UCLA
UCLA Electronic Theses and Dissertations

Title
CryoET reveals structures essential to infections by pathogens from viruses to protozoa

Permalink
https://escholarship.org/uc/item/8t23555z

Author
zhang, jiayan

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

CryoET Reveals Structures Essential to Infections by

Pathogens from Viruses to Protozoa

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Jiayan Zhang

2021

ABSTRACT OF THE DISSERTATION

CryoET Reveals Structures Essential to Infections by

Pathogens from Viruses to Protozoa

by

Jiayan Zhang

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2021

Professor Z. Hong Zhou, Chair

Infections by viruses and flagellated protozoa lead to many human diseases, ranging from mild cold sores to devastating cancers. These pathogens infect billions of people worldwide and afflict wild and domestic animals, placing an enormous burden on global public health and economy. To resolve their *in situ* structures essential for infection, we have firstly established the validity of our approach of cryoET with sub-tomogram averaging using virus as a gold standard organism, then applied the same approach to investigate the structural details of flagellum in more complex protozoa *Trypanosoma brucei*.

In this thesis work, we firstly established the validity of our cryoET approach by obtaining in situ structures of the vesicular stomatitis virus glycoprotein G trimer in prefusion and postfusion conformations, which agree with the known crystal structures of purified G trimers in both conformations. Later for herpesvirus HCMV, we resolved 3D structures of gB trimers in two distinct conformations at up to 21 Å resolution. We further

captured the prefusion gB in complex with an "L"-shaped density attributed to the gH/gL complex. Our results deepen the knowledge of membrane fusion mechanism during HCMV infection.

How does nuclear egress complex (NEC)—the mediator of viral capsid budding into the cytoplasm—interacts with the viral capsid and how proper curvature of the coat is achieved to enable budding? We applied our cryoET approach to report that binding of a capsid protein, UL25, promotes the formation of a pentagonal rather than hexagonal NEC arrangement. Our results suggest that during nuclear budding, interactions between UL25 bound to the pentagonal capsid vertices and NEC introduce pentagonal insertions into the hexagonal NEC array to yield an NEC coat of the appropriate size and curvature, leading to productive budding and egress of UL25-decorated capsids.

Furthermore, to understand structural foundations of eukaryotic flagella supporting motility and signaling necessary for infection, transmission, and pathogenesis, we then applied our cryoET approach to study *T. brucei*, a protozoan parasite in the Excavata lineage that causes African trypanosomiasis. From our resolved structure of its 96-nm axonemal repeats, we discovered several lineage-specific structures, including novel inter-doublet linkages and microtubule inner proteins. We also determined structures of the paraflagellar rod (PFR), PFR-axoneme connectors, and the axonemal central pair complex. Together, our findings fill the previously critical gap in understanding structural foundations of eukaryotic flagella, provide insights into flagellum-driven, non-planar helical motility of *T. brucei* and have broad implications ranging from cell motility and tensegrity in biology to engineering principles in bionics.

The dissertation of Jiayan Zhang is approved.

Jeffrey F. Miller

Kent L. Hill

Ren Sun

Peter Bradley

Z. Hong Zhou, Committee Chair

University of California, Los Angeles

2021

Acknowledgements

I'd like to first and foremost thank my mentor, Dr. Z Hong Zhou, for his unwavering support and guidance during my five-year PhD training. He works endlessly in his power to provide his trainees with the support, environment, and resources needed to accomplish cutting edge research. Hong is my role model as a scientist, who works very hard and is always energetic, full of passion and curiosity in exploring scientific questions, also with a big picture and a clear vision. I have been truly fortunate to have him as my mentor. I firmly believe that the valuable trainings I have received, the relationship I have built, and the work I have accomplished under Hong's mentorship would never be possible anywhere else. I will never forget when I first joined the lab, Dr. Zhou's kind words: "You will do a good job, and I will guide you along the way" since he knew that I came from a resource poor community and my parents did not even have the chance to get education. When I struggled how to present scientific works in proper English, he 'tortured' and stretched me to be an efficient and structured presenter, and I was lucky enough to be awarded as the 'best presenter' in one symposium.

I also want to express my deepest appreciation to my thesis committee members, Dr. Jeffrey Miller, Kent Hill, Ren Sun, and Hong Zhou. I would like to thank them for their valuable and critical advice on my research projects, and their always support and encouragement along the journey. I am deeply grateful that they always made them available to me from their tight schedule when I have a need and infinite thanks to the four gentlemen for their tolerance for my lots of stupid questions or even mistakes.

