pp150-C4) demonstrated GFP signal spread that was significantly attenuated compared to WT HCMV (Fig. 1F). While the lethal mutations were localized to the direct polar or hydrophobic interfaces between pp150 and SCP, R240 and K251 of pp150-C4 only interact with the far C-terminal region of SCP, while K255 interfaces with MCP, suggesting pp150-C4 is not implicated in SCP association. To determine whether mutations to the SCP or MCP interface were responsible for the attenuation, we generated separate pp150 R240E, K251E and pp150 K255E mutants (pp150-C4a and pp150-C4b respectively) leading to WT and attenuated phenotypes respectively (Fig.

1D-F). Viral replication, as exhibited by GFP spread, showed the pp150-C4a mutant to have wild-type kinetics, whereas pp150-C4b exhibited attenuated viral spread similar to pp150-C4 (Fig 1F).

4.4.2. Amino acid mutations at the SCP-pp150 interface disrupt pp150-capsid interactions

We next sought to determine whether disruption of the pp150-SCP interaction was responsible for the lethal and attenuated phenotypes we observed. To this end, we cloned pp150-hemagglutinin (HA) and FLAG-SCP tagged proteins into a pcDNA3 expression vector. The plasmid DNAs of both WT and mutant genes were co-transfected into ARPE-19 cells and fixed for immunofluorescence assays (IFA) (Fig. 2A). IFAs revealed colocalization of WT SCP to pp150 as cytosolic puncta of transfected ARPE-19 cells (Fig. 2A, yellow). By contrast, plasmids containing lethal SCP or pp150 mutants exhibited delocalized signal dispersed throughout the cytosol (Fig. 2A). The pp150-C4 mutant virus demonstrated SCP colocalization similar to WT, indicating pp150-C4 readily associates with SCP despite attenuated replication. This finding is consistent with the notion that the K255E mutation drives the attenuation via disruption of the pp150-MCP interface. To corroborate the IFA results, the plasmids were co-transfected into HEK-293 cells for co-immunoprecipitation assays. In WT controls, FLAG-tagged SCP was readily detected in the immunoblot; but the lethal mutations showed no FLAG signal (Fig. 2B). By comparison, pp150-C4 demonstrated SCP signal consistent with WT, further supporting our hypothesis that the interaction is not disrupted (Fig. 2B). These findings together demonstrate that lethal cluster mutations are those that disrupt the pp150-SCP interface and that disruptions to other interfaces can produce attenuating mutations.

4.4.3. Pp150-attenuated mutants maintain genome packaging and viral maturation

To characterize the intracellular consequences of the pp150-C4 mutant's impaired pp150-MCP interaction on viral assembly and replication, we enriched pp150-C4 mutant HCMV infected cells using fluorescence activated cell sorting (FACS) gated on GFP signal. These pp150-C4 infected cells were fixed, embedded in resin, and used to generate ultra-thin sections (~55 nm) for TEM analysis. These ultra-thin sections were transferred to formvar-carbon coated TEM grids, stained for large-field of view transmission electron microscopy with montaging in SerialEM [44]. Nuclear membranes were identified and nucleocapsids populations were categorized into three groups including (A-) empty, (B-) scaffold protein-containing, and (C-) genome-filled capsids, based on appearance when observed by electron microscopy (Fig. 4A). WT nuclei appeared with greater proportions of B- and C-capsids compared to A-capsids and this distribution was similar in pp150-C4 infected cells (Fig. 3A-B). In addition to having nucleocapsid distribution similar to WT, pp150-C4 infected cells had vAC associated viral products including enveloped C-capsids, suggesting efficient viral maturation unlike pp150-deletion mutants (Fig. 3C) [45].

4.4.4. Pp150 is translocated across the nuclear membranes to bind capsids in the nucleus

Because pp150 lacks a recognizable nuclear localization signal or obvious chaperones, it is unclear how pp150 is translocated into the nucleus and whether it binds to the capsids there or in the cytosol where it also accumulates [30,33]. Indeed, previous studies have demonstrated the nucleocapsids capacity for nuclear egress without pp150 [30].

After finding pp150-C4-HCMV had no obvious defects in vAC formation, we sought to determine whether pp150 was being properly incorporated into the nucleocapsids of the mutant virus. The low replication efficiency of pp150-C4 virus complicated efforts to purify complete virions from culture media and necessitated a modified culture procedure to purify nucleocapsids from infected cells. Similar to WT HCMV (Fig. 4A-C, left), the pp150-C4 HCMV nucleocapsids purified from cell nuclei were predominantly A-capsid or B-capsids (Fig 4A-C, right), consistent with previous findings [46,47].

Initial cryoEM capsid reconstructions from WT and pp150-C4 mutant samples showed no obvious densities at locations occupied by pp150 in WT virions (Figs. 4B-C). 3D reconstruction from a small subset of genome containing C-capsids also lacked pp150 densities (Fig. 4D). While these findings were consistent with previously reported structures [46], closer inspection of cryoEM images of nucleocapsid revealed conspicuously empty spaces between MCP towers; by contrast, in the cryoEM micrographs of virions, pp150 tegumentation prevented us from distinguishing between MCP towers (Fig. 4A-B). Using cryoEM images from virions as a template, we identified a visually distinguishable subset of 237 A-capsids and 61 B-capsids with density corresponding to the tegument in WT virus (Fig. 4E). When subjected to in-depth data processing, this subset of images yielded a 3D reconstruction with cryoEM densities corresponding to pp150 and tRNAs similar to those observed in previously published structures (Fig. 4E) [48]. While insufficient numbers of tegument bound particles were present in the pp150-C4 dataset for 3D reconstruction to resolve similar densities, capsid images with tegument like that of the WT were observed, suggesting pp150-capsid association in the mutant (Fig. 4F).

4.5. Discussion

Amongst herpesviruses, virion assembly is a tightly regulated process driven by virus-specific proteins, making it an attractive target for drug development. The β-herpesvirus specific tegument protein pp150 is integral to cytoplasmic assembly and requires nucleocapsid incorporation to carry out its functions [16,30,33,49]. Recently, the detailed interactions between pp150 and the nucleocapsid were elucidated with the publication of several high-resolution cryoEM structures of HCMV virions [16,17]. The current study leveraged these structures to carry out the mutational analysis of several well-conserved domains essential to the pp150-SCP interaction, identify a mutation which considerably attenuates viral replication, and confirm that pp150 associates with viral capsids within the nucleus of infected cells.

Previous functional investigations of pp150 have largely focused on CR1, CR2, and the highly conserved cysteine tetrad which were recognized decades ago [32,33]. Indeed, peptide mimics of CR2 have shown promise as inhibitors of pp150-capsid association, however the mechanism of inhibition remains unclear as CR1 and CR2 do not obviously interface with capsid proteins according to available CMV structures [33]. We found that amino acid substitutions in the conserved polar or hydrophobic domains directly involved in the pp150-SCP interface were detrimental to viral replication and these effects could be directly attributed to disruption of pp150-SCP association (Figs. 1 & 2). No mutations presented here overlapped with CR1 or CR2, however, many of the residues selected were well conserved across SCMV and MCMV (Fig. 1B). Thus, we suspect the well-conserved residues between amino acids 240-276 are essential for capsid association of pp150 and its homologues across cytomegaloviruses of various species, but this requires further study.

Without an obvious nuclear localization signal or chaperone, it is unclear how pp150 reaches the nucleus. Our IFA and coimmunoprecipitation assays demonstrate that pp150 interacts with SCP in the cytoplasm of co-transfected cells (Fig. 2A). Considering pp150 and SCP are expressed at similar timepoints during viral infection [50], it is conceivable that these interactions likewise occur in virus infected cells. As SCP is thought to be shuttled to the nucleus by scaffolding proteins along with MCP [51], pp150's cytoplasmic association to this complex would provide a plausible mechanism for nuclear transport, however testing this hypothesis will require a future investigation.

Despite evidence suggesting pp150 tegumentation occurs in the nuclei of infected cells, previous structural investigations of genome-free, nucleus derived HCMV nucleocapsids indicated the absence of pp150 [46,52]. We likewise observed that most nucleus derived capsids in our dataset lacked tegument in cryoEM images and 3D reconstructions (Fig. 4). Curiously, C-capsids purified from the nucleus, typically considered the most likely candidates for pp150 association, also lacked pp150 tegument as revealed by cryoEM reconstruction. Approximately 8% and 4% of A- and B-capsids respectively had tegument, consisting of pp150 and tRNA, observable in cryoEM images (Fig. 4E). As nucleus derived C-capsids are notoriously unstable [47], and in the infected cells, A-capsids compose only a small percentage of viral particles (Fig. 3 and [47], one may reasonably assume that many A-capsids in our purified sample arose from C-capsids whose genomes ejected prematurely. Therefore, the greater proportion of tegumented A-capsids we observed may reflect preferential association of pp150 to C-capsids over B-capsids.

Prior studies demonstrated that pp150-deletion renders HCMV unable to mature beyond the cytoplasmic stage, with few cytoplasmic capsids being observed and a highly vesiculated vAC [30]. By contrast, our pp150-C4 mutant HCMV appeared capable of efficient genome packaging, nuclear egress, and vAC formation despite significant attenuation to viral spread (Figs. 1 & 3). This would appear to suggest that the effects of such a mutation are either too subtle for detection in our TEM surveys or that pp150-C4 mutants are impacted at another point during their replication cycle. The large number of complex processes comprising capsid maturation and incomplete characterization of pp150's roles throughout makes it difficult to pinpoint where along that process the K255E mutation impacts. However, based on capsid structures, it would appear the pp150-MCP interaction is important, though not essential, for the pp150's attachment to the capsid (Fig. 1A-C). Together, our results suggest pp150-C4 produces complete virions with significantly reduced infectivity, potentially offering a novel platform for development of a live-attenuated vaccine.

Contemporary live-attenuated HCMV vaccine candidates rely on gain-of-function from less pathogenic laboratory strains, however, these strains harbor genetic mutations and deletions which may unpredictably impact the fidelity of the induced immune response [4-6,12,13]. Here a single-amino-acid mutation to the highly conserved pp150 tegument protein results in significantly attenuated viral replication, without the need to delete entire genes. Further, the pp150 K255E mutation does not appear to prevent pp150-capsid association (Fig. 4G), meaning that this important antigen is still presented to the host. Thus, the pp150 K255E mutation could, in theory, be introduced to clinical strains and be expected to similarly attenuate viral replication with minimal impact on immune response. Additionally, because pp150 K255 is conserved in CMV of other species, this mutation may be more amenable to studies in animal models (Fig. 1B). While this mutation alone may not be the answer to the development of a live-attenuated vaccine, we believe this is an example of how rational design strategies may produce viable vaccine candidates to serve as prophylactics against HCMV transmission. Further, the work reported here represents the first time structure-guided design and BAC mutagenesis has led to identification of lethal, as well as an attenuated HCMV mutant and raises important questions regarding the role of the pp150-SCP and pp150-MCP interactions in viral replication. The latter produces far less infectious virus compared the wild-type virus and warrants further study in the pursuit as candidate of live-attenuated vaccine.

5. Conclusion

In summary, guided by atomic structure of HCMV, we demonstrated the sensitivity of the pp150-SCP interface to mutations, while also providing the first structural evidence of pp150 binding to capsids inside nucleus. While many of these interacting residues are essential to HCMV replication, we identified a single amino-acid mutation that significantly attenuates intercellular transmission of the virus in vitro, which opens the door for further exploration vis-a-vis developing a live attenuated HCMV vaccine.

4.6. CRediT authorship contribution statement

Alex Stevens: Overall experimental design and execution of the electron microscopy-related components, data analysis and result interpretation, manuscript drafting and figure preparation, paper editing. **Ruth Cruz-Cosme**: BAC-related experiments, co-IP, and western blot assays. **Najealicka Armstrong**: molecular cloning to generate plasmids and immunofluorescence assays. **Qiyi Tang**: Design and supervision of BAC-related experiments, co-IP experiments, fluorescence microscopy, and manuscript editing. **Z. Hong Zhou**: Overall project design and supervision of execution, research fund acquisition and management, manuscript draft and editing.

4.7. Declaration of competition interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

4.8. Data availability

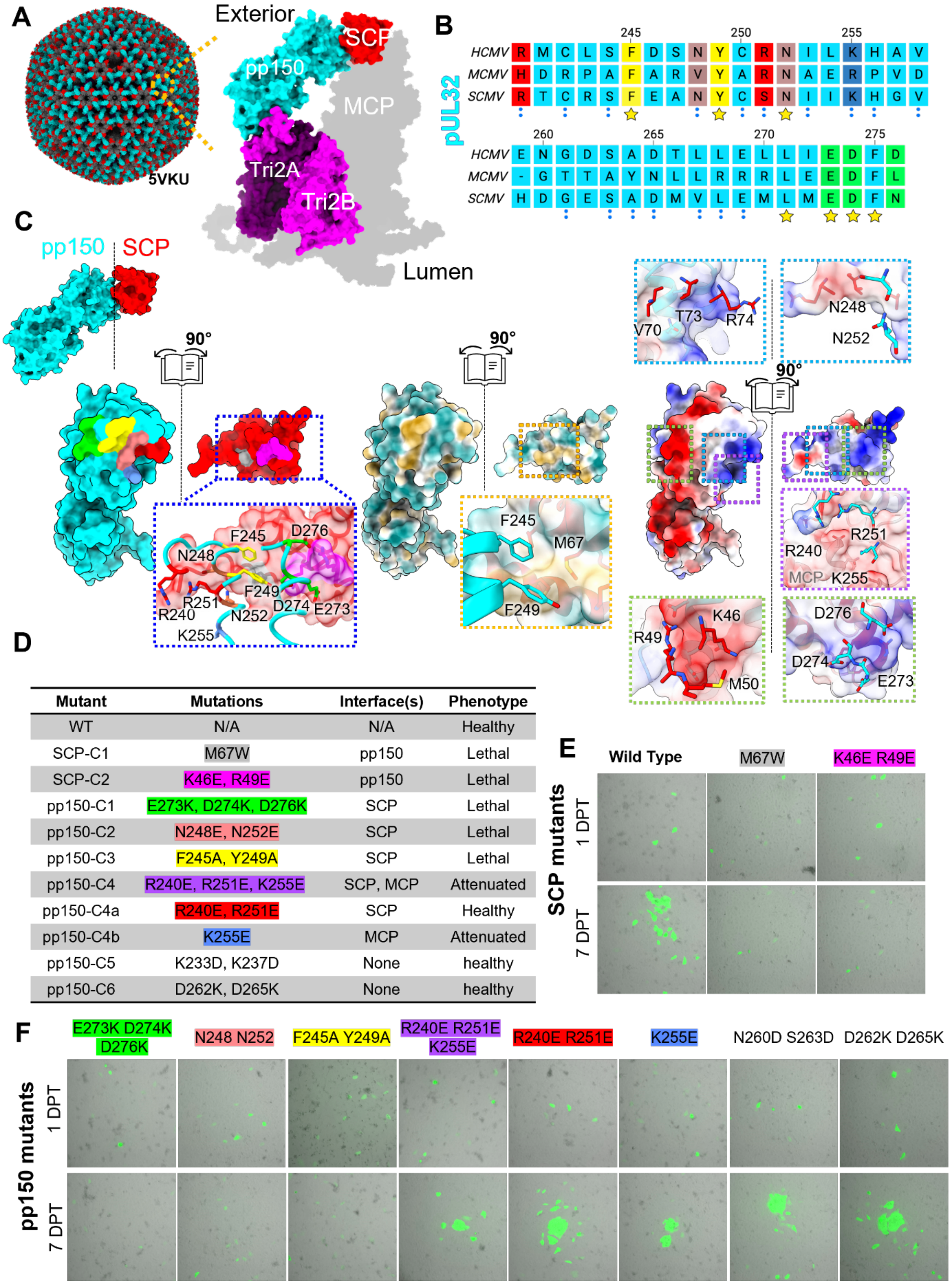
The cryoEM structures reported here are available at the Electron Imaging Database under the accession code of EMD-44647 (mutant A-capsid), EMD-44648 (mutant B-capsid), EMD-44636 (WT A-capsid), EMD-44640 (WT B-capsid).

4.9. Acknowledgements

We thank Sergey Ryazantsev & Chunni Zhu for the cell embedding and thin-sectioning service. This work was supported by grants from the US National Institutes of Health (R01DE028583 to ZHZ). A.S. Received support from NIH Ruth L. Kirschstein National Research Service Award AI007323. We acknowledge the use of resources in the Electron Imaging Center for NanoSystems (EICN) supported by UCLA, National Science Foundation (DMR-1548924 and DBI-1338135) and National Institutes of Health (S10RR23057 and U24GM116792). We acknowledge support from the UCLA AIDS

Institute, the UCLA Brain Research Institute Electron Microscopic Core, the James B. Pendleton Charitable Trust, and the McCarthy Family Foundation.