My special thanks to all the past and present members of the Zhou lab, for their support, camaraderie, and accompany during my time at UCLA. I am grateful that many

senior members in the group for helping me with my questions about the computer and microscope: Dr. Peng Ge, Ivo Atanasov, Wong Hoi, Dr. Yanxiang Cui, Dr. Kang Zhou. Your almost 24/7 support made it possible and so smooth for every group member to get access to the well-built computational server and the Titan Krios. I also would like to thank Dr. Yuntao Liu, Dr. Chi-min Ho, Dr. Kevin Huynh, Dr. Ana Lucia Alvarez-Cabrera, Dr. Wei Liu, Dr. Xinghong Dai, Dr, Xian Xia, Dr. Yao He, Dr. Shiheng Liu, Dr. Zhu Si, Dr. Xuekui Yu, Paul, Alex, Jonathan, Xueting, Shiqing, Hui Wang, and others for supporting me in different aspects along the way.

I also would like to thank the Molecular Biology Institute (MBI) and department of Immunity, Microbes, and Molecular Pathogenesis (IMMP) at UCLA for nominating me for the Whitcome Fellow, twice (Big Thank you Dr. Philip Whitcome!) and the dissertation year fellowship, both fellowships I am deeply grateful to have. These prestigious fellowships provide fantastic opportunities to international students like me to explore the beauty of science.

I am deeply grateful to my collaborators Dr. Kent Hill, Dr. Rob Gunsalus, and Dr. Katya Heldwein for their unending support and mentorship, for being a delight to work with, and for welcoming me into the world of *T. brucei*, *M. Hungatei*, and HSV-1 nuclear egress with open arms. The works in this thesis were not possible without the extensive knowledge, hard work, and the insightful discussions from all contributors and co-authors, including Dr. Zhu Si, Dr. JJW Wong, Dr. Elizabeth Draganova, Dr. Simon Imhof, and Hui Wang, Dr. Katya Heldwein, Dr. Theodore Jardetzky, Dr. Kent Hill, Dr. Hong Zhou, and others.

I have been fortunate enough to have had a long history of supportive and giving mentors throughout my school journey, who have always supported me, believed in me,

Table of Contents

List of Figures

Chapter 4

List of Tables

Vita

Education

- PhD student, Molecular Biology, University of California, Los Angeles, USA

  2016 – 2021

- Bachelor in Horticulture, Tang Aoqing Honors Program in Science, Jilin University, China   2012 – 2016

Publications

- **Jiayan Zhang***, Hui Wang*., *et al*. Structures of trypanosome flagellum paraflagellar rod and mechanism of eukaryotic cell motility. ***Cell Discovery***. CELLDISC-01632-T. DOI: 10.1038/s41421-021-00281-2.

- Imhof Simon*., **Zhang, Jiayan***., Wang Hui., *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*** 2019;8:e52058 DOI: 10.7554/eLife.52058.

- Si, Zhu*, **Zhang, Jiayan***., Sakar Shivakoti., Atanasov I., Tao CL., *et al*. Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. ***PLOS Pathogens*** 14(12): e1007452. https://doi.org/10.1371/journal.ppat.1007452.

- Draganova Elizabeth., **Zhang, Jiayan**., ZH Zhou, Ekaterina Heldwein. Structural basis for capsid recruitment and coat formation during HSV-1 nuclear egress. ***eLife*** 2020;9:e56627 DOI: 10.7554/eLife.56627

- Wang, J.J.W., Young, T.A., **Zhang, Jiayan**. *et al*. Monomeric ephrinB2 binding induces allosteric changes in Nipah virus G that precede its full activation. ***Nature Communications*** 8, 781 (2017).

https://doi.org/10.1038/s41467-017-00863-3

- ZH Zhou., WH Hui., **Zhang, Jiayan**., *et al*. In situ Structure of Viral RNA by Cryo Electron Tomography with Volta Phase Plate, Energy Filtering and Direct Electron Counting. ***Microscopy and Microanalysis*** 22(S3):74-75. DOI: 10.1017/S1431927616001227

- B Jia., Z Li., J Liu., U Sun., X Jia., YH Xuan., **J Zhang**., *et al*. A zinc-dependent protease AMZ-tk from a thermophilic archaeon is a new member of the archaemetzincin protein family. ***Front Microbiol***. 2015 Dec 17;6:1380. doi: 10.3389/fmicb.2015.01380. eCollection 2015.

- X Xie., X Li., Y Tian., M Su., **J Zhang**., *et al*. Identification and characterization of microRNAs and their targets from expression sequence tags of Ribes nigrum. ***Canadian Journal of Plant Science*** 96(6). DOI: 10.1139/CJPS-2016-0013

*: These authors contribute equally to this research.

Awards and Honors (selected)

- UCLA Dissertation Year Fellowship   2020 – 2021

- UCLA Whitcome Pre-doctoral Fellowship in Molecular Biology   2019 – 2021

- Best Student Presentation (1st place), Southern California Society for Microscopy and Microanalysis Annual Spring Symposium.   2019

- First place in the IXL Innovation Olympics Competition   2017

- First place in UCLA/L.E.K. Advanced Degree Case Competition   2017

- Top 10 Undergraduates in Jilin University (10/60,000, ranked 1st among the 10)   2015

- National Scholarship in China   2013, 2014

Chapter 1: Introduction

1.1 Introduction to cryogenic electron tomography

1.1.1 Cryogenic electron tomography

Single-particle cryogenic electron microscopy (cryoEM) has been elevated as the mainstream of structural biology propelled by three technological innovations that together allow for the successful imaging and analyzing of biological samples at near-atomic resolution using phase contrast transmission electron microscopy (TEM). The three key innovations—structure preservation in amorphous ice, sensitive direct electron detector, and advanced computer algorithms for imaging processing—enabled the resolution revolution in cryoEM. For their contributions in these three areas, Richard Henderson, Jacques Dubochet, and Joachim Frank, earned the Nobel Prize in Chemistry in 2017 [1].

For structural biologists, the technological advancements of cryoEM have led to the visualization, identification, and interpretation of uncovered structural details of biological molecules and complex multicomponent assemblies which were purified, engineered, or fractionated, that have provided insights into molecular biology. However, awareness has grown in recent years that biological functions can rarely be attributed to individual macromolecules. Most cellular functions arise from their concerted action, thus there is a need for methods enabling structural studies performed in situ, ideally in unperturbed cellular environments [2]. Cryogenic electron tomography (cryoET) is a groundbreaking technology for three-dimensional (3D) visualization and structural investigations of "biomolecules sociology" of cells—single proteins and their interactions and spatial arrangement—in their native state [2].

1.1.2 Workflow of cryoET
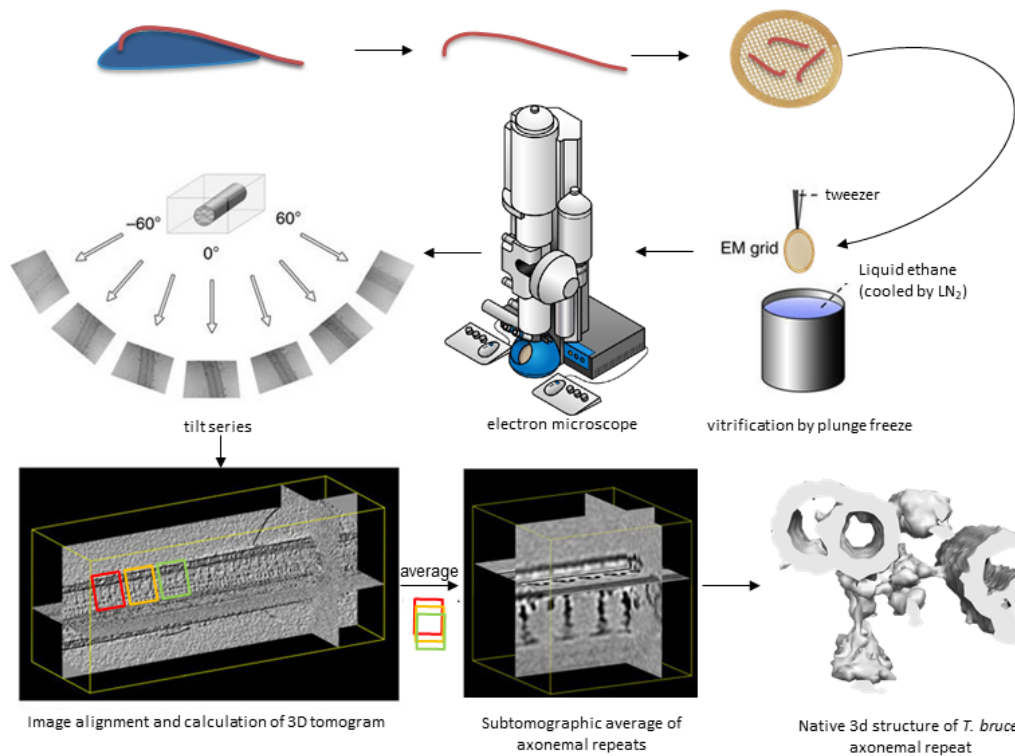
Figure 1- 1. CryoET workflow. CryoET workflow from sample preparation (top panel), tomogram acquisition (middle panel), reconstruction and sub-tomogram averaging (bottom panel).