4.10. Figures



Mutant	Mutations	Interface(s)	Phenotype
WT	N/A	N/A	Healthy
SCP-C1	M67W	pp150	Lethal
SCP-C2	K46E, R49E	pp150	Lethal
pp150-C1	E273K, D274K, D276K	SCP	Lethal
pp150-C2	N248E, N252E	SCP	Lethal
pp150-C3	F245A, Y249A	SCP	Lethal
pp150-C4	R240E, R251E, K255E	SCP, MCP	Attenuated
pp150-C4a	R240E, R251E	SCP	Healthy
pp150-C4b	K255E	MCP	Attenuated
pp150-C5	K233D, K237D	None	healthy
pp150-C6	D262K, D265K	None	healthy

Figure 4- 1. Structure-guided mutational analysis of the pp150-SCP interface. (A) Atomic model of the HCMV capsid with pp150 coat (left). The inset (right) shows zoomed-in view of subunits colored to emphasize different subunit interactions. (B) Sequence alignment of HCMV, murine CMV (MCMV), and simian CMV (SCMV) pp150 homologues between pp150 residues 240-276. (C) surface interfaces between pp150 and SCP, with mutations colored by clusters to enhance detail (left). Surfaces with Hydrophilic, neutral, and hydrophobic areas colored in cyan white and gold respectively (middle) while positive, neutral, and negative electrostatic potentials are shown in blue, white, and red respectively (right). (D) Table listing cluster mutations, and their effect on virus replication, as observed in transfected cells. (E) Representative Fluorescence microscopy images of cells transfected with SCP mutant BACs, 1 (top) or 7 (bottom) days post transfection (DPT). (F) Same as (E) but from pp150 mutant BACs at 1 or 7 DPT.

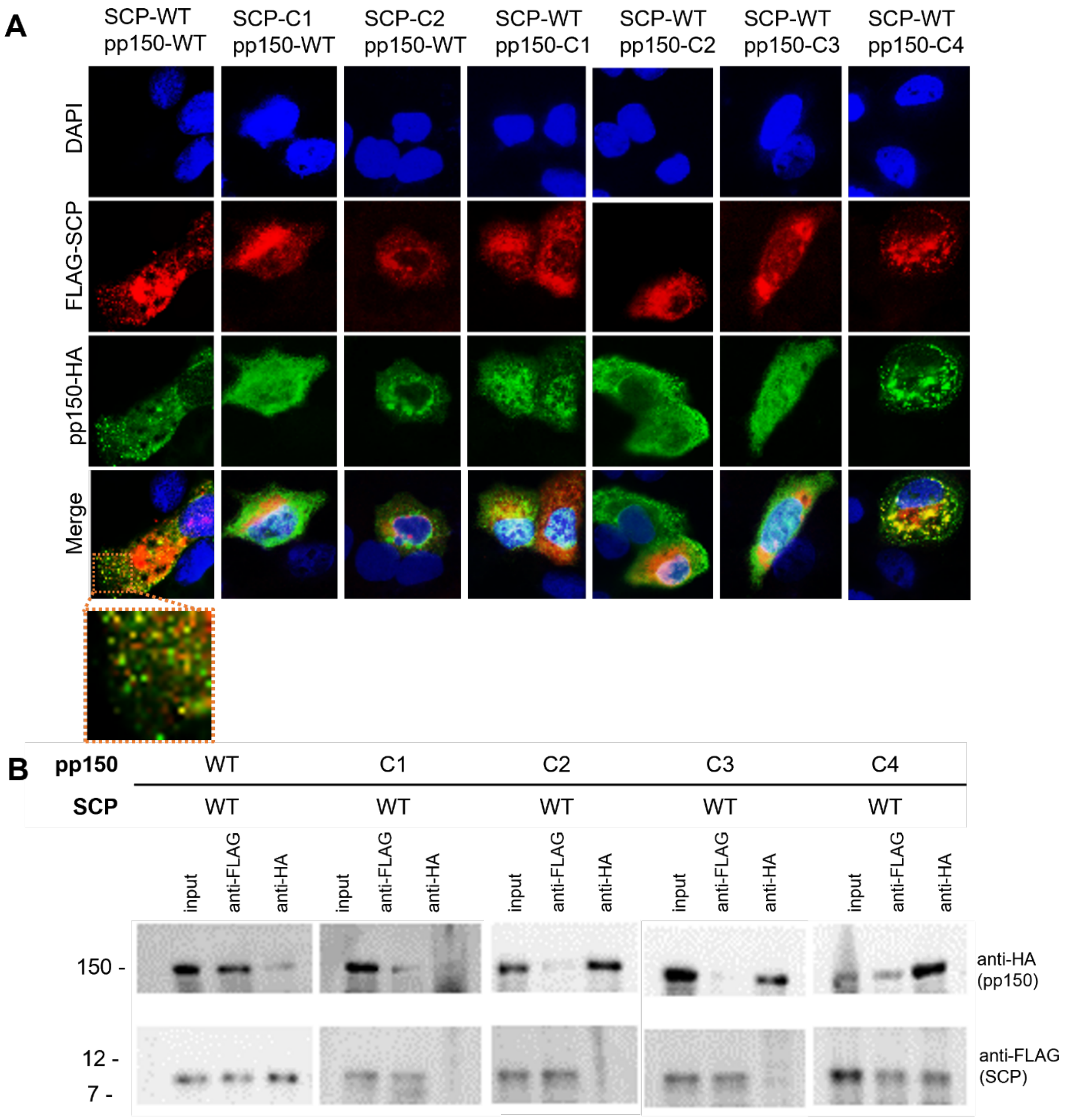
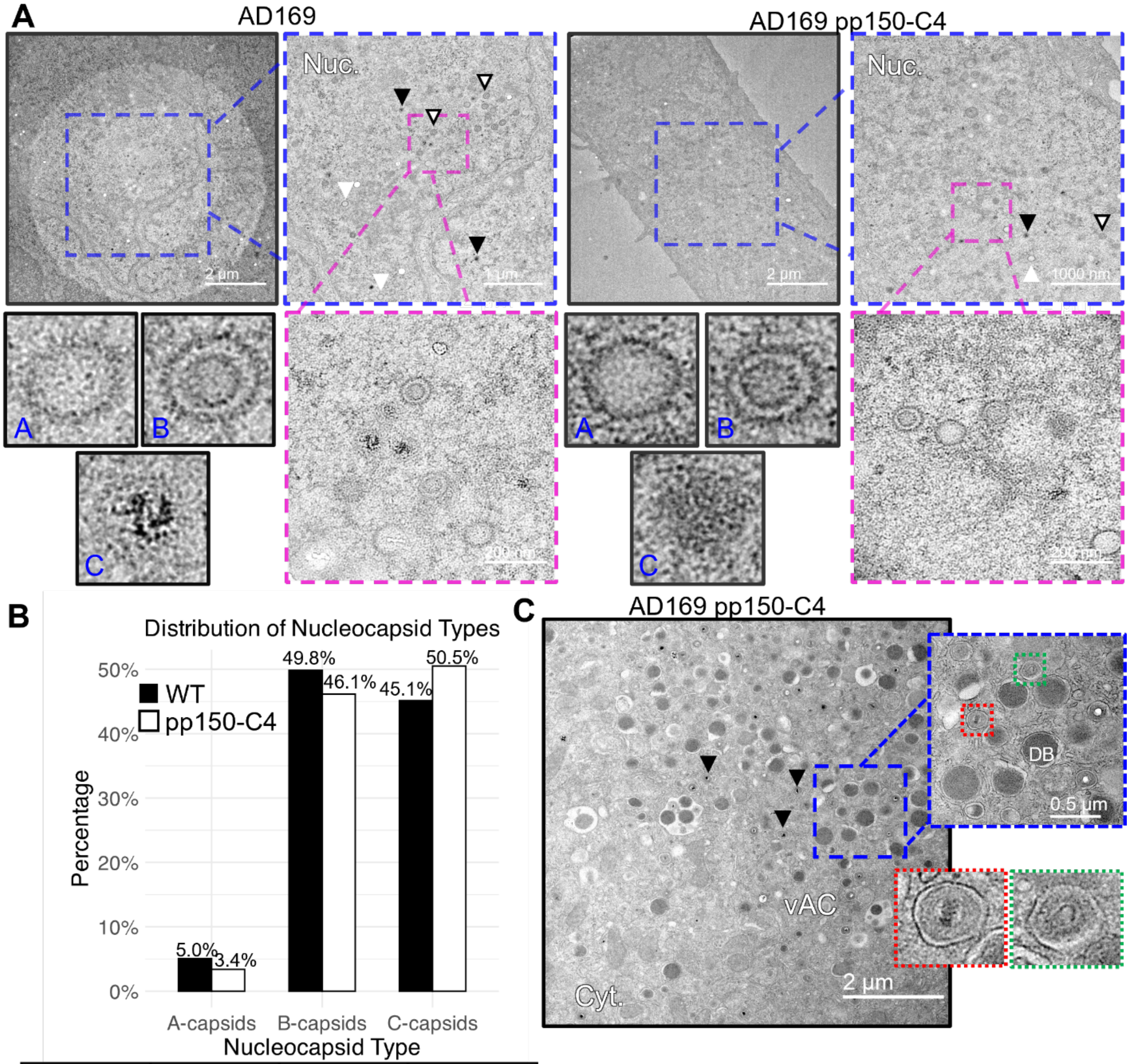


Figure 4- 2. Characterization of mutant pp150 and SCP interactions.

(A) Immunofluorescence assays of ARPE-19 cells transfected with either WT or cluster mutations on pp150 or SCP showing colocalization between anti-HA (green) and anti-FLAG (red) with DAPI stained nuclei (blue). (B) Coimmunoprecipitation (CoIP) assays showing interactions of pp150 cluster mutants with WT SCP. Protein extracts were precipitated with anti-HA or anti-FLAG antibodies and resolved by SDS-PAGE. Protein-protein interactions were immunodetected using anti-HA and anti-FLAG antibodies.



Virus	A-capsid	B-capsid	C-capsid
WT	98	969	878
Mutant 4	58	790	865

Figure 4- 3. Thin section TEM of ARPE-19 cells infected with WT and mutant HCMV (A) Representative TEM micrographs of WT (left) and AD169 pp150-C4 infected cells (right) with zoomed in views highlighting different capsid types, with some A- (white arrowheads), B- (white with black outline arrowheads), and C-capsids (black arrowheads) indicated. (B) Distribution of A-, B-, and C-nucleocapsids as percentages of total counted and totals in accompanying table (right). (D) Representative micrograph of vAC from AD169 pp150-C4 infected cells with zoomed in views of particles undergoing envelopment (right, top) along with dense bodies (DB) and examples of enveloped particles (right, bottom).

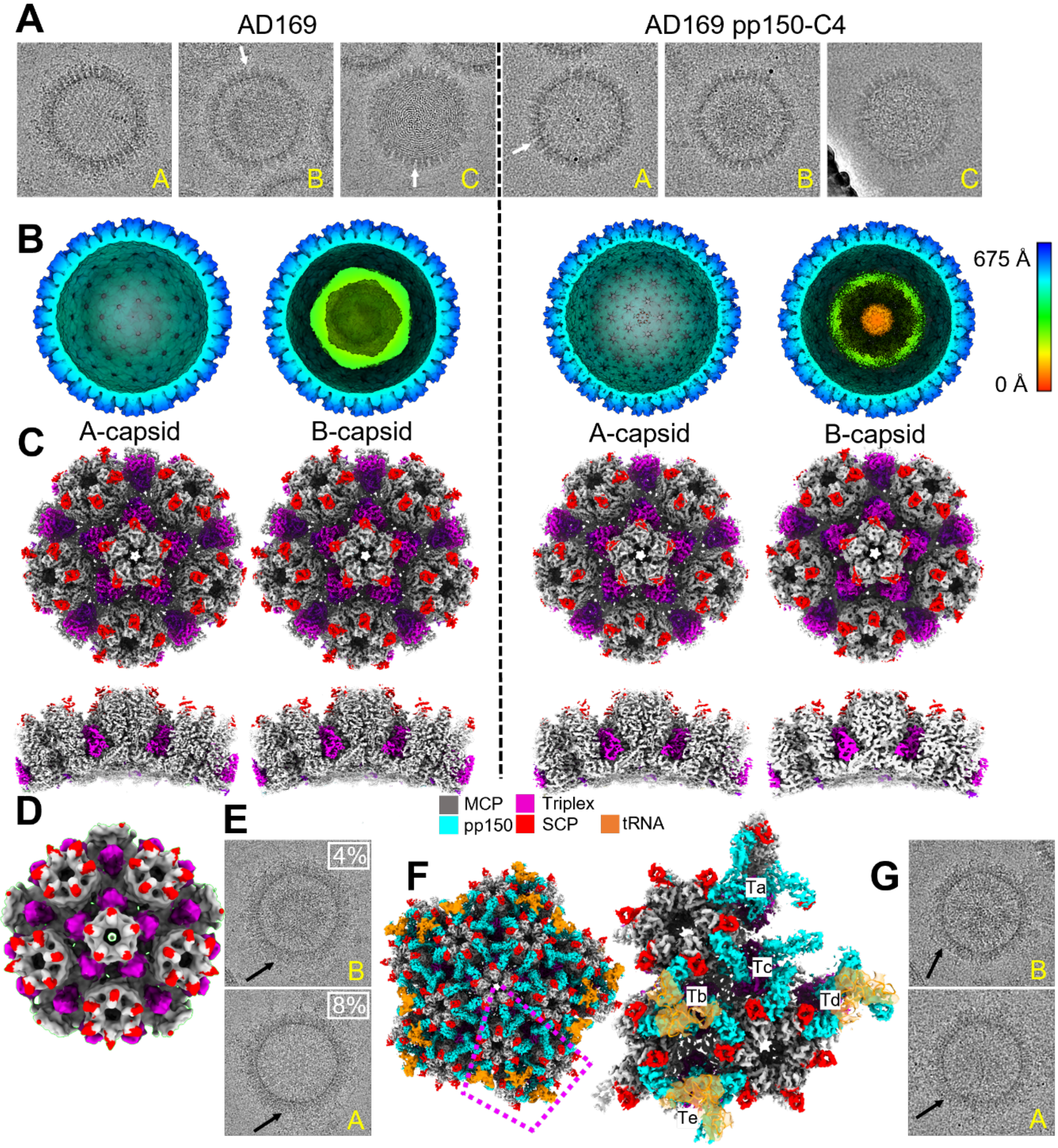


Figure 4- 4. 3D reconstructions of nucleocapsids reveal variable tegumentation.

(A) Representative cryoEM images of A-, B-, and C-capsids purified from WT AD169 (left) and AD169 pp150-C4 mutant infected cells (right) and a virion particle (lower middle). Comparison of MCP towers in 2D electron micrographs (inset greyscale images, blue arrowhead) and empty spaces (black outlined arrowheads) to equivalent regions from whole capsid 3D reconstructions (insets) from B. (B) 3D Reconstruction of WT (left) and AD169 pp150-C4 A- and B-capsids colored by distance from center of capsid with MCP towers (blue arrowheads) and the empty spaces between them (black outlined arrowheads). (C) Capsid vertex sub-particle reconstruction from particles in B, colored by capsid proteins to convey nucleocapsid composition, viewed from outside the capsid and as a cross section (bottom). (D) Low resolution vertex sub-particle reconstruction from the WT C-capsids showing lacking pp150 density. (E) Representative cryoEM images demonstrating fuzzy tegument (black arrow) around B- (top) and A-capsids (bottom) with percentages from total B- and A-capsids observed with tegument and (F) a vertex sub-particle generated from particles in E (right) showing the presence of pp150 density (cyan) and asymmetric unit with distribution of pp150 and tRNA (right). (G) Representative images of pp150-C4 mutant capsids with fuzzy tegument (black arrows).

Mutant	**Mutations**	**Interface(s)**	**Phenotype**	**mmCSM-PPI ΔΔG (kcal/mol)**
WT	N/A	N/A	Healthy	NA
SCP-C1	M67W	pp150	Lethal	-0.91
SCP-C2	K46E, R49E	pp150	Lethal	-2.09
pp150-C1	E273K, D274K, D276K	SCP	Lethal	+0.09
pp150-C2	N248E, N252E	SCP	Lethal	-0.49
pp150-C3	F245A, Y249A	SCP	Lethal	-1.08
pp150-C4	R240E, R251E, K255E	SCP, MCP	Attenuated	-0.03
pp150-C4a	R240E, R251E	SCP	Healthy	0.05
pp150-C4b	K255E	MCP	Attenuated	0
pp150-C5	K233D, K237D	N/A	healthy	-0.03
pp150-C6	D262K, D265K	N/A	healthy	-0.03

Table 4- S1. In-silico mutational analysis of pp150-SCP interface.
Table from Figure 1D extended to include predicted change to binding affinity (ΔΔG) in kcal/mol by either SCP or pp150 mutant clusters. Decreased affinity values are negative and colored red and increased affinity are positive colored in blue.

4.11. References

1. Zuhair, M. et al. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. in *Reviews in Medical Virology* Vol. 29 e2034 (2019).
2. Manicklal, S., Emery, V.C., Lazzarotto, T., Boppana, S.B. & Gupta, R.K. The “Silent” Global Burden of Congenital Cytomegalovirus. in *Clinical Microbiology Reviews* Vol. 26 86-102 (2013).
3. Britt, W. Maternal Immunity and the Natural History of Congenital Human Cytomegalovirus Infection. *Viruses* **10**, 405 (2018).
4. Das, R. et al. Safety, efficacy, and immunogenicity of a replication-defective human cytomegalovirus vaccine, V160, in cytomegalovirus-seronegative women: a double-blind, randomised, placebo-controlled, phase 2b trial. *The Lancet Infectious Diseases* **23**, 1383-1394 (2023).
5. Murata, S., Oshima, N., Iwasa, T., Fukao, Y. & Sawata, M. Safety, Tolerability, and Immunogenicity of V160, a Conditionally Replication-Defective Cytomegalovirus Vaccine, in Healthy Japanese Men in a Randomized, Controlled Phase 1 Study. *Antibodies* **12**(2023).
6. Adler, S.P. et al. Phase 1 Clinical Trial of a Conditionally Replication-Defective Human Cytomegalovirus (CMV) Vaccine in CMV-Seronegative Subjects. *The Journal of Infectious Diseases* **220**, 411-419 (2019).