CryoET can be applied to a broad range of biological samples from isolated macromolecules to multicellular organisms. The biological sample is firstly purified and mixed with fiducial gold beads in certain ratio and then applied on the Quantifoil holey carbon grids which was freshly glow discharged. Excess of the sample mixture is blotted away with a filter paper with controlled environmental humidity, temperature, blotting force, and blotting time, and vitrified by immediately plunging into liquid nitrogen-cooled liquid ethane with an FEI Mark IV Vitrobot cryo-sample plunger. Vitrified buffer film is not crystallized that electron beam can pass through it without strong diffraction. For high-resolution cryoET tilt series acquisition, vitrified specimens are transferred with a

cryo-holder into an FEI Titan Krios 300kV TEM equipped with a Gatan imaging filter (GIF) and a direct electron detector. The electrons coming from the gun will be navigated by a series of lenses to pass through the biomolecules frozen inside the vitrified buffer to finally generate a projected image on the electron detector. Samples are imaged under low-dose condition to reduce the radiation damage issue and the cryoET tilt series are recorded with software packages like *SerialEM[3]* by tilting the specimen stage from the maximum -70° to -70° with 2° or 3° increments.

Frames in each movie of the raw tilt series are drift-corrected, coarsely aligned and averaged with *Motioncorr* [4] which produced a single image for each tilting angle. The tilt series images are reconstructed into 3D tomograms by weighted back projections using the *IMOD* software package [5]. To improve the signal-to-noise ratio and enhance the resolution, sub-tomograms containing targeted components were extracted from the raw tomograms. Sub-tomogram averaging and the missing-wedge compensation are performed using PEET program (Figure 1-1) [6, 7].

1.1.3 State of the art in cryoET

CryoET has provided insights into the structure and function of viruses [8-12], nuclear pore complexes[13-16], cytoskeletal elements[17-19], flagellar motors[20], S-layer proteins [21] and others [2].

Structural studies of viruses have been driving methods developments in 3D EM since the early days[22]. CryoET and subtomogram averaging has been used widely for studies of pleomorphic viruses. Examples include the murine leukemia retrovirus [23, 24], rubella virus [25], herpes simplex virus type I [8, 26], and the severe acute respiratory syndrome coronavirus 2 [12, 27]. John Briggs group applied cryoET and subtomogram averaging of viral lattices with high local order to determine and report

the first sub-nanometer resolutions for the ex situ Ebola nuclear capsid [9] and fullerene-like HIV-1 capsid [10]. Later, Biggs group resolved the capsid and spacer peptide 1 region of Gag—the key regulator of assembly and maturation and the target of maturation inhibitors—within assembled immature HIV-1 particles at 3.9 Å resolution and built an atomic model. The atomic-resolution structure reveals a network of intra- and intermolecular interactions mediating immature HIV-1 assembly [11]. Impressively, the resolution of such structure was further improved to 3.4 Å by applying 3D CTF correction [28] by Briggs group, and more recently to 3.1 Å with *emClarity*, a sub-tomogram averaging package [29]. In addition to sub-tomogram averaging and data-collection advances, new insights into the assembly and maturation of viral capsids have been obtained. In 2020, Briggs group imaged intact SARS-CoV-2 virions and determined the high-resolution structure, conformational flexibility, and distribution of S trimers *in situ* on the virion surface. This work provided a basis from which to understand interactions between S and neutralizing antibodies during infection or vaccination [12].

Application of cryoET and subtomogram averaging approach has resulted in the first 4.8 Å *in situ* structure of Caulobacter crescentus surface layer proteins (S-layers) [21] (Fig. 4A). S-layers are 2D arrays of proteins that provide protection and structural support to many bacteria and archaea. Tanmay Bharat group presented a near-atomic resolution *in situ* structure of the complete S-layer using an integrative approach combining cryoET and subtomogram averaging, single-particle averaging, and native mass spectrometry (MS). This provided insights into the S-layer lattice organization, its interaction with surface lipopolysaccharides and dependence on calcium ions, and exemplifies the progress EM has made in the past three decades, since the first electron crystallography structure of C. crescentus S-layer in 1992 [30].

The promise of cryoET has been partially reflected by the applications introduced

above. For a long time, investigating isolated and purified molecular complexes was the only possible method of understanding structural details to uncover their functions. However, technological advancements in cryoET allow *in situ* structural studies at 3–4 nm resolution to even sub-nanometer resolutions with averaging methods. We now just begin to unlock the potential of cryoET and subtomogram averaging as a productive approach to study the *in situ* "biomolecules sociology" of cells.

1.2 Infection

An infection is the invasion and growth of microorganism in the body, such as bacteria, viruses, fungi, and parasites. These infectious microscopic organisms are known as pathogens which can spread in several different ways including skin contact, the transfer of bodily fluids, contact with feces, ingesting contaminated food or water, inhaling airborne particles or droplets, etc. The microorganism uses the host's body to sustain itself, reproduce, and colonize. The immune system is an effective barrier against infectious agents. However, pathogens may sometimes overwhelm the immune system's ability to fight them off. At this stage, an infection becomes harmful which can cause symptoms including fever, chills, cough, diarrhea, ear pain, skin rash, etc.

1.2.1 Bacterial Infection

Bacterial infections remain the leading cause of death in children, the elderly, and immunocompromised patients. However, antibiotics and vaccines have drastically abridged the mortality rate caused by bacterial infections worldwide [31]. Antibiotics are usually the main treatment for bacterial infections and have dramatically extended the average human lifespan by 23 years [32] in just over 100 years since the first antibiotic, salvarsan was deployed in 1910. Vaccines are available that can help prevent bacterial

infections, for example, the DTaP and Tdap vaccines both protect against three bacterial infections: diphtheria, tetanus, and pertussis. While in the current climate of frequently occurring antibiotic resistance, new technologies such as genome mining and editing strongly suggests promising signs for a bright future of antibiotic discovery [32, 33] from natural products [34]. Therefore, my thesis work mainly focuses on two medical-significant infections—viral infection and parasitic infection due to the limited space, although I have made progress on determining the high-resolution structures of the surface layer proteins and sheath of archaea *M. hungatei*.

1.2.2 Viral Infection

Viral infections are the main agents of infection-causing cancers in humans [35]. Viral diseases are leading causes of mortality and morbidity among infants requiring care in the neonatal intensive care unit [36].

In this thesis work, I applied the approach of cryoET and sub-tomogram averaging to determine the *in situ* structures of fusion gB on a herpesvirus, Human cytomegalovirus (HCMV) to provide insights on the essential viral infection mechanism. Also, this thesis work described the capsid recruitment and coat formation during HSV-1 nuclear egress.

Infection by herpesviruses leads to many human diseases, ranging from mild cold sores to devastating cancers. HCMV is a member of the β-herpesvirus subfamily of the Herpesviridae, with a high seroprevalence in the global human population via direct contact with bodily secretions, ranging from 50% in industrialized countries [37] to as high as 99% in resource-poor communities, and developing countries [38]. HCMV is the leading viral cause of devastating birth defects and life-threatening complications among immunocompromised individuals, such as infants, elders, AIDS and transplant

patients. Like other human herpesviruses, HCMV is never completely cleared and remains latent throughout the life of the host. Current anti-HCMV therapies are toxic to neonatal patients and often fail due to the rapid development of viral resistance [39]. As such, rational design of new therapeutics and vaccines against HCMV are urgently needed.