7. Scarpini, S. et al. Development of a Vaccine against Human Cytomegalovirus: Advances, Barriers, and Implications for the Clinical Practice. *Vaccines* **9**, 551 (2021).
8. Zammarchi, L. et al. Treatment with valacyclovir during pregnancy for prevention of congenital cytomegalovirus infection: a real-life multicenter Italian observational study. *American Journal of Obstetrics & Gynecology MFM* **5**, 101101-101101 (2023).
9. Chou, S. Advances in the genotypic diagnosis of cytomegalovirus antiviral drug resistance. *Antiviral Research* **176**, 104711-104711 (2020).
10. Scarpini, S. et al. Development of a Vaccine against Human Cytomegalovirus: Advances, Barriers, and Implications for the Clinical Practice. in *Vaccines* Vol. 9 551 (2021).
11. Takahashi, M. et al. LIVE VACCINE USED TO PREVENT THE SPREAD OF VARICELLA IN CHILDREN IN HOSPITAL. *The Lancet* **304**, 1288-1290 (1974).
12. Wilkinson, G.W.G. et al. Human cytomegalovirus: taking the strain. *Medical Microbiology and Immunology* **204**, 273-284 (2015).
13. Wang, D. et al. A replication-defective human cytomegalovirus vaccine for prevention of congenital infection. *Science Translational Medicine* **8**, 1-8 (2016).
14. McSharry, B., Avdic, S. & Slobedman, B. Human Cytomegalovirus Encoded Homologs of Cytokines, Chemokines and their Receptors: Roles in Immunomodulation. *Viruses* **4**, 2448-2470 (2012).

15. Dai, X. et al. The Smallest Capsid Protein Mediates Binding of the Essential Tegument Protein pp150 to Stabilize DNA-Containing Capsids in Human Cytomegalovirus. in *PLoS Pathogens* Vol. 9 (ed. Gibson, W.) e1003525 (2013).
16. Yu, X., Jih, J., Jiang, J. & Zhou, Z.H. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. in *Science* Vol. 356 eaam6892 (2017).
17. Li, Z., Pang, J., Dong, L. & Yu, X. Structural basis for genome packaging, retention, and ejection in human cytomegalovirus. in *Nature Communications* Vol. 12 4538 (Springer US, 2021).
18. McGeoch, D.J. Human Herpesviruses. in *Trends in Microbiology* Vol. 1045 (eds Kawaguchi, Y., Mori, Y. & Kimura, H.) 31-32 (Springer Singapore, Singapore, 2018).
19. Close, W.L., Anderson, A.N. & Pellett, P.E. Betaherpesvirus Virion Assembly and Egress. 167-207 (2018).
20. Dai, X. et al. CryoEM and mutagenesis reveal that the smallest capsid protein cements and stabilizes Kaposi's sarcoma-associated herpesvirus capsid. in *Proceedings of the National Academy of Sciences* Vol. 112 E649-E656 (2015).
21. Borst, E.M. et al. A Unique Role of the Human Cytomegalovirus Small Capsid Protein in Capsid Assembly. in *mBio* Vol. 13 (ed. Damania, B.) (American Society for Microbiology, 2022).
22. Muller, C., Alain, S., Baumert, T.F., Ligat, G. & Hantz, S. Structures and Divergent Mechanisms in Capsid Maturation and Stabilization Following Genome

Packaging of Human Cytomegalovirus and Herpesviruses. in *Life* Vol. 11 150 (2021).

23. Bauer, D.W., Huffman, J.B., Homa, F.L. & Evilevitch, A. Herpes Virus Genome, The Pressure Is On. in *Journal of the American Chemical Society* Vol. 135 11216-11221 (2013).

24. Brandariz-Nuñez, A., Liu, T., Du, T. & Evilevitch, A. Pressure-driven release of viral genome into a host nucleus is a mechanism leading to herpes infection. in *eLife* Vol. 8 1-20 (2019).

25. Dai, X. & Zhou, Z.H. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. in *Science* Vol. 360 eaao7298 (2018).

26. Wang, J. et al. Structure of the herpes simplex virus type 2 C-capsid with capsid-vertex-specific component. in *Nature Communications* Vol. 9 3668 (Springer US, 2018).

27. Sae-Ueng, U. et al. Major capsid reinforcement by a minor protein in herpesviruses and phage. *Nucleic Acids Research* **42**, 9096-9107 (2014).

28. Liu, Y.-T.T., Jih, J., Dai, X., Bi, G.-Q.Q. & Zhou, Z.H. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. in *Nature* Vol. 570 257-261 (Springer US, 2019).

29. Tandon, R. & Mocarski, E.S. Cytomegalovirus pUL96 Is Critical for the Stability of pp150-Associated Nucleocapsids. in *Journal of Virology* Vol. 85 7129-7141 (2011).

30. Tandon, R. & Mocarski, E.S. Control of Cytoplasmic Maturation Events by Cytomegalovirus Tegument Protein pp150. in *Journal of Virology* Vol. 82 9433-9444 (2008).

31. AuCoin, D.P., Smith, G.B., Meiering, C.D. & Mocarski, E.S. Betaherpesvirus-Conserved Cytomegalovirus Tegument Protein ppUL32 (pp150) Controls Cytoplasmic Events during Virion Maturation. in *Journal of Virology* Vol. 80 8199-8210 (2006).

32. Baxter, M.K. & Gibson, W. Cytomegalovirus Basic Phosphoprotein (pUL32) Binds to Capsids In Vitro through Its Amino One-Third. in *Journal of Virology* Vol. 75 6865-6873 (2001).

33. Mitra, D., Hasan, M.H., Bates, J.T., Bidwell, G.L. & Tandon, R. Tegument Protein pp150 Sequence-Specific Peptide Blocks Cytomegalovirus Infection. in *Viruses* Vol. 13 2277 (MDPI, 2021).

34. Warden, C., Tang, Q. & Zhu, H. Herpesvirus BACs: Past, Present, and Future. in *Journal of Biomedicine and Biotechnology* Vol. 2011 1-16 (2011).

35. Wang, D. & Shenk, T. Human cytomegalovirus UL131 open reading frame is required for epithelial cell tropism. *J Virol* **79**, 10330-8 (2005).

36. Dai, X. & Zhou, Z. Purification of Herpesvirus Virions and Capsids. in *Bio-Protocol* Vol. 4 2-6 (2014).

37. Peng, L., Ryazantsev, S., Sun, R. & Zhou, Z.H. Three-Dimensional Visualization of Gammaherpesvirus Life Cycle in Host Cells by Electron Tomography. in *Structure* Vol. 18 47-58 (2010).

38. Mastronarde, D.N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. in *Microscopy and Microanalysis* Vol. 9 1182-1183 (2003).

39. Kremer, J.R., Mastronarde, D.N. & McIntosh, J.R. Computer Visualization of Three-Dimensional Image Data Using IMOD. in *Journal of Structural Biology* Vol. 116 71-76 (1996).

40. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. in *Nature Methods* Vol. 14 290-296 (2017).

41. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. in *Nature Methods* Vol. 16 1153-1160 (Springer US, 2019).

42. Scheres, S.H.W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. in *Journal of Structural Biology* Vol. 180 519-530 (Elsevier Inc., 2012).

43. Rodrigues, C.H.M., Pires, D.E.V. & Ascher, D.B. mmCSM-PPI: predicting the effects of multiple point mutations on protein–protein interactions. in *Nucleic Acids Research* Vol. 49 W417-W424 (Oxford University Press, 2021).

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

45. Tandon, R. & Mocarski, E.S. Control of Cytoplasmic Maturation Events by Cytomegalovirus Tegument Protein pp150. *Journal of Virology* **82**, 9433-9444 (2008).

46. Li, Z. et al. Cryo-electron microscopy structures of capsids and in situ portals of DNA-devoid capsids of human cytomegalovirus. in *Nature Communications* Vol. 14 2025 (Springer US, 2023).

47. Tandon, R., Mocarski, E. & Conway, J. The A, B, Cs of Herpesvirus Capsids. in *Viruses* Vol. 7 899-914 (2015).

48. Liu, Y.-t., Strugatsky, D., Liu, W. & Zhou, Z.H. Structure of human cytomegalovirus virion reveals host tRNA binding to capsid-associated tegument protein pp150. *Nature Communications*, 1-9 (2021).

49. Zydek, M., Hagemeier, C. & Wiebusch, L. Cyclin-Dependent Kinase Activity Controls the Onset of the HCMV Lytic Cycle. *PLoS Pathogens* **6**, e1001096 (2010).

50. Ye, L. et al. Functional Profile of Human Cytomegalovirus Genes and Their Associated Diseases: A Review. *Frontiers in Microbiology* **11**, 1-13 (2020).

51. Gibson, W. Structure and Formation of the Cytomegalovirus Virion. in *Current Topics in Microbiology and Immunology*, Vol. 325 (eds. Shenk, T.E. & Stinski, M.F.) 187-204 (Springer, Berlin, 2008).

52. Trus, B.L., Gibson, W., Cheng, N. & Steven, A.C. Capsid Structure of Simian Cytomegalovirus from Cryoelectron Microscopy: Evidence for Tegument Attachment Sites. in *Journal of Virology* Vol. 73 4530-4530 (1999).

5. Chapter 5: Structures of Native Doublet Microtubules from Trichomonas vaginalis Reveal Parasite-Specific Proteins as Potential Drug Targets

Alexander Stevens[1,2,3,#], Saarang Kashyap[1, 2,#], Ethan H. Crofut[1,2], Shuqi E. Wang[1], Katherine A. Muratore[1], Patricia J. Johnson[1*], Z. Hong Zhou[1,2,3*]

[1]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA; [2]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA; [3]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA,

[#]Contributed Equally

*Corresponding authors

Keywords: cryoEM, doublet microtubule, *Trichomonas vaginalis,* trichomoniasis, antiparasitic

5.1. Abstract

Doublet microtubules (DMTs) are flagellar components required for the protist *Trichomonas vaginalis* (*Tv*) to swim through the human genitourinary tract to cause trichomoniasis, the most common non-viral sexually transmitted disease. Lack of DMT structures has prevented structure-guided drug design to manage *Tv* infection. Here, we determined the cryoEM structure of native *Tv*-DMTs, identifying 30 unique proteins, including 19 microtubule inner proteins and 9 microtubule outer proteins. While the A-tubule is simplistic compared to DMTs of other organisms, the B-tubule features specialized, parasite-specific proteins, such as *Tv*FAP40 and *Tv*FAP35 that form filaments near the inner and outer junctions, respectively, to stabilize DMTs and enable

Tv locomotion. Notably, a small molecule, assigned as IP6, is coordinated within a pocket of *Tv*FAP40 and has characteristics of a drug molecule. This first atomic model of the *Tv*-DMT highlights the diversity of eukaryotic motility machinery and provides a structural framework to inform rational design of therapeutics.

5.2. Introduction

Trichomonas vaginalis (*Tv*) is a flagellated, extracellular parasite of the human genitourinary tract and causative agent of trichomoniasis, the most common non-viral sexually transmitted infection (STI), with 250 million infections per annum and global prevalence over 3%[1-3]. *Tv* infection is linked to increased rates of preterm delivery and mortality, genitourinary cancers, and HIV transmission, with a disproportionate impact on women in developing countries[1-5]. Though the antibiotic metronidazole can be curative, concerns about its carcinogenic potential, increasing metronidazole resistance in *Tv*, and rising frequency of reinfection underscore the need for alternative precision therapies[1,6-8]. *Tv* relies on its four anterior and one membrane-bound, recurrent flagellum to propel itself through the genitourinary tract and attach to the mucosa of its human hosts, making the mechanisms driving locomotion potential therapeutic targets[9]. Unfortunately, no high-resolution structures related to *Tv* flagella are currently available, and even tubulin remains uncharacterized in *Tv* despite a putative role in antimicrobial resistance[10-12].

As observed in low-resolution, thin-section transmission electron microscopy (TEM) studies[13], the locomotive flagella originate from cytosolic basal bodies, and extend into the flagellar membrane with decorations along the microtubule filaments that stabilize the tubules and facilitate intraflagellar transport. The flagellar core, or axoneme, conforms to the canonical “9+2” arrangement wherein a central pair of singlet microtubules (MTs) is connected via radial spokes to nine surrounding doublet-microtubules (DMTs) which transduce force through the flagella (Fig. 1)[13,14]. Studies in other organisms revealed DMTs are coated with different combinations of microtubule inner and outer proteins (MIPs and MOPs) that facilitate assembly, stability, and function (Fig. 2a)[15-20].

Dozens of MIPs and MOPs have been identified across numerous studies of eukaryotic flagella, of which about half are conserved[15-19]. DMTs from multicellular eukaryotes incorporate more complex MIP arrangements, particularly along the highly variable ribbon protofilaments (PFs) that compose the inner and outer junctions (IJ and OJ) where the A- and B-tubules meet (Fig. 1a)[15-19]. In sperm flagella, filamentous tektin bundles near the ribbon PFs are thought to reinforce the long flagella as they swim through the viscous milieu of the genitourinary tract[21,22]. Though the *Tv* genome lacks tektin genes, the parasite swims through the same environment as sperm, coordinating its much shorter flagella into a distinct beating pattern[23]. Considering these apparent differences, it is unclear how the parasite propagates motion under these conditions and suggests a species-specific adaptation that may be exploited for therapeutic development.

Here, we leveraged mass spectrometry, cryogenic electron microscopy (cryoEM), and artificial intelligence to analyze the DMTs derived from *Tv* parasites and elucidate the structures of the proteins that compose them. Our structure contains 29 distinct proteins, including the α- and β-tubulin, 19 MIPs and 9 MOPs. Among these, we identified four *Tv*-specific proteins, including one bound to a ligand not observed in the DMTs of other organisms. This first structure of the *Tv* flagellum highlights remarkable simplicity in the species' DMT architecture compared to more complex organisms such as mammals, as well as other protists like *Tetrahymena thermophila*. Despite this simplicity, *Tv* can still traverse the same viscous environment as the more complex mammalian sperm, suggesting a key to parasite locomotion lies in the short list of *Tv*-DMT proteins.

5.3. Results

5.3.1. T. vaginalis DMTs feature both familiar and novel MIPs

We optimized a protocol to isolate DMTs from *T. vaginalis* and limit perturbations to the internal structures, then subjected them to single-particle analysis using cryoEM. The resultant cryoEM maps of the 48 nm repeat DMT had a global resolution of 4.2 Å and focused refinement improved local resolution to between 3.2 Å and 3.8 Å (Fig. 1b, Table S1). Reconstructions of the 16 nm and 96 nm repeat structures were resolved to 3.8 Å and 4.3 Å respectively. We also collected mass spectrometry data for our cryoEM sample to produce a library of potential *Tv*-DMT proteins and utilized cryoID to identify most likely candidates for certain map densities[24]. AlphaFold predicted structures served as initial models for atomic modeling of both conserved and species-specific cryoEM map densities[25,26]. From our structures we identified 30 unique proteins, including 19 MIPs, 9 MOPs and the α/β tubulin of *Tv* (Movie S1, Table S2). Of these proteins, 15 MIPs and all 9 MOPs are conserved between *Tv* and previous DMT structures, whereas 4 MIPs are novel. There are also 4 unassigned MIP and 3 MOP densities that appear to play an important role in DMT function, but for which we lacked sufficient resolution to model.

Consistent with their ~80% sequence identities, the atomic models of *Tv*'s α- and β-tubulin are nearly identical to those of their human homologs (Fig. 1c), including the region of β-tubulin where many antiparasitic, benzimidazole-derived drugs (BZs) bind (Fig. 1d). Previous studies in *Tv* suggest mutations to aromatic residues Tyr168 and Phe201 in β-tubulin confer BZ resistance[12,27,28]. Indeed, like human β-tubulin's Phe169 and Tyr202, *Tv* orients Tyr168 and Phe201 into the BZ binding pocket where they are stabilized by Aro-Met-Aro interactions with adjacent Met234 and Phe21 residues and sterically occlude BZ drugs like thiabendazole (TBZ) (Fig. 1d). To corroborate this, we performed docking experiments using AutoDock Vina and found TBZ docked β-tubulin produced large positive binding free energy values (∆G) (Fig. S2). By contrast a virtual β-tubulin Y168F, mutant, reflecting the putative TBZ binding site from susceptible organisms like *Ascaris lumbricoides*, exhibited a significantly lower binding free energy when TBZ was docked (Fig. S2). Interestingly, we observe the swapped positions of phenylalanine and tyrosine residues between human and *Tv* β-tubulin, which may help to explain species-specific sensitivity to different BZs.

Like other organisms, the α/β tubulin heterodimers polymerize and assemble into rings of 13 and 10 PFs that compose the A- and B-tubules respectively (Fig. 2b). Within the A-tubule, molecular rulers FAP53, FAP127, and Rib43a impose a 48 nm MIP periodicity and facilitate the organization of other MIPs like FAP67 and RIB72 (Fig. 2e-h). Consistent with studies in *T. thermophila*[18], FAP115 repeats every 32 nm and creates a mismatch with the 48 nm periodicity of the ruler proteins, leading to 96 nm periodicity (Fig. 2f). Interestingly, FAP141 from other organisms is replaced by the smaller *Tv*FAP12 which lashes FAP67 to the A-tubule lumen like the N-terminal helices of FAP53 and FAP127 (Fig. 2e)[15]. Along with the N-terminal helices of FAP53 and FAP127, *Tv*FAP12 passes into the B-tubule to maintain 16 nm a repeating crosslink between the A- and B-tubules as observed in FAP141 expressing organisms[29]. Unlike other species, the *Tv* ribbon PFs (A11-A13) that divide A- and B-tubules are sparsely decorated with A-tubule MIPs, suggesting alternative strategies of ribbon arc stabilization.

In the B-tubule lumen, we found assembly-related MIPs FAP45, CCDC173, enkurin, FAP77, FAP52, FAP20, and PACRGA/B that are conserved amongst other organisms. Interestingly, along the B-tubule side of the ribbon arc, we identified the filamentous MIPs *Tv*FAP35 and *Tv*FAP40, which run lengthwise along the A11 and A13 PFs respectively and may compensate for the dearth of MIPs along the ribbon arc in the A-lumen (Fig. 2g). Further, we observed globular MIPs that span PFs B3-B4 and B5-B6 and exhibit 96 nm periodicity (Fig. 2h). While the map resolution was insufficient to model these proteins, their interactions with neighboring ruler proteins like CCDC173 indicate an enforced periodicity of 96 nm, which is the largest repeating periodicity of any DMT MIP to date. Though we observed several novel proteins, the *Tv*-DMTs have the simplest MIP organization in the A-tubule with just 10 MIPs (nine identified and one unidentified) compared to the next simplest species of record, *C. reinhardtii*, with 22 A-tubule MIPs[15]. The comparatively simple MIP organization observed in *Tv* suggests the few novel MIPs may play a substantial role in flagellar function.

5.3.2. T. vaginalis microtubules reinforce the inner junction with species-specific protein

The DMT IJs of other organisms are typically composed of FAP52, enkurin/FAP106, PACRG isoforms (PACRGA and PACRGB), and FAP20, while *Tetrahymena* and mammalian DMTs include globular proteins atop FAP52 that mediate interactions with PF A13[18,22,30]. Interestingly, the *Tv*-DMT cryoEM map revealed the long, filamentous protein *Tv*FAP40, running atop PF A13 at the IJ which alters the topography of this important protofilament. *Tv*FAP40 monomers repeat every 16 nm and are arranged head-to-tail, where head-tail polarity corresponds to the – and +-ends of the DMT respectively (Fig, 3B-C). Each *Tv*FAP40 monomer consists of a globular N-terminal ‘head’-domain (residues 1-145) connected to a coiled-coil ‘tail’ (residues 149-361). The tail consists of 3 coiled-helices (α7-9) where a proline-rich kink connects α7 and α8 while a 180° turn at the linker between α8 and α9 forms the ‘tip’ of the tail. The tip includes α8 and neighboring residues of α9 (residues 246-282), with both a polar face oriented towards the MT and a hydrophobic face oriented towards a neighboring *Tv*FAP40 monomer (Fig. 3c-d). As the kink reaches into the cleft between tubulin heterodimers, α8 is brought into close contact with tubulin and establishes electrostatic interactions. The kink also offsets α7 from α8, creating an overhang to bind the head of a neighboring *Tv*FAP40 monomer which may help stabilize the interaction (Fig. 3d).

*Tv*FAP40’s unique location along PF A13 has not been seen in other MIPs and alters conserved MIP interactions at the inner junction. In other organisms, the PACRGB N-terminus binds the groove between A11 and A13, but in our structure, *Tv*FAP40 blocks this groove and replaces A13 as the binding partner. Additionally, the *Tv*FAP40 C-terminus hooks around α2 of enkurin, where *Tv*FAP40’s C-terminal tyrosine (Tyr377) participates in hydrophobic interactions with adjacent aromatic residues from enkurin (Tyr155 and Trp156) (Fig. 3f). This C-terminal hook acts in concert with the *Tv*FAP40 head that binds the other side of enkurin α2 and restricts it such that the bottom end of the helix is 1 nm closer to A13 than in other structures.

5.3.3. Tv-specific FAP40 head domain binds a stabilizing ligand

In addition to binding neighboring monomers, *Tv*FAP40 incorporates a unique ligand binding pocket. Our cryoEM maps indicate the *Tv*FAP40 head-domain binds a six-pointed, star-shaped ligand, and our atomic model indicates this pocket is positively charged (Fig. 4a-c, Movie S2). Indeed, the putative binding site features seven positively charged side chains oriented towards the points of the star, and density from a likely metal cation coordinated by adjacent arginine residues (Fig. 4c), which suggests negatively charged functional groups (Fig. 4c-g). Sequence and structural homology searches within UniProt or the RCSB protein database could not identify similar binding sites[31,32], but the high local resolution of our map in this pocket revealed the stereochemistry of the functional groups at the points of the star, consistent with bonding to a non-planar six-membered ring. Together these features suggested an inositol polyphosphate, in this case inositol hexakisphosphate (IP6), which was a good fit for the map density (Fig. 4c).

IP6 is an abundant cellular polyanion known to stabilize positive interfaces such as the pore of HIV nucleocapsids[33], a trait that may be useful to DMT reinforcing proteins. To confirm whether IP6 was a reasonable ligand assignment, we carried out *in silico* molecular docking using Swissdb's AutoDock Vina webserver[34-37]. Restricting the docked ligand to the observed binding pocket resulted in docked arrangements consistent with the observed ligand density, and binding energies (ΔG) of -3.5 kcal/mol or less (Fig. 4h). The docking experiments suggested interactions with the same arginine and lysine residues as the real ligand structure in the binding pocket. These results support the notion that IP6 acts as a ligand within *Tv*FAP40 and may stabilize the head to reinforce its interactions with enkurin and the tail of its neighboring monomer and microtubule PF. Interestingly, in zebrafish embryos the IP6-producing enzyme (Ipk1) was found to localize to basal bodies of cells, and Ipk1 knockdown disrupted cilia growth and beating[38,39]. Combined with our structures, these studies point to an uncharacterized role for IP6 in flagellar function and stability.

5.3.4. TvFAP35 secures FAP77 and buttresses the outer junction of Tv-DMTs

Directly above *Tv*FAP40 in the B-tubule we identified a novel, filamentous density along the ribbon PF A11 as *Tv*FAP35, another *Tv*-specific protein (Fig. 5a-c). In the B-tubule, *Tv*FAP35 repeats every 16 nm in a head-to-tail fashion with the heads and tails oriented to the – and +-ends respectively (Fig. 5b). The *Tv*FAP35 tail domain has the same 'kinked-coiled-coil' fold as *Tv*FAP40 including the tips that mediate MT binding and dimerization (Fig. 5c-f). The head-domain of *Tv*FAP35 differs from *Tv*FAP40, as it includes only a flexible N-terminus and helix-turn-helix (Fig. 5c), as opposed to the six helices found in the *Tv*FAP40 head (Fig. 3c).

The position of *Tv*FAP35 along A11 is similar to that of tektin-like protein 1 (TEKTL1) which is thought to reinforce OJ stability during flagellar beating in sperm DMTs[22]. Further, the coiled-coil structure of *Tv*FAP35 resembles the 3-helix bundle architecture of TEKTL1. Unlike TEKTL1, the coiled coils of *Tv*FAP35 include a proline-rich kink that occupies the cleft between tubulin heterodimers. As PFs bend, gaps form at the interface between tubulin heterodimers[40], and the *Tv*FAP35 kink may create stress relief points along A11 by acting as a flexible linker which accommodates bending. Thus, like TEKTL1, the coiled-coils of *Tv*FAP35 may provide structural stability to the DMT while the kink allows bending and greater flexibility. *Tv*FAP35 also interacts with FAP77, a MIP that aids in B-tubule formation and tethers complete A- and B-tubules together at the OJ (Fig. 5g & h)[18]. The FAP77 helix-turn-helix motif (residues 140-164) is braced to PF A11 via electrostatic interactions with the coiled-coil of *Tv*FAP35 (Fig. 5b, c, and f). Further, the tail-domain of *Tv*FAP35 passes over residues 238-246 of *Tv*FAP77 which run between a cleft of the A11 PF and reinforces *Tv*FAP77's association to A11 (Fig. 5g). These observations suggest that, like *Tv*FAP40, *Tv*FAP35 plays an integral role in the stabilization of the ribbon PFs and their associated MIPs. Additionally, because FAP77 is implicated in B-tubule assembly[18], the interactions of *Tv*FAP35 with FAP77 and A11 suggest that *Tv*FAP35 may also contribute to DMT assembly.

5.3.5. Novel Tv proteins share an ancient MT binding motif

Upon comparison, we noticed both *Tv*FAP35 and *Tv*FAP40 have kinked-coiled-coils composed of three helices (α1-3, Fig. S2) with similar lengths, dimerization domains, and MT binding motifs (Fig. 3-4). This similarity prompted us to search for homologous proteins via amino acid sequence alignment, but this returned few candidates. Interestingly, the coiled-coils of *Tv*FAP35 and *Tv*FAP40 share just 23% identity despite similar folds. We next turned to structural alignment using *FoldSeek*[41], and identified numerous structural homologs. After curating homolog candidates by removing those without kinked-coiled-coil domains or with TM-scores below 0.4, 31 homolog candidates were selected for further comparison. Initial analysis revealed all kinked-coiled-coil-containing homologs belonged to the group *Bikonta*, which includes many protists, and excludes animals, fungi and amoebozoans.

Six kinked-coiled-coil homologs from protists, representing different clades were selected for multiple sequence alignment with *Tv*FAP35 and *Tv*FAP40 (Fig. S3a) which revealed several conserved residues from the dimerization and MT binding domains. Based on the *Tv*FAP35 sequence, the dimerization domains include hydrophobic residues at Val213 and aromatic residues at Tyr103 and Tyr218, which form hydrophobic interfaces between neighboring monomers (Figs. S2g). Further, the MT binding motif on α2 has a high proportion of charged residues, which are likely important in tubulin binding (Fig. S3f). Outside of the dimerization and MT binding motifs, the kinked-coiled-coils exhibit an average sequence conservation of ~20% which may be necessary to accommodate different MIPs, as observed in our novel proteins (Fig. 3 and 5).

5.3.6. T. vaginalis microtubule outer proteins exhibit 8 nm periodicities

Considering the marked simplicity of the *Tv*-DMT MIP arrangement, we expected to find comparably simple MOP organization. Along the A-tubule we observed both the canonical N-DRC and radial spoke complexes that mediate inter-axoneme connections and flagellar bending (Fig. 6a-c). We see that, like other DMT structures, the axoneme-related proteins exhibit 96 nm periodicity enforced by the molecular ruler proteins CCDC39 and CCDC40, which coil their way between PFs A3 and A2 (Fig. 6b-e). Besides N-DRC and radial spoke proteins, a diverse arrangement of filamentous MOPs occupies the clefts and the surface of several PFs. Previous DMT structures from other species found the shortest MOP periodicity to be 24 nm[18,22]. *Tv*MOP1 is 24 nm repeating MOP that arranges head-to-tail in the furrows between PFs A3, A4, B8, B9, & B10 and contacts the flexible C-terminal tails of α and β tubulin in the B-tubule (Fig. 6f). Interestingly, though the exterior of the outer junction is sparsely decorated in DMTs structures from other organisms[15,22], we found this area to contain a large filamentous protein that repeats every 8 nm and a smaller filament that runs in a zig-zag beneath it and between A10 and B1 (Fig. 6g). The large protein density fashions an ankyrin-like domain seated atop a large coiled-coil domain, which spans the gap between PF A9 and B1 (Fig. 6g).

Due to limited local resolution, we were unable to confidently assign the identities of these proteins and instead dubbed them *Tv* outer junction microtubule outer protein 1 and 2 (*Tv*OJMOP1 and *Tv*OJMOP2) for the large and zig-zag MOPs respectively. *Tv*OJMOP1 exhibits an 8 nm periodicity like that of tubulin heterodimers, an unusual repeat length amongst DMT MOPs, and crosslinks PF B1 to A9 and A10 (Fig. 6g). *Tv*OJMOP1 was observed in only 1/5 of particles, suggesting that some may have been lost during DMT isolation or that *Tv*OJMOP1 localizes to certain regions of the axoneme. Exhaustive search through AlphaFold predicted structures from our proteomic data using a strategy similar to that of DomainFit[42] yielded the following 5 candidate proteins which contain both ankyrin and coiled-coil domains: TVAGG3_0305310, TVAGG3_0421180, TVAGG3_0431750, TVAGG3_0596110, and TVAGG3_0415080. However, none of these candidates could fully account for the observed density, and so it remains unclear if *Tv*OJMOP1 is composed of one or more of these proteins. Recent work in *C. reinhardtii* has demonstrated that anterograde intraflagellar transport (IFT) brings IFT-B complexes directly over the area where *Tv*OJMOP localizes (Fig. 6h)[43]. However, as components are often lost during DMT isolation they are unlikely candidates. *As Tv*OJMOP1 features an ankyrin domain oriented towards the would-be IFT-B cargo (Fig. 6g-h), it may interact with TPR-rich proteins of IFT-B to stabilize the cargo. Additionally, others have documented the tendency for cytoplasmic dynein motors to jump between PFs[44]. *Tv*OJMOP1 may help segregate transport tracks and keep the dynein motors on their preferred A-tubule PFs.

5.4. Discussion

T. vaginalis pathogenesis relies on the parasites' locomotive flagella to establish infection and spread between human hosts[23,45]. This study reports the first high-resolution structure of *Tv* flagellar doublet microtubules, elucidating their molecular composition, architectural arrangement, atomic structures, and small molecule ligands. In addition to the first atomic structures of the *Tv* tubulin subunits comprising the DMTs, we have identified 19 MIPs and 9 MOPs distributed across the A- and B-tubules. These MIPs and MOPs mediate *Tv*-DMT function in the flagella with several novel proteins. As the first near-atomic structure of flagellar microtubules in the major human parasite *Trichomonas vaginalis*, our results provide a structural framework to understand the parasite's distinct locomotion, offer insights into antibiotic drug resistance, and identify new targets for precision medicine.

With a relatively short list of both conserved and *Tv*-specific MIPs, the *Tv*-DMT is perhaps the simplest among known DMT structures. Notably, the *Tv* A-tubule fashions the fewest MIPs of any characterized organism (Fig. 2). Among them, the ruler proteins Rib43a, FAP53, and FAP127 are conserved, but lack many of the interacting partners of their homologs in other species, such as mammalian sperm, that traverse the same environment. The sparsity of A-tubule ribbon proteins in *Tv* suggests these proteins are less essential for locomotion in the human genitourinary tract, which contrasts with the complex MIP arrangement of sperm-specific proteins and tektin bundles seen in the A-tubule of mammalian sperm[22,30]. While *Tv* and sperm exhibit distinct flagellar beating patterns, the sinusoidal beating pattern of the recurrent flagella in *Tv* suggests the additional MIP complexity observed in sperm is not essential to this style of beating. However, human sperm swim five times faster than *Tv* and must propagate beating over longer flagella, so sperms' complex A-tubule MIP arrangement may facilitate rapid propulsion through their viscous environment[9,46].

Remarkable specialization is observed in the B-tubule, where several novel proteins reinforce the ribbon arc in a manner similar to tektin bundles from other organisms[15]. Like tektin, the *Tv*FAP35 and *Tv*FAP40 proteins exhibit 16 nm periodicity and similarly interact with other MIPs along their respective protofilaments (Fig. 3 and 5). However, unlike tektin, *Tv*FAP35 and *Tv*FAP40 have variable head domains which seemingly confer different functionalities. To this end, the positively charged pocket of *Tv*FAP40 putatively binds an IP6 pocket factor (Fig. 4). In HIV, IP6 acts as a pocket factor to stabilize the nucleocapsid lattice[33]. Considering *Tv*FAP40 likely plays a stabilizing role along the ribbon arc, IP6 binding may augment that stabilization by reinforcing the interactions between the monomers at the head-tail interface. Binding abundant biomolecules is a common strategy amongst pathogens, particularly viruses[47-49], but this is the first instance such pocket factors have been documented in DMTs.

The *Tv*-specific MIPs and MOPs are particularly significant in light of their role in propagating the pathogenesis of the most widespread non-viral STI[3]. Specialization differentiates the parasite's DMTs from those of other organisms, including their human hosts, thus drugs targeting these specialized components would have minimal toxicity. For instance, the unique cofactor binding pocket found in the *Tv*-specific *Tv*FAP40 protein has a structure with no known homologs and appears to specifically bind IP6 (Fig. 4). This pocket could be targeted by antimicrobial compounds to destabilize parasite DMTs with limited off-target effects. Notably, the only homologous proteins to *Tv*FAP40 belonged to other Bikonts and include other human-borne parasites like *T. brucei* and *Leishmania donovani*, that may incorporate similar species-specific proteins (Fig. S3). While this study represents the first of its kind on the DMT from a human-borne parasite, it has demonstrated the power of *in situ* cryoEM over other structural or *in silico* methods, to open new avenues for rational drug design. Together, our findings provide a basis to explore the contribution of microtubule-associated proteins to the unique aspects that allow *T. vaginalis* to swim through the human genitourinary tract, and the diversity of eukaryotic motility in general. Moreover, the atomic details revealed in species-specific proteins and bound small molecules can inform the rational design of therapeutics.

5.5. Author contributions

Z.H.Z and P.J. designed and supervised the project. K.A.M., S.E.W., and A.S. prepared samples. S.E.W. conducted mass spectrometry work. A.S. and S.K. performed cryoEM imaging and prepared 3D reconstructions. Under the guidance of Z.H.Z., A.S., S.K., and E.H.C. built the atomic models, interpreted the structures, and made the figures and wrote the paper; all authors reviewed and approved the paper.

5.6. Conflict of interest

The authors declare no competing interests.

5.7. Acknowledgements

We thank Yanxiang Cui for his preliminary work on this project, along with Xian Xia and Kent Hill for their expert advice in the development of this work. This work was supported in part by grants from the National Institutes of Health (R01GM071940 to Z.H.Z. and R01AI103182/R33AI119721 to P.J.J.). K.A.M. and A.S. received support from NIH Ruth L. Kirschtein National Research Service Award AI007323. We acknowledge the use of resources at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057, 1S10OD018111) and NSF (DBI-1338135 and DMR-1548924). We acknowledge support from the UCLA AIDS Institute, the James B. Pendleton Charitable Trust, and the McCarthy Family Foundation.