1.23 Parasitic Infection

A parasite is an organism that lives on or in a host organism and gets its food from or at the expense of its host. The term parasite encompasses all infectious microbes, but most commonly is used to refer to eukaryotic organisms, and this is how the term is used in this thesis. Parasites affect hundreds of millions of people worldwide, cause tremendous suffering and death, and afflict wild and domestic animals, placing an enormous burden on global public health and posing a barrier to economic growth especially in less-developed countries [40, 41]. Worldwide, several million people die each year from parasitic diseases, with the bulk of those deaths coming from malaria and parasitic protozoans, which are microscopic, one-celled organisms. The major protozoan parasites include *Plasmodium* species which cause malaria, *Entamoeba histolytica* which cause amebiasis, *Leishmania* species which cause leishmaniasis, and *Trypanosoma* species, in specific, *T. brucei* which is the etiologic agents of nagana disease in cattle and sleeping sickness in humans. No effective vaccines exist, and current treatments are old, toxic and increasingly ineffective [42].

T. brucei is one of the focus of my thesis work. Various sub-species of T. brucei are transmitted by tsetse flies of the genus Glossina and cause fatal sleeping sickness in humans and Nagana in domestic animals in sub-Saharan Africa. The flagellum is an essential and defining feature of the trypanosome cell. In this thesis work, I applied the

approach of cryoET and sub-tomogram averaging to determine the *in situ* structures of trypanosome flagellar and insights into non-planar motility of eukaryotic cells.

1.3 Thesis Outline

This thesis is organized into six chapters as follows: Chapter 1 provides the background to, and motivations behind, cryoET structures from viruses to protozoa, with a particular focus on the mechanism of infection. It also summarizes the objectives and main results of the thesis. In chapter 2, we used the virus to establish the technology of cryoET with sub-tomogram averaging and successfully determined the *in situ* structures of fusion gB revealed on HCMV by cryoET with Volta phase plate, in different functional states. Additionally, chapter 3 describes the structural basis for capsid recruitment and coat formation during HSV-1 nuclear egress. We then applied the technology to more complex situations, parasite *trypanosome brucei* involved infection, the Africa Sleeping Sickness, Chapter 4 describes the novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei* by cryoET with Volta phase plate. Chapter 5 describes the structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. Finally, Chapter 6 summarizes the work presented in Chapters 2, 3, 4, and 5, and discusses the direction of future work.

1.4 References

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

2. Turk, M. and W. Baumeister, *The promise and the challenges of cryo-electron tomography.* FEBS Lett, 2020. **594**(20): p. 3243-3261.

3. Mastronarde, D.N., *Automated electron microscope tomography using robust prediction of specimen movements.* J Struct Biol, 2005. **152**(1): p. 36-51.

4. Li, X., et al., *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* Nat Methods, 2013. **10**(6): p. 584-90.

5. Kremer, J.R., D.N. Mastronarde, and J.R. McIntosh, *Computer visualization of three-dimensional image data using IMOD.* J Struct Biol, 1996. **116**(1): p. 71-6.

6. Nicastro, D., et al., *The molecular architecture of axonemes revealed by cryoelectron tomography.* Science, 2006. **313**(5789): p. 944-8.

7. Heumann, J.M., A. Hoenger, and D.N. Mastronarde, *Clustering and variance maps for cryo-electron tomography using wedge-masked differences.* J Struct Biol, 2011. **175**(3): p. 288-99.

8. Grunewald, K., et al., *Three-dimensional structure of herpes simplex virus from cryo-electron tomography.* Science, 2003. **302**(5649): p. 1396-8.

9. Wan, W., et al., *Structure and assembly of the Ebola virus nucleocapsid.* Nature, 2017. **551**(7680): p. 394-397.

10. Mattei, S., et al., *The structure and flexibility of conical HIV-1 capsids determined within intact virions.* Science, 2016. **354**(6318): p. 1434-1437.

11. Schur, F.K., et al., *An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation.* Science, 2016. **353**(6298): p. 506-8.

12. Ke, Z., et al., *Structures and distributions of SARS-CoV-2 spike proteins on intact virions.* Nature, 2020. **588**(7838): p. 498-502.

13. Albert, S., et al., *Proteasomes tether to two distinct sites at the nuclear pore complex.* Proc Natl Acad Sci U S A, 2017. **114**(52): p. 13726-13731.

14. Beck, M., et al., *Nuclear pore complex structure and dynamics revealed by cryoelectron tomography.* Science, 2004. **306**(5700): p. 1387-90.