5.8. Methods and Data Availability

5.8.1. Cell culture

T. vaginalis strain G3 was cultured in Diamond's modified trypticase-yeast extract-maltose (TYM) medium supplemented with 10% horse serum (Sigma-Aldrich), 10 U/mL penicillin, 10 µg/ml streptomycin (Gibco), 180 µM ferrous ammonium sulfate, and 28 µM sulfosalicylic acid[50]. 2L of parasites, grown at 37 °C and passaged daily, were harvested by centrifugation, and washed twice with phosphate-buffered saline and pelleted at low speed. Cells were resuspended in 50 mL lysis buffer (2% IGEPAL CA-630, 2% Triton X-100, 10% glycerol, 10 mM Tris, 2 mM EDTA, 150 mM KCl, 2 mM MgSO4, 1 mM dithiothreitol [DTT], 1× Halt protease inhibitors [pH 7.4]) and lysed in a Stansted cell disrupter, operated at 30 lb/in^2 front pressure and 12 lb/in^2 back pressure.

Cytoskeletal elements were harvested similar to what has been previously described[51]. Lysates were recovered and maintained at 4 °C for all subsequent steps. Nuclei were removed via low-speed centrifugation (1000 x g) for 10 mins to generate pellet 1 (P1) and lysate 1 (L1). L1 was centrifuged (10,000 x g for 40 mins) to pellet cytoskeletal components into P2 and L2. Cytoskeleton pellets (P2) were resuspended in 1 mL low salt (LS) buffer (150 mM NaCl, 50 mM Tris, 2 mM $MgCl_2$, 1 mM DTT, 1× complete protease inhibitor (Sigma-Aldrich) [pH 7.4]) and centrifuged at low speed (1000 x g, 10 mins) to pellet cellular debris into P3. The resulting lysate (L3) was placed a sucrose cushion (30% w/v sucrose in LS buffer) and centrifuged at low speed (1,800 x g, 10 mins). The supernatant atop the cushion was collected and resuspended in 1 mL LS buffer prior to centrifugation (5,000 x g 15 mins) to pellet larger cytoskeletal components (P4). The lysate was finally centrifuged at high speed (16,600 x g, 40 minutes) to pellet axoneme related cytoskeletal elements (P5). The P5 was then resuspended in minimal volume of LS buffer supplemented with 5 mM ATP and left at RT for 1 hour.

5.8.2. In-solution digestion, Mass Spectrometry Data Acquisition and Analysis

T. vaginalis cytoskeleton pellets P4 and P5 resuspended in low salt (LS) buffer were mixed with 4× volume of ice-cold acetone and kept at -20°C for 2 h. The mixtures were centrifuged at 4°C with 14,000 rpm for 15 min and supernatants discarded. The air-dried pellets were fully dissolved in 8 M Urea in 100 mM Tris-HCl (pH 8) at 56 °C and the proteins reduced with 10 mM Tris(2-carboxyethyl) Phosphine for 1 h at 56 °C. The reduced proteins were then alkylated with 40 mM iodoacetamide for 30 min in dark at room temperature and the reaction was quenched with Dithiothreitol at a final concentration of 10 mM. The alkylated samples were subsequently diluted with 7× volume of 100 mM Tris-HCl pH 8, to 1M Urea concentration. To generate peptides, Pierce Trypsin Protease (Thermo Fisher Scientific) was added to the samples and the ratio of trypsin:protein was 1:20 (w/w). The digestion reaction was incubated at 37 °C overnight, and the residue detergents in the protein samples were removed using a HiPPR Detergent Removal Spin Column Kit (Thermo Fisher Scientific) on the next day. Prior to the mass spectrometry assay, the samples were desalted with Pierce C18 Spin Columns (Thermo Fisher Scientific) and lyophilized.

Three biological replicates were prepared and trypsin-digested following the steps above for fractions P4 and P5, respectively. The lyophilized protein pellets were dissolved in sample buffer (3% Acetonitrile with 0.1% formic acid) and ~1.0 μg protein from each sample was injected to an ultimate 3000 nano LC, which was equipped with a 75μm x 2 cm trap column packed with C18 3μm bulk resins (Acclaim PepMap 100, Thermo Fisher Scientific) and a 75µm x 15 cm analytical column with C18 2µm resins (Acclaim PepMap RSLC, Thermo Fisher Scientific). The nanoLC gradient was 3−35% solvent B (A = H2O with 0.1% formic acid; B = acetonitrile with 0.1% formic acid) over 40 min and from 35% to 85% solvent B in 5 min at flow rate 300 nL/min. The nannoLC was coupled with a Q Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA), operated with Data Dependent Acquisition mode (DDA) with inclusion list for the target peptides. The ESI voltage was set at 1.9 kV, and the capillary temperature was set at 275 °C. Full spectra (m/z 350 - 2000) were acquired in profile mode with resolution 70,000 at m/z 200 with an automated gain control (AGC) target of 3 × 106. The most abundance 15 ions were subjected to fragmentation by higher-energy collisional dissociation (HCD) with normalized collisional energy of 25. MS/MS spectra were acquired in centroid mode with resolution 17,500 at m/z 200. The AGC target for fragment ions is set at 2×10^4 with maximum injection time of 50 ms. Charge states 1, 7, 8, and unassigned were excluded from tandem MS experiments. Dynamic exclusion was set at 45.0 s.

The raw data was searched against total *T. vaginalis* annotated proteins (version 63) downloaded from TrichDB, using ProteomeDiscoverer 2.5. Following parameters were set: precursor mass tolerance ±10 ppm, fragment mass tolerance ±0.02 Th for HCD, up to two miscleavages by semi trypsin, methionine oxidation as variable modification, and cysteine carbamidomethylation as static modification. Protein abundance was quantified using Top 3 approach, i.e., the sum of the three most intense peptides coming from the same protein. Only proteins that were detected in all three replicates of P4 or P5 were included for further analyses, which resulted in a total of 386 and 311 proteins identified from P4 and P5, respectively. Among these common proteins, contaminants that are obviously not cytoskeletal proteins were identified from the datasets based on the GO terms and function annotations. For instance, proteins annotated as histone, kinase or DNA binding proteins or proteins located in subcellular compartments e.g., translational apparatus, nucleus, plasma membrane, were removed from the datasets. As a consequence, the numbers of putative cytoskeletal proteins identified in P4 and P5 were reduced to 303 and 239, respectively. The union of dataset P4 and P5, which consists of 371 distinct proteins, represent the entire cytoskeletal proteome of *T. vaginalis* identified by this study. DeepCoil 2.0 program was employed to predict coiled-coil domains (ccds) from the 371 putative cytoskeletal proteins based on protein sequence[52]. Three indices, i.e., number of ccds within each protein, the average length of ccds in each protein and percentage of total protein length occupied by ccds were calculated based on the output of DeepCoil 2.0. In addition to the cytoskeletal proteome in this study, the presence of ccds was also investigated for the hydrogenosome proteome of *T. vaginalis* and a randomly picked *T. vaginalis* protein dataset [53].

5.8.3. CryoEM sample preparation and image aquisition

To prepare DMTs for single particle analysis, 2.5 µL of DMT lysate was applied to glow discharged carbon holey grids (R2/1) (Ted Pella) and incubated on the grid for 1 minute prior to blotting and plunge freezing into a 50:50 mixture of liquid ethane and propane using a Vitrobot Mark IV (Thermo-Fisher). Flash frozen grids were stored under liquid nitrogen until cryoEM imaging.

Dose fractionated cryoEM movies were recorded on a K3 direct electron detector (Gatan) equipped Titan Krios electron microscope (FEI/Thermo-Fisher) fitted with a Gatan Imaging Filter (GIF) and operated at 300 keV. Movies were recorded at a nominal magnification of 81,000 x and calibrated pixel size of 0.55 Å at the specimen level, operated in super resolution mode. Using SerialEM[54], 30,834 movies were recorded with a cumulative electron dose of ~ 45 e^-/A^2.

5.8.4. CryoEM image processing and 3-dimensional reconstruction

Movie frame alignment and motion correction were performed in CryoSPARC[55], to generate cryoEM micrographs from each movie. Patch-aligned and dose weighted micrographs were binned 2X to improve processing speeds and transferred for processing in Relion 4.0 and Topaz automated particle picking, using the filament option "-f" incorporated by Scheres and colleagues[56-58]. Picked particles coordinates were extracted in Relion using the particle extract job with helical option enabled to extract particles every 8.2 nm along the picked filaments. The extracted particles were transferred back to CryoSPARC for further analysis and 3D reconstruction. 942,986 DMT particles were initially screened for quality using 2D classification job type, and those classes with good features were chosen for further data processing leaving 868,683 particles. Initial 3D reconstructions were made using 2X binned particles to expedite data processing. CryoSPARC's Helix refine job type was used to refine the DMT particles and prevent particles from the same filament from being placed in different half-sets during refinment. With half sets determined, the particles were then subjected to non-uniform refinement to yield an initial DMT reconstruction based on the 8.2 nm repeating tubulin heterodimer organization.

We next carried out focused classification and refinements as described previously[15], using CryoSPARC. Briefly, cylindrical masks over MIPS or MOPs with known periodicities were used to relax the 16, 48, and 96 nm periodicity from the initial 8.2 nm repeating DMT structure in stepwise fashion. To improve local resolutions, we performed focused local refinements wherein cylindrical masks were placed over specific protofilaments so that CryoSPARC could be used to align those protofilaments and their MIP and MOP features. This resulted in 8, 16, 48, and 96 nm reconstructions with 3.8, 3.8, 4.2, and 4.3 Å global resolutions, respectively. Local resolutions were improved using the local refinement job types in CryoSPARC, with maps over the regions of interest.

5.8.5. Atomic Modeling and Docking

The tubulin models were built using AlphaFold predicted models of α- and β-tubulin and using molecular dynamics flexible fitting software in UCSF ChimeraX[59,60]. To model MIPs, homologs from other organisms with existing structures roughly fit into our DMT maps before using NCBI's basic local alignment search tool (BLAST) to identify homologs in *Tv* and confirmed their identity using our mass spectrometry data[61]*.* For densities lacking homologous proteins, initial models were built using DeepTracer[62], followed by refinement in Coot[63].

The identities of unknown densities were confirmed using automated building in ModelAngelo and standard Protein BLAST of the predicted amino acid sequences against TrichDB database[61,64,65]. Alternatively, or often in combination with ModelAngelo predicted models, cryoID was used to identify the most likely candidates for cryoEM densities[24]. Further attempts to fit proteins in low resolution regions were made using a strategy similar to that of the DomainFit software package[42]. Briefly, visual inspection of AlphaFold predicted structures also aided in matching of candidates with map density shapes to assess potential matches. Models were fit using Coot and ISOLDE as described previously [59,66] and refined using Phenix Real Space Refinement[67].

Docking of thiabendazole (SMILES: C1=CC=C2C(=C1)NC(=N2)C3=CSC=N3) into *Tv* β-tubulin was performed using SwissDock tools and AutoDock Vina version 1.2.0[35,36]. was mutated to phenylalanine using the "swap amino acid" function in UCSF ChimeraX[60]. The box center was placed at 361 – 472 – 277 for each run with dimensions 10 – 10 – 15 and sampling exhaustivity set to the default value of 4.

We used the same software as above for docking IP6 (SMILES: C1(C(C(C(C(C1OP(=O)(O)O)OP(=O)(O)O)OP(=O)(O)O)OP(=O)(O)O)OP(=O)(O)O)OP(=O)(O)O) into *Tv*FAP40, except that the box center was placed at 393 – 277 – 260 and box size was left at default of 20 – 20 – 20 with sampling exhaustivity of 4. Grid box size was chosen to constrain ligands to putative binding sites from previous studies (thiabedazole) or observed localization (IP6) [27]. Structure visualization and figure preparation were done with UCSF ChimeraX[60] and Adobe illustrator, respectively.

5.8.6. Data availability

The cryoEM maps of the 16, 48, and 96 nm repeats have been deposited at the Electron Microscopy Data Bank (EMDB) and can be found under accession numbers EMD-46642, EMD-46643 and EMD-46636 respectively. The locally refined cryoEM map of the 48 nm repeat was deposited at the EMDB under the accession number EMD-46580 and coordinates for the complete atomic models under the accession number 9D5N. Focused maps for the 8 nm *Tv*OJMOP1, 16 nm *Tv*-FAP40 and the 96 nm NDRC focused refinements have been deposited at the EMDB under accession numbers EMD-46634, EMD-46633, and EMD-46635, respectively.

5.9. Figures

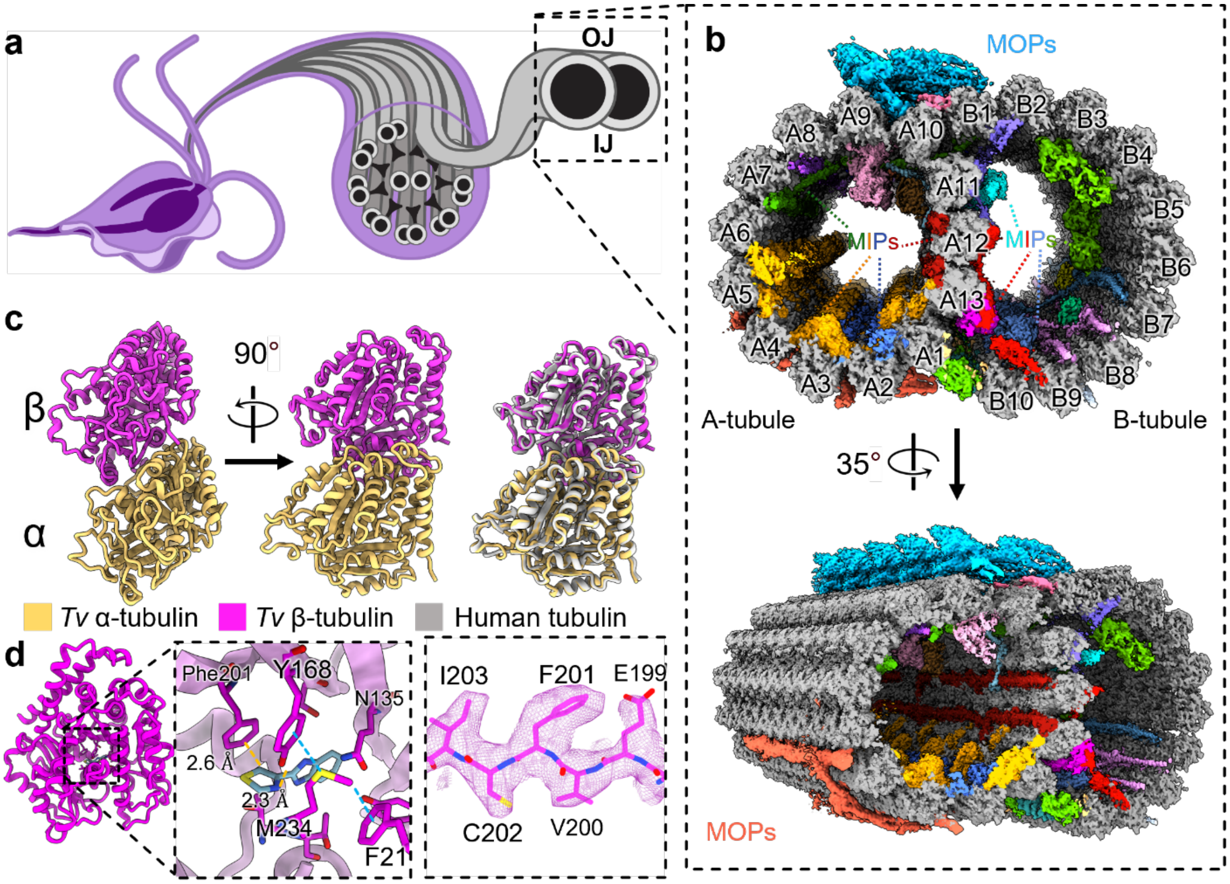


Figure 5- 1. CryoEM reconstruction of the doublet microtubules from Tv. (a) Diagram of axoneme from the flagella of *T. vaginalis.* (**b**) Cross-section of *Tv*-DMTs with microtubule inner proteins (MIPs) and microtubule outer proteins (MOPs) indicated with various colors. A- and B-tubules, as well as protofilaments, are labeled. (**c**) Atomic models of α and β tubulin, superimposed with human tubulin (right). **(d)** Alternate view of *Tv* β tubulin (left) and docked thiabendazole molecule (blue) fit into putative binding site with adjacent residues shown (right) with cryoEM map density. **IJ:** inner junction; **OJ:** outer junction.

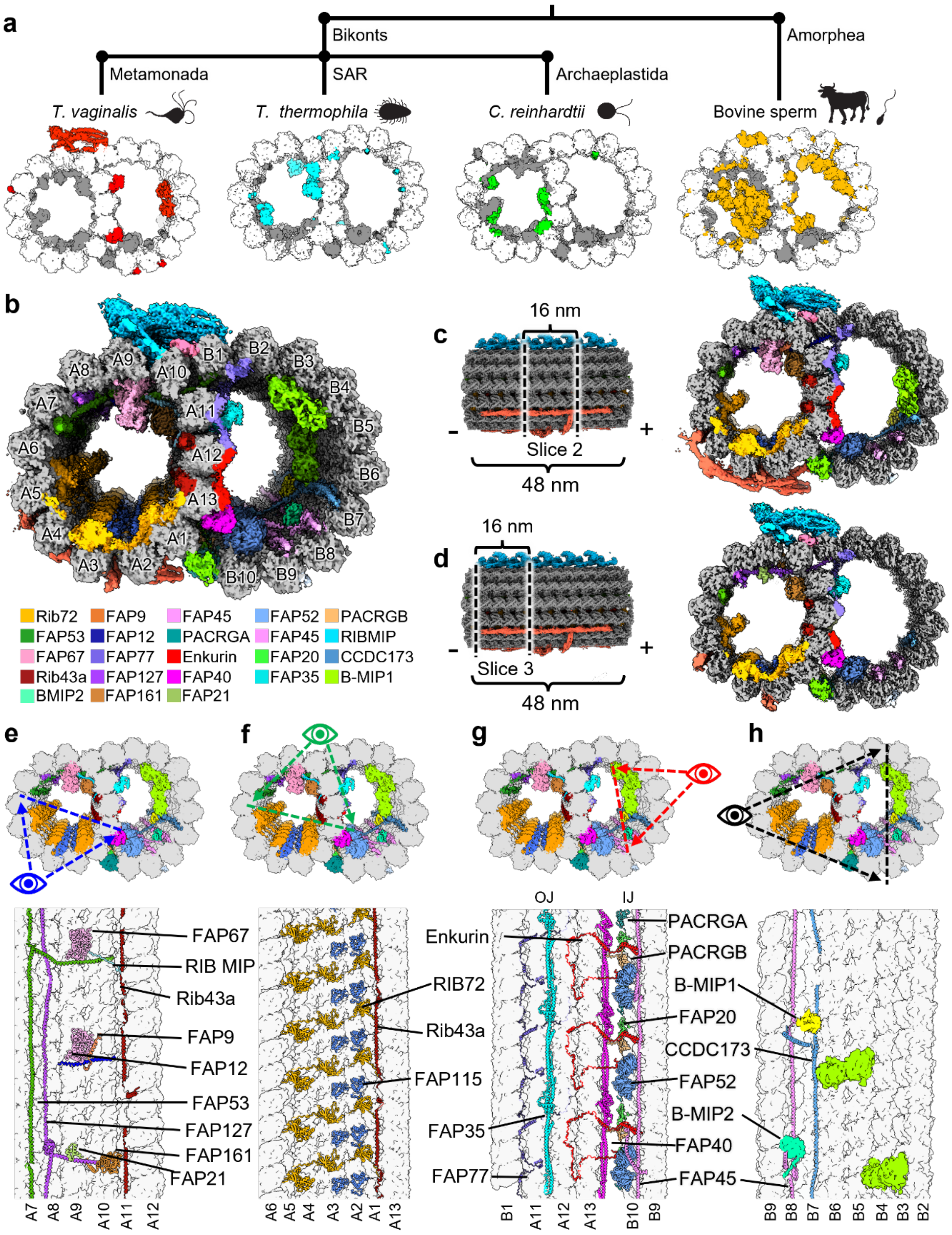


Figure 5- 2. Tv-DMTs reveal conserved and novel MIPs. (**a**) Phylogeny tree illustrating proposed divergence between Bikonts and Amorphea (top), with example

organisms from these branches and accompanying DMTs (bottom) with tubulin (white), conserved flagella associated proteins (FAPs) (grey), and species-specific FAPs (colored) (**b**) Cross-sectional view of cryoEM reconstruction of 48 nm repeat with MIP protein densities colored to demonstrate arrangement. (**c and d**) Cross-sectional view of DMTs from the 48 nm repeat map, shown as different 16 nm long sections throughout the DMT. (**e-h**) Cross-sectional views of *Tv*DMTs from different perspectives to illustrate MIP arrangement and periodicity.

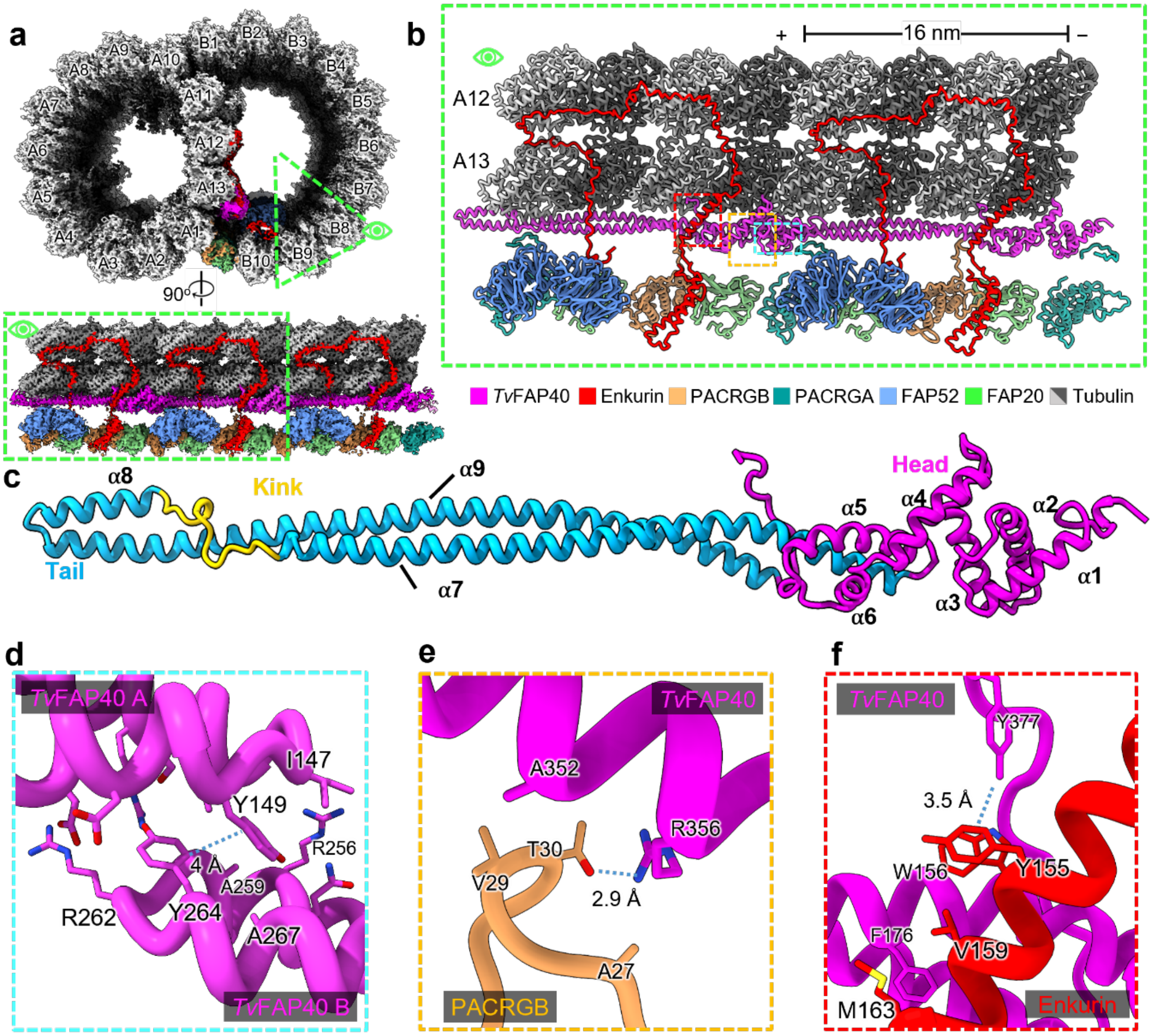


Figure 5- 3. TvFAP40 alters the inner junction arrangement in parasite DMTs. (**a**) Cross-sectional view of cryoEM reconstruction of 16nnm repeat with protofilaments labeled and proteins near inner junction colored (top) and cutaway view of region of interest (bottom). (**b**) View of atomic models built from map in **a**. (**c**) atomic model of *Tv*FAP40 colored by domain. (**d**) Zoomed-in view of dimerization domain between two *Tv*FAP40 monomers (labeled *Tv*FAP40 A and B). (**e** and **f**) Close-up view of interaction between PACRGB (tan) and *Tv*FAP40 and Enkurin (red), with residues shown to highlight interactions.

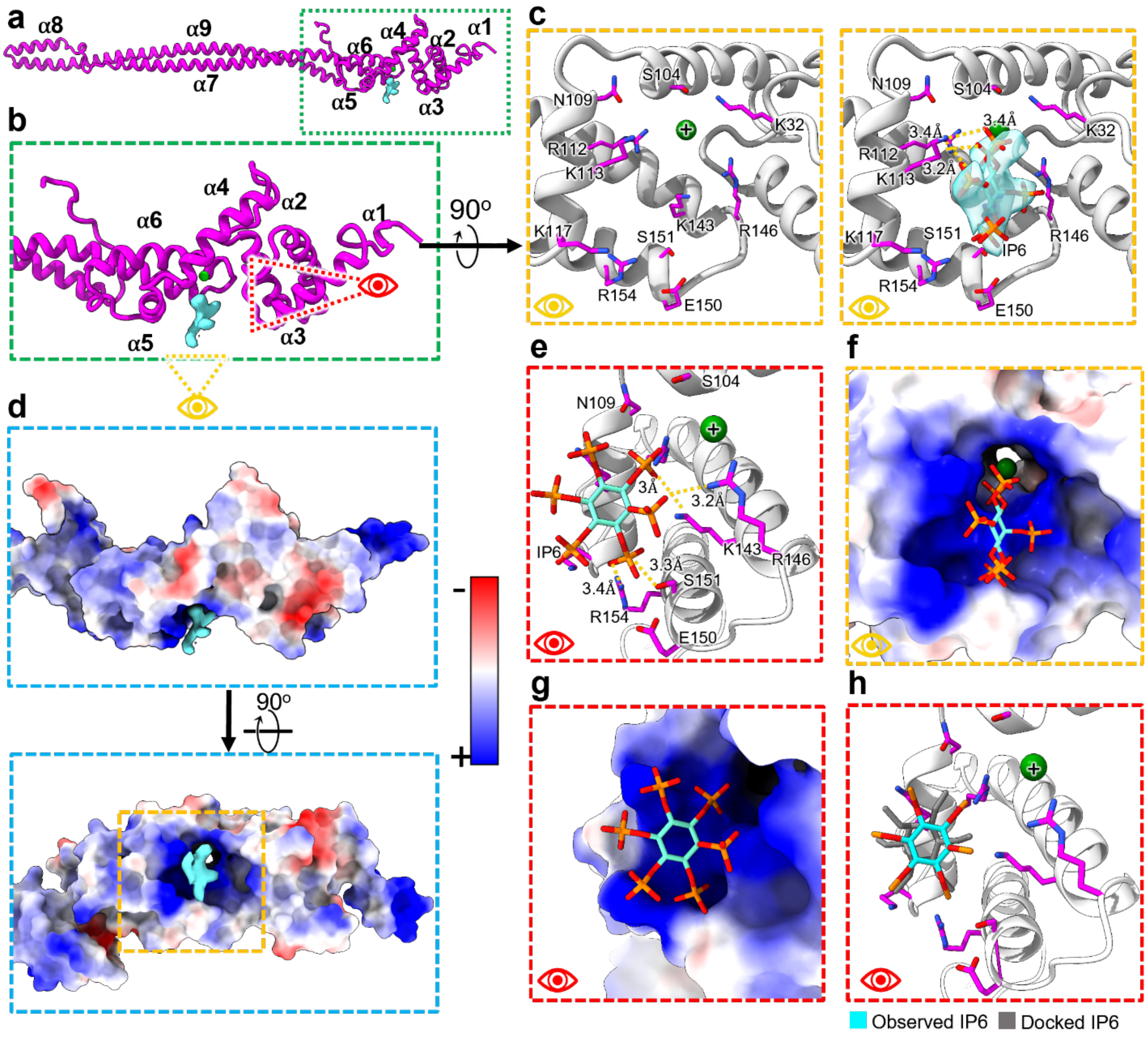


Figure 5- 4. TvFAP40 binds IP6 in a positively charged pocket. (**a**) Atomic model of *Tv*FAP40 with (**b**) zoomed-in view of the head domain and (**c**) perspectives of the putative IP6 binding site with (right) and without (left) IP6 fit into the cryoEM map. (**d**) Coulombic potential map of head domain from B (top) and rotated (bottom) views with blue and red indicating positive and negative coulombic potentials respectively. (**e**) Side-view of IP6 in binding pocket with adjacent residues shown. (**f-g**) Views from C and E

shown with electrostatic potential maps of Tv-FAP40. (**h**) Comparison of observed IP6 binding site and docked IP6.

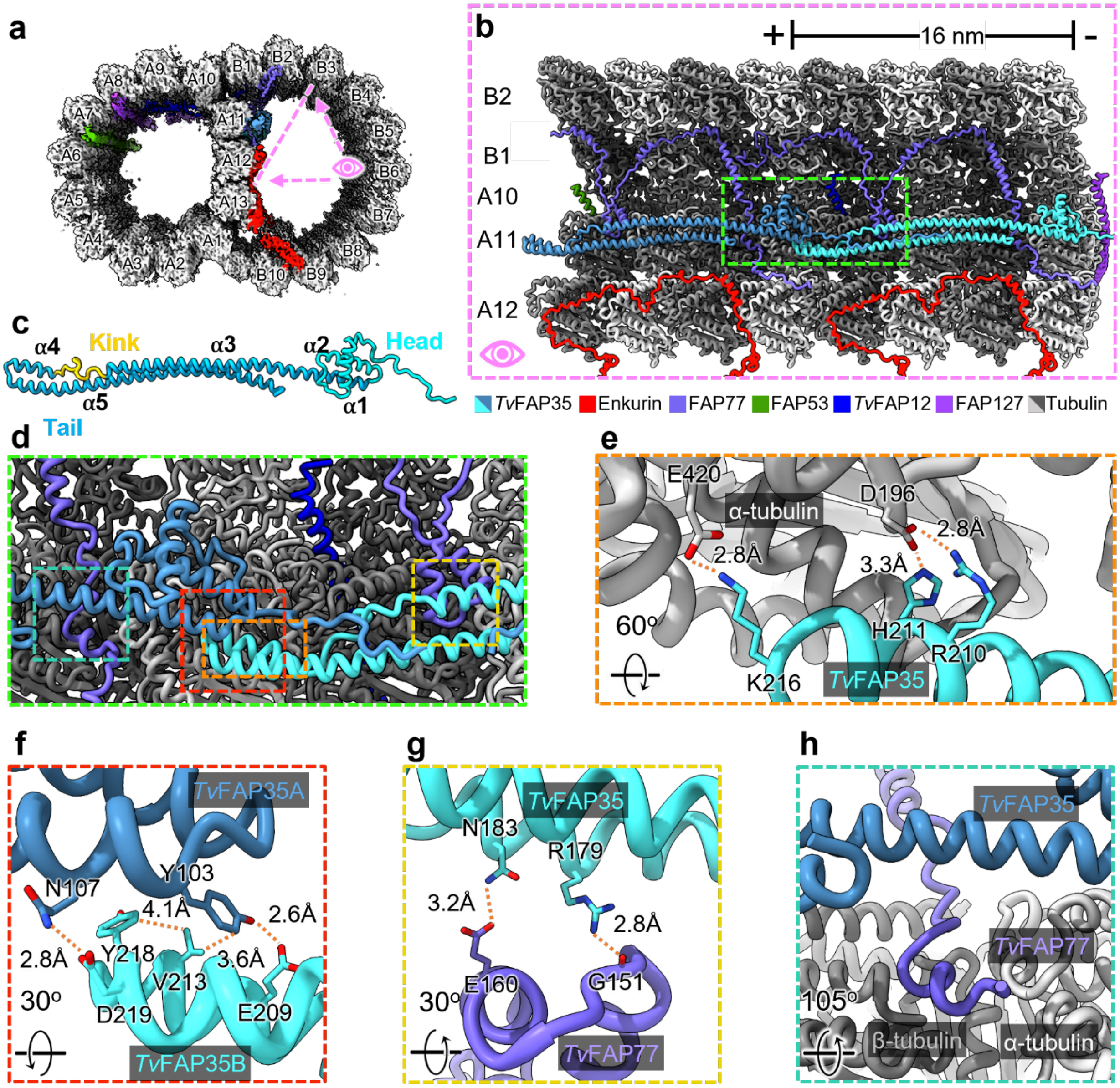


Figure 5- 5. TvFAP35 stabilizes ribbon PF A11 and outer junction proteins. (**a**) Cross-sectional view of the *Tv*DMT cryoEM map with enkurin and outer junction proteins colored. (**b**) 32 nm section of protofilaments A10, A11, A12, B1, and B2, along with their associated MIPs, shown with atomic models. (**c**) *Tv*FAP35 monomer labeled

with head (cyan), tail (blue), and kink (yellow), with helix numbers. (**d**) Zoomed-in view including important interactions of *Tv*FAP35. (**e**) Electrostatic interactions at the MT-binding motif of *Tv*FAP35. (**f**) Mixed residue interactions at the dimerization interface between *Tv*FAP35 monomers. (**g**) Interactions between TvFAP35 and the helix-turn-helix (residues 140-164) of *Tv*FAP77. (**h**) Residues 238-246 of *Tv*FAP77 pass near the *Tv*FAP35 coiled-coil. Residues 255 and after of *Tv*FAP77, which stretch further down, are omitted for clarity.