15. von Appen, A., et al., *In situ structural analysis of the human nuclear pore complex.* Nature, 2015. **526**(7571): p. 140-143.

16. Zhang, Y., et al., *Molecular architecture of the luminal ring of the Xenopus laevis nuclear pore complex.* Cell Res, 2020. **30**(6): p. 532-540.

17. Sorrentino, S., et al., *Toward correlating structure and mechanics of platelets.* Cell Adh Migr, 2016. **10**(5): p. 568-575.

18. Watanabe, R., et al., *The In Situ Structure of Parkinson's Disease-Linked LRRK2.* Cell, 2020. **182**(6): p. 1508-1518 e16.

19. Le Guennec, M., et al., *A helical inner scaffold provides a structural basis for centriole cohesion.* Sci Adv, 2020. **6**(7): p. eaaz4137.

20. Chang, Y., et al., *Molecular mechanism for rotational switching of the bacterial flagellar motor.* Nat Struct Mol Biol, 2020. **27**(11): p. 1041-1047.

21. von Kugelgen, A., et al., *In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer.* Cell, 2020. **180**(2): p. 348-358 e15.

22. De Rosier, D.J. and A. Klug, *Reconstruction of three dimensional structures from electron micrographs.* Nature, 1968. **217**(5124): p. 130-4.

23. Forster, F., et al., *Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography.* Proc Natl Acad Sci U S A, 2005. **102**(13): p. 4729-34.

24.   Qu, K., et al., *Structure and architecture of immature and mature murine leukemia virus capsids.* Proc Natl Acad Sci U S A, 2018. **115**(50): p. E11751-E11760.

25.   Mangala Prasad, V., T. Klose, and M.G. Rossmann, *Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus.* PLoS Pathog, 2017. **13**(6): p. e1006377.

26.   Zeev-Ben-Mordehai, T., et al., *Two distinct trimeric conformations of natively membrane-anchored full-length herpes simplex virus 1 glycoprotein B.* Proc Natl Acad Sci U S A, 2016. **113**(15): p. 4176-81.

27.   Turonova, B., et al., *In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges.* Science, 2020. **370**(6513): p. 203-208.

28.   Turonova, B., et al., *Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A.* J Struct Biol, 2017. **199**(3): p. 187-195.

29.   Himes, B.A. and P. Zhang, *emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging.* Nat Methods, 2018. **15**(11): p. 955-961.

30.   Smit, J., et al., *The S-layer of Caulobacter crescentus: three-dimensional image reconstruction and structure analysis by electron microscopy.* J Bacteriol, 1992. **174**(20): p. 6527-38.

31.   Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014. **384**(9945): p. 766-81.

32.   Hutchings, M.I., A.W. Truman, and B. Wilkinson, *Antibiotics: past, present and future.* Curr Opin Microbiol, 2019. **51**: p. 72-80.

33. Blin, K., et al., *antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.* Nucleic Acids Res, 2019. **47**(W1): p. W81-W87.

34. Zhang, L., et al., *Effect of Andrographolide and Its Analogs on Bacterial Infection: A Review.* Pharmacology, 2020. **105**(3-4): p. 123-134.

35. Kuper, H., H.O. Adami, and D. Trichopoulos, *Infections as a major preventable cause of human cancer.* J Intern Med, 2000. **248**(3): p. 171-83.

36. Barford, G., A.C. Rentz, and R.G. Faix, *Viral infection and antiviral therapy in the neonatal intensive care unit.* J Perinat Neonatal Nurs, 2004. **18**(3): p. 259-74.

37. Manicklal, S., et al., *The "silent" global burden of congenital cytomegalovirus.* Clin Microbiol Rev, 2013. **26**(1): p. 86-102.

38. Akinbami, A.A., et al., *Seroprevalence of cytomegalovirus antibodies amongst normal pregnant women in Nigeria.* Int J Womens Health, 2011. **3**: p. 423-8.

39. Chou, S., *Antiviral drug resistance in human cytomegalovirus.* Transpl Infect Dis, 1999. **1**(2): p. 105-14.

40. Langousis, G. and K.L. Hill, *Motility and more: the flagellum of Trypanosoma brucei.* Nat Rev Microbiol, 2014. **12**(7): p. 505-18.