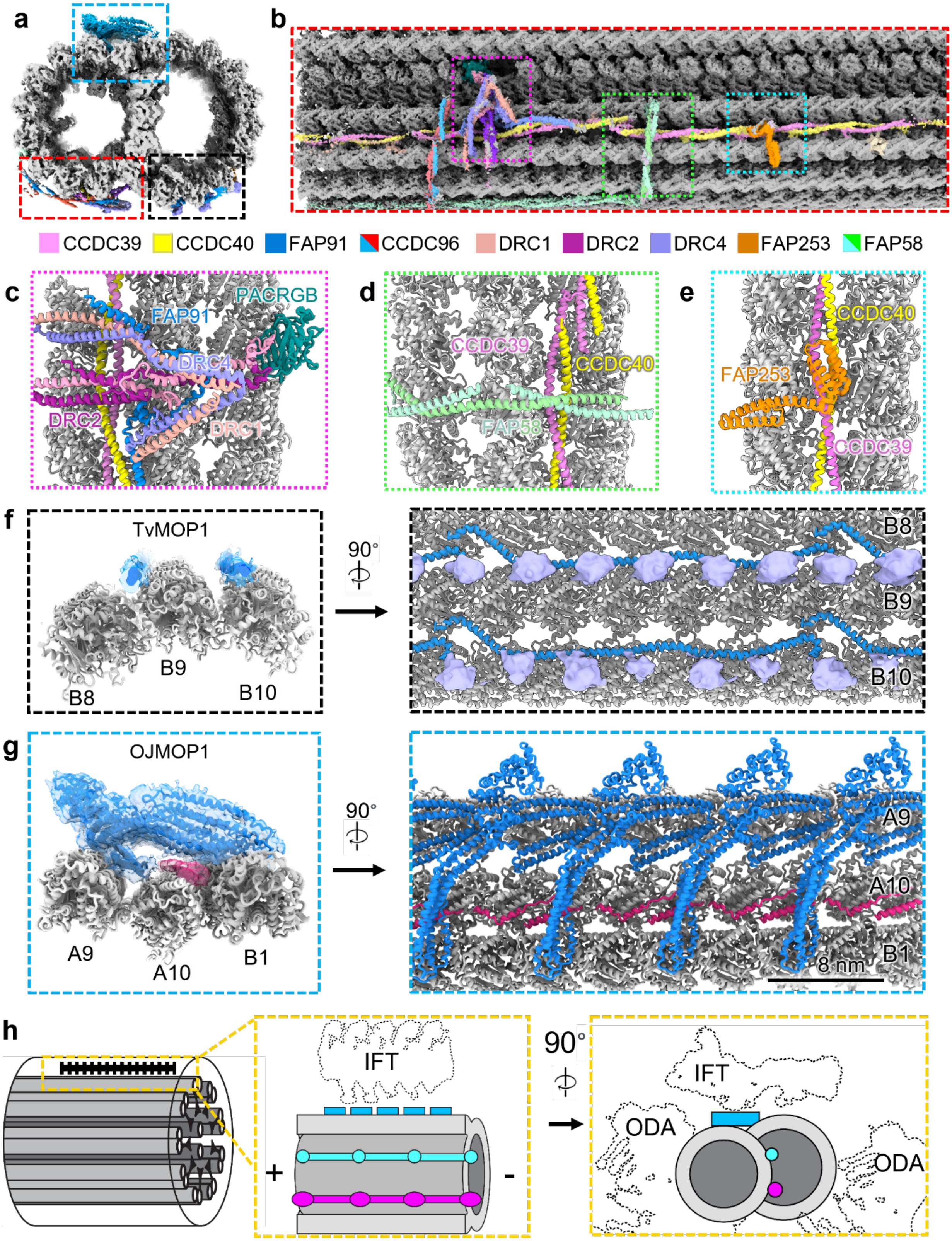
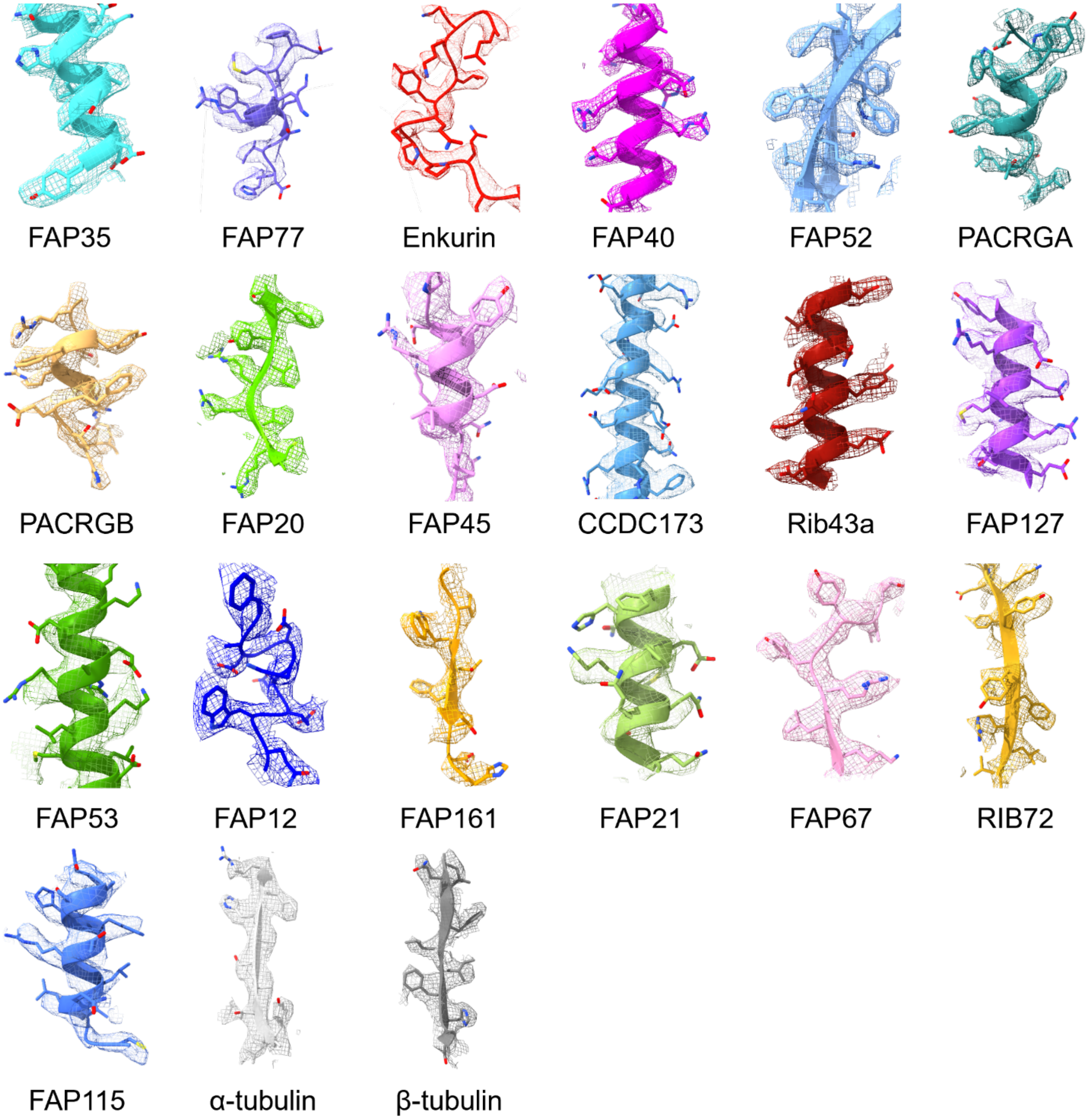
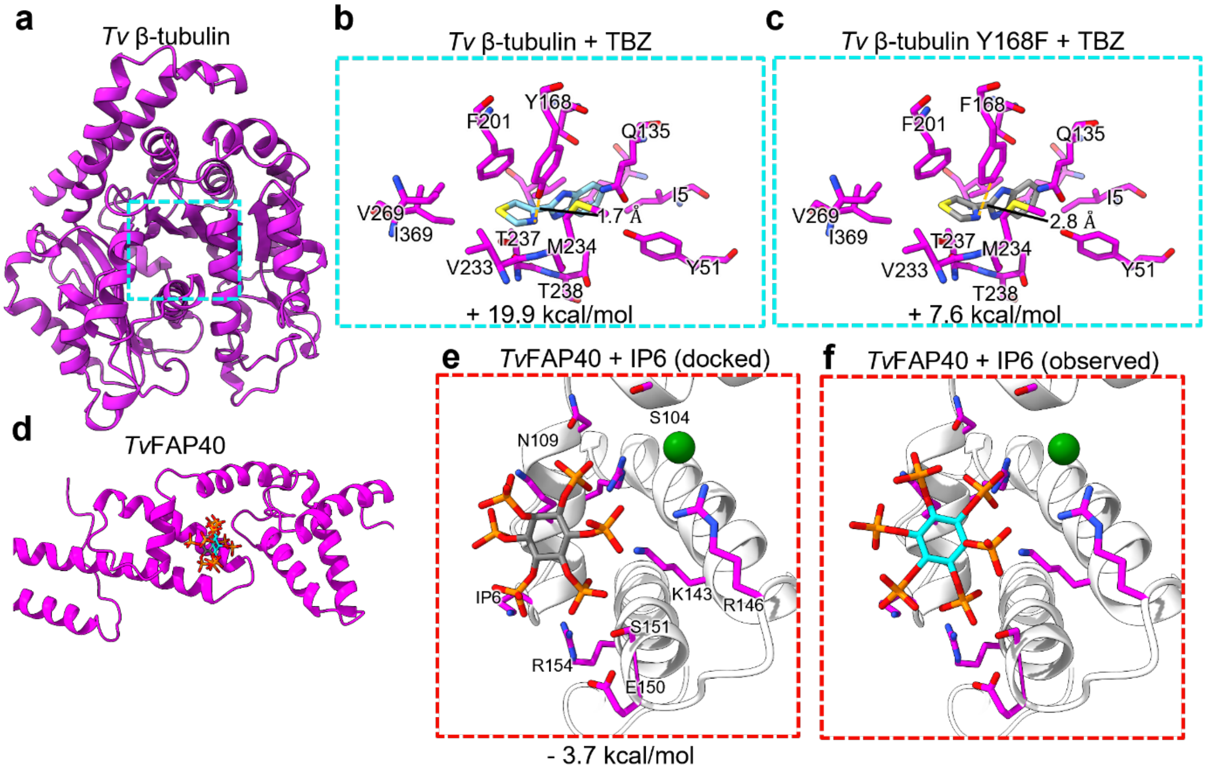

a
b
CCDC39
CCDC40
FAP91
CCDC96
DRC1
DRC2
DRC4
FAP253
FAP58
c
PACRGB
FAP91
DRC4
DRC2
DRC1
d
CCDC39
CCDC40
FAP58
e
CCDC40
FAP253
CCDC39
f
TvMOP1
B8
B9
B10
90°
B8
B9
B10
g
OJMOP1
A9
A10
B1
90°
A9
A10
B1
8 nm
h
IFT
+
-
90°
IFT
ODA
ODA

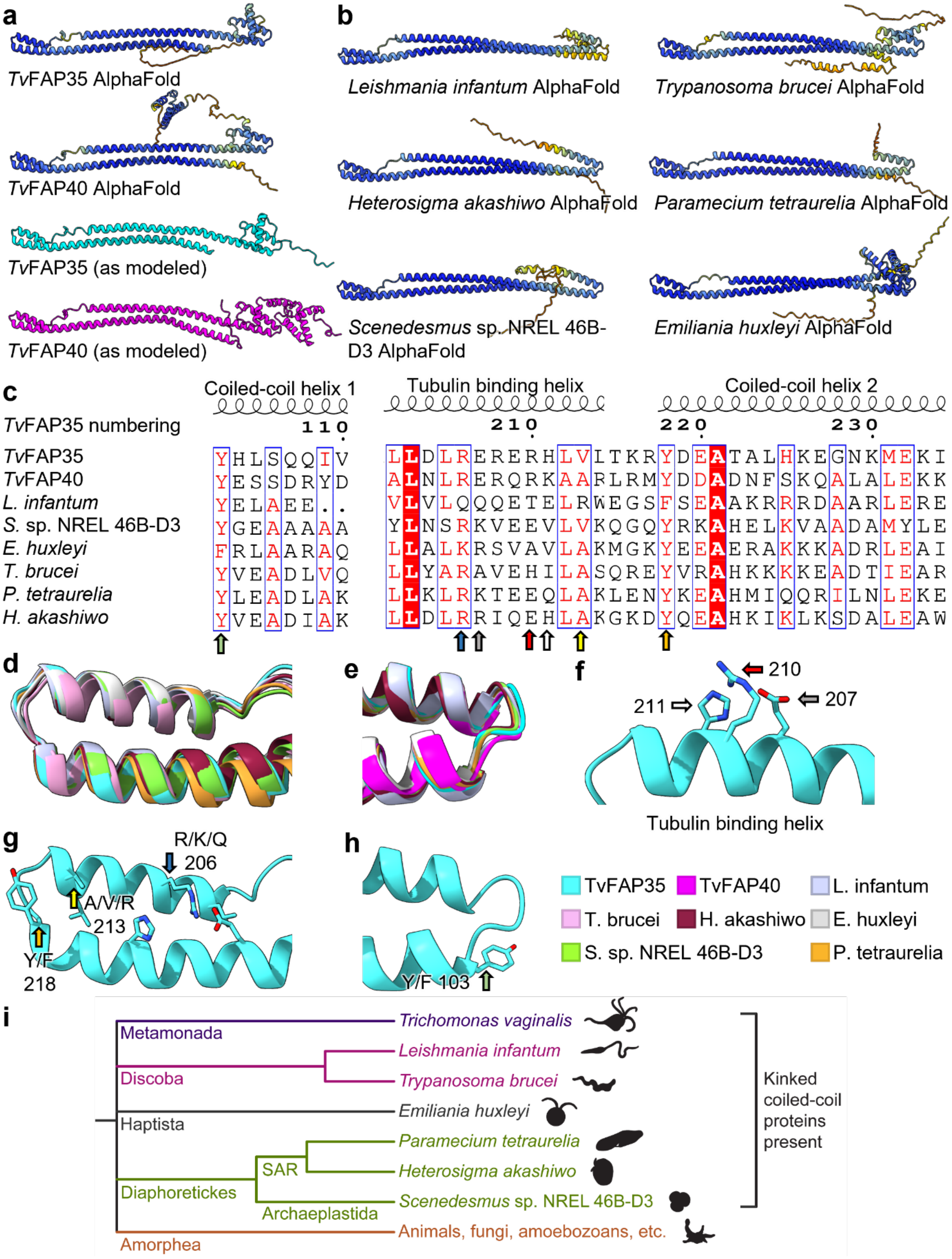
Figure 5- 6 Microtubule organization reveals novel 8nm periodicity. (**a**) Cross-sectional view of 96 nm repeat map, colored by MOP. (**b**) external view of *Tv*DMT and zoomed in views of MOPs (**c-e**). (**f**) *Tv*MOP1 demonstrating 24 nm periodicity as cross-section (left) and external view (right). (**g**) *Tv*OJMOP1 demonstrating 8nm periodicity with cross-sectional (left) and external views (right). (**h**) Schematic view of *Tv*-DMT organization with dotted lines to indicate positions of IFT and inner and outer dynein arm attachment (IDA and ODA).



Supplementary Figure 5- 1. Fitted models in cryoEM densities. Examples of cryoEM maps with fitted atomic models of MIP and MOP proteins.



Supplementary Figure 5- S2. Docking experiments of β-tubulin and TvFAP40. (**a**) Atomic model of β-tubulin with putative BZ drug binding site boxed. **(b)** WT *Tv* β-tubulin with docked thiabendazole (TBZ), fit into putative binding site. (**c**) *Tv* β-tubulin Y168F mutant with docked TBZ in putative binding site. (**d**) Atomic model of *Tv*FAP40 with putative IP6 binding site boxed. (**e**) *Tv*FAP40 binding pocket with docked IP6. (**f**) *Tv*FAP40 binding pocket with observed IP6.


a
TvFAP35 AlphaFold
TvFAP40 AlphaFold
TvFAP35 (as modeled)
TvFAP40 (as modeled)
b
Leishmania infantum AlphaFold
Trypanosoma brucei AlphaFold
Heterosigma akashiwo AlphaFold
Paramecium tetraurelia AlphaFold
Scenedesmus sp. NREL 46B-D3 AlphaFold
Emiliania huxleyi AlphaFold
c
Coiled-coil helix 1
Tubulin binding helix
Coiled-coil helix 2
TvFAP35 numbering
110
210
220
230
TvFAP35 YHLSQQIV LLDLRERERHLVLTKRYDEATALHKEGNKMEKI
TvFAP40 YESSDRYD ALNLRERQRKAARLRMYDDADNFSKQALALEKK
L. infantum YELAEE.. VLVLQQQETELRWEGSFSEAAKRRRDAARLERE
S. sp. NREL 46B-D3 YGEAAAAA YLNSRKVEEVLVKQGQYRKAHELKVAADAMYLE
E. huxleyi FRLAARAQ LLALKRSVAVLAKMGKYEEAERAKKKADRLEAI
T. brucei YVEADLVQ LLYARAVEHILASQREYVRAHKKKKEADTIEAR
P. tetraurelia YLEADLAK LLKLRKTEEQLAKLENYKEAHMIQQRILNLEKE
H. akashiwo YVEADIAK LLDLRRIQEHLAKGKDYQEAHKIKLKSDALEAW
d
e
f
210
211
207
Tubulin binding helix
g
R/K/Q
206
A/V/R
213
Y/F
218
h
Y/F 103
TvFAP35
TvFAP40
L. infantum
T. brucei
H. akashiwo
E. huxleyi
S. sp. NREL 46B-D3
P. tetraurelia
i
Metamonada
Trichomonas vaginalis
Discoba
Leishmania infantum
Trypanosoma brucei
Haptista
Emiliania huxleyi
SAR
Paramecium tetraurelia
Heterosigma akashiwo
Diaphoretickes
Archaeplastida
Scenedesmus sp. NREL 46B-D3
Amorphea
Animals, fungi, amoebozoans, etc.
Kinked coiled-coil proteins present

Supplementary Figure 5- S3. Analysis of TvFAP40 and TvFAP35 and structural homologs. (**a**) AlphaFold-predicted models of *Tv*FAP35 and *Tv*FAP40 (top) colored by AlphaFold confidence interval (blue more confident, red less confident) and their atomic models (bottom) colored in cyan and magenta respectively. **(b)** AlphaFold-predicted structures for structural homologs from selected species, colored by AlphaFold confidence interval. (**c**) Sequence alignment of dimerization and MT binding domain regions from proteins in **a** and **b** aligned to *Tv*FAP35, with conserved residues highlighted and those at the active site indicated with arrows. (**d and e**) α-carbon backbone aligned models from the MT-binding and dimerization domains of the kinked-coiled-coil domains. (**f**) Conserved proteins from **c** shown at their locations at the MT-binding interface on *Tv*FAP35. (**g-h**) Same as **f** but based on both faces of the dimerization domain. (**i**) Phylogeny tree including organisms in which FoldSeek identified similar protein structures.

Table 5- S1 CryoEM data collection

	WT *Tv*-DMT 16 nm repeat (EMD-46642)	WT *Tv*-DMT 48 nm repeat (EMD-46643)	WT *Tv*-DMT 96 nm repeat (EMD-46636)	WT *Tv*-DMT 48 nm composite (EMD-46580) (PDB 9D5N)
Data collection and processing				
Magnification	81,000	81,000	81,000	81,000
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	45	45	45	45
Defocus range (μm)	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5
Pixel size (Å)	1.1	1.1	1.1	1.1
Symmetry imposed	C1	C1	C1	C1
particle images (no.)	425,317	148,707	76,082	425,317
Map resolution (Å)	3.8	4.2	4.4	
FSC threshold	0.143	0.143	0.143	
Repeat unit (nm)	16	48	96	48
Symmetry imposed	C1	C1	C1	C1

Table 5- S2 MIPS and MOPS of *T. vaginalis*: Conserved and Novel MIPs in *Tv*, * indicates identified proteins unmodeled in *Tv* microtubule pdb, contact author for AlphaFold fit of these unmodeled proteins, Red text indicates Tv-specific proteins**)**

	Location	**Protein**	**Uniprot ID**	**Copy number in 96 nm repeat**	**Length (residues)**	**C. reinhardtii orthomolog**	**T.thermophila ortholog**	**Human ortholog**
1	Tubulin	□ tubulin	A2E8B1	276	452	α tubulin	α tubulin	α tubulin
2	Tubulin	□ tubulin	A2DC16	276	447	β tubulin	β tubulin	β tubulin
3	A-tubule	Rib72	A2GCC1	12	595	Rib72	Rib72	EFHC1/2
4	B-tubule	FAP45	A2ETR1	2	465	FAP45	CFAP45	FAP45
5	B-tubule	FAP52	A2FVE3	6	605	FAP52	CFAP52	FAP52
6	Inner Junction	PACRGB	A2EJQ5	6	241	PACRG	PACRG	PACRG
7	A-tubule	FAP21	A2F5C9	2	386	FAP21	FAP21	FAP21
8	Inner Junction	PACRGA	A2DAX1	6	236	PACRG	PACRG	PACRG
9	A-tubule	FAP53	A2G223	2	482	FAP53	CFAP53	FAP53
10	A-tubule	TvFAP12	A2F1C6	2	108	_	_	_
11	B-tubule	CCDC173	A2EFC9	2	455	FAP210	CCDC173	FAP210
12	A-tubule	FAP115	A2G843	3	927	FAP115	CFAP115	_
13	A-tubule	FAP67	A2E829	2	375	FAP67	CFAP67A	NME7
14	B-tubule	Tv-FAP35	A2DUL4	6	306	_	_	_
15	B-tubule	Tv-FAP40	A2DSS2	6	377	_	_	_
16	Ribbon	Rib43a	A2DEG5	2	383	Rib43a	Rib43a	RIBC2
17	B-tubule	FAP77	A2FTW1	6	283	FAP77	CFAP77	FAP77
18	A-tubule	TvFAP9	A2DVD4	2	75	_	_	_
19	A-tubule	FAP127	A2FH94	2	490	FAP127	CFAP127	MNS1
20	B-tubule	Enkurin	A2EMB8	6	241	FAP106	A8I9E8	ENKUR
21	Inner Junction	FAP20	A2EAE1	12	194	FAP20	CFAP20	FAP20
22	N-DRC	DRC4*	A2FWB4	1	512	DRC4	DRC4	DRC4
23	N-DRC	DRC2*	A2GIM1	1	461	DRC2	DRC2	DRC2

24	N-DRC	DRC1*	A2DEK4	1	633	DRC1	DRC1	DRC1
25	External coiled coils	CFAP58*	A2FE28	1	870	FAP189	CFAP58	FAP58
26	Radial Spoke Base	FAP253*	A2DGT8	1	396	FAP253	CFAP253	IQUB
27	Radial Spoke Base	CCDC96*	A2DTI0	1	363	FAP184	CCDC96	CCDC96
28	External coiled coils	CCDC39*	A2EKX3	1	998	FAP59	CCDC39	CCDC39
29	External coiled coils	CCDC40*	A2DSS7	1	889	FAP172	CCDC40	CCDC40
30	N-DRC/radial spoke	FAP91*	A2F6E6	1	602	FAP91	CFAP91	FAP91

5.10. References

1. Kissinger, P. Trichomonas vaginalis: a review of epidemiologic, clinical and treatment issues. *BMC Infectious Diseases* **15**, 307 (2015).
2. Johnston, V.J. & Mabey, D.C. Global epidemiology and control of Trichomonas vaginalis. in *Current opinion in infectious diseases* Vol. 21 56-64 (2008).
3. WHO. Global incidence and prevalence of selected curable sexually transmitted infections. (2012).
4. Kissinger, P. & Adamski, A. Trichomoniasis and HIV interactions: a review. in *Sexually Transmitted Infections* Vol. 89 426-433 (2013).
5. Zhang, Z. et al. The correlation between Trichomonas vaginalis infection and reproductive system cancer: a systematic review and meta-analysis. *Infectious Agents and Cancer* **18**, 15-15 (2023).
6. Kissinger, P. et al. Early Repeated Infections with Trichomonas vaginalis among HIV-Positive and HIV-Negative Women. *Clinical Infectious Diseases* **46**, 994-999 (2008).
7. Kirkcaldy, R.D. et al. Trichomonas vaginalis antimicrobial drug resistance in 6 US cities, STD Surveillance Network, 2009-2010. *Emerging infectious diseases* **18**, 939-43 (2012).
8. National Toxicology Program. Report on Carcinogens. Fifteenth Edition. (U.S. Department of Health and Human Services, Research Triangle Park, NC, 2021).
9. Malli, S., Loiseau, P.M. & Bouchemal, K. Trichomonas vaginalis Motility Is Blocked by Drug-Free Thermosensitive Hydrogel. *ACS Infectious Diseases* **6**, 114-123 (2020).

10. Korosh, T. et al. Potential of bisbenzimidazole-analogs toward metronidazole-resistant Trichomonas vaginalis isolates. *Chemical Biology & Drug Design* **90**, 489-495 (2017).

11. Aguirre, G. et al. Novel antiprotozoal products: Imidazole and benzimidazole N-oxide derivatives and related compounds. *Archiv der Pharmazie* **337**, 259-270 (2004).

12. Katiyar, S.K. & Edlind, T.D. β-Tubulin genes of Trichomonas vaginalis. *Molecular and Biochemical Parasitology* **64**, 33-42 (1994).

13. Coceres, V.M. et al. Ultrastructural and Functional Analysis of a Novel Extra-Axonemal Structure in Parasitic Trichomonads. in *Frontiers in Cellular and Infection Microbiology* Vol. 11 1-21 (2021).

14. Manton, I. & Clarke, B. Electron Microscope Observations on the Spermatozoid of Fucus. *Nature* **166**, 973-974 (1950).

15. Ma, M. et al. Structure of the Decorated Ciliary Doublet Microtubule. in *Cell* Vol. 179 909-922.e12 (Elsevier Inc., 2019).

16. Walton, T., Wu, H. & Brown, A. Structure of a microtubule-bound axonemal dynein. *Nature Communications* **12**, 477-477 (2021).

17. Walton, T. et al. Axonemal structures reveal mechanoregulatory and disease mechanisms. *Nature* **618**, 625-633 (2023).

18. Kubo, S. et al. Native doublet microtubules from Tetrahymena thermophila reveal the importance of outer junction proteins. *Nature Communications* **14**, 2168-2168 (2023).

19. Gui, M. et al. De novo identification of mammalian ciliary motility proteins using cryo-EM. in *Cell* Vol. 184 5791-5806.e19 (Elsevier Inc., 2021).

20. Imhof, S. et al. Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen Trypanosoma brucei. *eLife* **8**, 1-30 (2019).

21. Chen, Z. et al. De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking. *Cell* **186**, 5041-5053.e19 (2023).

22. Tai, L., Yin, G., Huang, X., Sun, F. & Zhu, Y. In-cell structural insight into the stability of sperm microtubule doublet. *Cell Discovery* **9**, 116-116 (2023).

23. Lenaghan, S.C., Nwandu-Vincent, S., Reese, B.E. & Zhang, M. Unlocking the secrets of multi-flagellated propulsion: drawing insights from Tritrichomonas foetus. *Journal of The Royal Society Interface* **11**, 20131149-20131149 (2014).

24. Ho, C.-M. et al. Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. in *Nature Methods* Vol. 17 79-85 (Springer US, 2020).

25. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. in *Nature* Vol. 596 583-589 (Springer US, 2021).

26. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439-D444 (2022).

27. Aguayo-Ortiz, R. et al. Towards the identification of the binding site of benzimidazoles to β-tubulin of Trichinella spiralis: Insights from computational

and experimental data. *Journal of Molecular Graphics and Modelling* **41**, 12-19 (2013).

28. Juliano, C., Martinotti, M.G. & Cappuccinelli, P. "In vitro" effect of microtubule inhibitors on Trichomonas vaginalis. *Microbiologica* **8**, 31-42 (1985).
29. Ma, M. et al. Structure of the Decorated Ciliary Doublet Microtubule. *Cell* **179**, 909-922.e12 (2019).
30. Leung, M.R. et al. Structural specializations of the sperm tail. *Cell* **186**, 2880-2896.e17 (2023).
31. Berman, H.M. et al. The Protein Data Bank. in *Nucleic Acids Research* Vol. 28 235-242 (2000).
32. Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523-D531 (2023).
33. Mallery, D.L. et al. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife* **7**(2018).
34. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* **39**, W270-W277 (2011).
35. Eberhardt, J., Santos-Martins, D., Tillack, A.F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling* **61**, 3891-3898 (2021).
36. Bugnon, M. et al. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Research*, 1-9 (2024).

37. Trott, O. & Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455-461 (2010).

38. Sarmah, B., Winfrey, V.P., Olson, G.E., Appel, B. & Wente, S.R. A role for the inositol kinase Ipk1 in ciliary beating and length maintenance. *Proceedings of the National Academy of Sciences* **104**, 19843-19848 (2007).

39. Sarmah, B., Latimer, A.J., Appel, B. & Wente, S.R. Inositol Polyphosphates Regulate Zebrafish Left-Right Asymmetry. *Developmental Cell* **9**, 133-145 (2005).

40. Fedorov, V.A. et al. Mechanical properties of tubulin intra- and inter-dimer interfaces and their implications for microtubule dynamic instability. *PLOS Computational Biology* **15**, e1007327-e1007327 (2019).

41. van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**, 243-246 (2024).

42. Gao, J. et al. DomainFit: Identification of protein domains in cryo-EM maps at intermediate resolution using AlphaFold2-predicted models. *Structure*, 1-12 (2024).

43. Lacey, S.E., Foster, H.E. & Pigino, G. The molecular structure of IFT-A and IFT-B in anterograde intraflagellar transport trains. *Nature Structural & Molecular Biology* **30**, 584-593 (2023).

44. Reck-Peterson, S.L. et al. Single-Molecule Analysis of Dynein Processivity and Stepping Behavior. *Cell* **126**, 335-348 (2006).

45. de Miguel, N., Riestra, A. & Johnson, P.J. Reversible association of tetraspanin with Trichomonas vaginalis flagella upon adherence to host cells. *Cellular Microbiology* **14**, 1797-1807 (2012).

46. Beauchamp, P.J., Galle, P.C. & Blasco, L. Human Sperm Velocity and Postinsemination Cervical Mucus Test in the Evaluation of the Infertile Couple. *Archives of Andrology* **13**, 107-112 (1984).

47. Smyth, M., Pettitt, T., Symonds, A. & Martin, J. Identification of the pocket factors in a picornavirus. *Archives of Virology* **148**, 1225-1233 (2003).

48. Hardy, J.M. et al. A unified route for flavivirus structures uncovers essential pocket factors conserved across pathogenic viruses. *Nature Communications* **12**, 3266-3266 (2021).

49. Flatt, J.W., Domanska, A., Seppälä, A.L. & Butcher, S.J. Identification of a conserved virion-stabilizing network inside the interprotomer pocket of enteroviruses. *Communications Biology* **4**, 250-250 (2021).

50. Clark, C.G. & Diamond, L.S. Methods for Cultivation of Luminal Parasitic Protists of Clinical Importance. in *Clinical Microbiology Reviews* Vol. 15 329-341 (2002).

51. Stevens, A., Muratore, K., Cui, Y., Johnson, P.J. & Zhou, Z.H. Atomic Structure of the Trichomonas vaginalis Double-Stranded RNA Virus 2. *mBio* **12**, 1-17 (2021).

52. Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V. & Dunin-Horkawicz, S. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790-2795 (2019).

53. Garg, S. et al. Conservation of Transit Peptide-Independent Protein Import into the Mitochondrial and Hydrogenosomal Matrix. *Genome Biology and Evolution* **7**, 2716-2726 (2015).
54. Mastronarde, D.N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microscopy and Microanalysis* **9**, 1182-1183 (2003).
55. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. in *Nature Methods* Vol. 14 290-296 (2017).
56. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. in *Nature Methods* Vol. 16 1153-1160 (Springer US, 2019).
57. Scheres, S.H.W. Single-particle processing in RELION. in *Manuals* 1-21 (2013).
58. Lövestam, S. & Scheres, S.H.W. High-throughput cryo-EM structure determination of amyloids. *Faraday Discussions* **240**, 243-260 (2022).
59. Croll, T.I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology* **74**, 519-530 (2018).
60. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14-25 (2018).
61. McGinnis, S. & Madden, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* **32**, W20-W25 (2004).

62. Pfab, J., Phan, N.M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proceedings of the National Academy of Sciences* **118**(2021).

63. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**, 486-501 (2010).

64. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024).

65. Alvarez-Jarreta, J. et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in 2023. *Nucleic Acids Research* **52**, D808-D816 (2024).

66. Yu, I. et al. Building atomic models based on near atomic resolution cryoEM maps with existing tools. in *Journal of Structural Biology* Vol. 204 313-318 (Elsevier, 2018).

67. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology* **74**, 531-544 (2018).

6. Chapter 6: Conclusion and future perspectives

As structural biologists seek to address a broader range of questions, the need for versatile and broadly applicable methods has become increasingly evident. In this work, we successfully employed cryoEM modalities to characterize pathogenically relevant complexes and elucidate the mechanisms underpinning infection. By focusing on dsRNA viruses, we aimed to determine the capsid features essential to replication and immune evasion within the host cytosol, using TVV virus from *T. vaginalis* as a model of minimal complexity. Our 3.6 Å structure of its viral capsid revealed that TVV's intracellular lifestyle does not necessitate tight binding between capsid subunits compared to extracellularly transmitted viruses. Further, insights derived from the structure of the ARV core enabled us to identify architectural changes that form an integral part of transcriptional initiation and elongation during viral replication. Because the genome organization of ARV correlates with TEC organization, we were also able to describe the structure of several of the outer layers of the genome.

We also utilized previously determined cryoEM structures to guide targeted mutagenesis experiments in the common human herpesvirus HCMV. Immunofluorescence imaging from this work suggested that the tegument protein pp150 may hitch a ride along with the SCP-MCP-scaffolding complex for nuclear trafficking. The discovery of the pp150 K255E mutation, which appreciably slowed viral spread in cultured cells, demonstrated the potential to modulate viral replication based on insights gained from cryoEM studies. Our results in this mutagenesis work highlights the power of cryoEM to inform rational vaccine and drug design.

Furthermore, we showcased the capacity of cryoEM to resolve the structures of doublet microtubules (DMTs) from the infectious parasite *T. vaginalis.* Our analysis of these DMT structures revealed the MIPs essential for the species-specific motility of *T. vaginalis*. Among the newly identified kinked-coil-coil proteins, *Tv*-FAP40 stood out for its unique ligand binding site, which lacks apparent homologues from other species. which makes it an attractive drug target that may reduce the chance of off target effects.

The exploding popularity of single particle cryoEM, coupled with advancements in data acquisition and analysis tools, promises to improve our understanding of a wide array of systems. Exciting new developments are emerging in cryogenic electron tomography (cryoET), which, unlike cryoEM, does not require identical particles to resolve detailed structures. Though the resolution achievable with cryoET still lags behind cryoEM, its ability to visualize cross sections of whole cells, and describe their molecular organization represents a significant leap in contextualizing these systems we have studied for decades and will open the door to many more. The impending era of molecular sociology promises to bring with it more questions that will undoubtedly continue to drive the field forward for decades.