41. Ginger, M.L., N. Portman, and P.G. McKean, *Swimming with protists: perception, motility and flagellum assembly.* Nat Rev Microbiol, 2008. **6**(11): p. 838-50.

42. Alsford, S., et al., *High-throughput decoding of antitrypanosomal drug efficacy and resistance.* Nature, 2012. **482**(7384): p. 232-6.

Chapter 2: Functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate

Zhu Si [1,2,¶], Jiayan Zhang[1,3, ¶], Sakar Shivakoti[1], Ivo Atanasov[4], Chang-Lu Tao[2], Wong H. Hui[4], Kang Zhou[4], Xuekui Yu[1], Weike Li[5], Ming Luo[5], Guo-Qiang Bi[2], Z. Hong Zhou[1,3,4,*]

Affiliations:

[1]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, California, United States of America

[2]School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, P.R. China

[3]Molecular Biology Institute, UCLA, Los Angeles, CA, USA

[4]California NanoSystems Institute, UCLA, Los Angeles, CA, USA

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

¶ These authors contribute equally to this project

*Corresponding author

Email: Hong.Zhou@ucla.edu

2.1 Abstract

Human cytomegalovirus (HCMV) enters host by glycoprotein B (gB)-mediated membrane fusion upon receptor-binding to gH/gL-related complexes, causing devastating diseases such as birth defects. Although an X-ray crystal structure of the recombinant gB ectodomain at postfusion conformation is available, the structures of prefusion gB and its complex with gH/gL on the viral envelope remain elusive. Here, we demonstrate the utility of cryo electron tomography (cryoET) with energy filtering and the cutting-edge technologies of Volta phase plate (VPP) and direct electron-counting detection to capture metastable prefusion viral fusion proteins and report the structures of glycoproteins in the native environment of HCMV virions. We established the validity of our approach by obtaining cryoET *in situ* structures of the vesicular stomatitis virus (VSV) glycoprotein G trimer (171 kD) in prefusion and postfusion conformations, which agree with the known crystal structures of purified G trimers in both conformations. The excellent contrast afforded by these technologies has enabled us to identify gB trimers (303kD) in two distinct conformations in HCMV tomograms and obtain their *in situ* structures at up to 21 Å resolution through subtomographic averaging. The predominant conformation (79%), which we designate as gB prefusion conformation, fashions a globular endodomain and a Christmas tree-shaped ectodomain, while the minority conformation (21%) has a columnar tree-shaped ectodomain that matches the crystal structure of the "postfusion" gB ectodomain. We also observed prefusion gB in complex with an "L"-shaped density attributed to the gH/gL complex. Integration of these structures of HCMV glycoproteins in multiple functional states and oligomeric forms with existing biochemical data and domain organization of other class III viral fusion proteins suggests that gH/gL receptor-binding triggers conformational changes of gB endodomain, which in turn triggers two essential steps to actuate virus-cell membrane fusion: exposure of gB fusion loops and unfurling of gB ectodomain.

2.2 Author Summary

Infection by herpesviruses leads to many human diseases, ranging from mild cold sores to devastating cancers. Human cytomegalovirus (HCMV) is among the most medically significant herpesviruses and causes birth defects and life-threatening complications in immuno-suppressed individuals. HCMV infection begins with cellular membrane fusion, a dynamic process involving receptor-binding to gH/gL complexes and drastic transformation of fusion protein gB trimer from the metastable prefusion conformation to the stable postfusion conformation. We have used cryo electron tomography incorporating cutting-edge technologies to observe the three-dimensional structures of gB and gH/gL in their native environments of HCMV particles. Visualizations of gB in both prefusion and postfusion conformations, together with structures of other class III viral fusion proteins and molecular dynamics flexible fitting (MDFF), facilitate a prediction for the structure of HCMV gB in its prefusion conformation. With the further observation of the contact of prefusion gB with gH/gL complex, the conformational changes of gB from pre- to postfusion state lead to a better understanding of herpesvirus fusion mechanism.

2.3 Introduction

Human cytomegalovirus (HCMV), a member of the *Betaherpesvirinae* subfamily of the *Herpesviridae* family, is a leading viral cause of birth defects [1, 2] and a major contributor to life-threatening complications in immunocompromised individuals. As one of the largest membrane-containing viruses, HCMV shares a common multilayered organization with all other herpesviruses, composed of an icosahedrally ordered nucleocapsid enclosing a double-stranded DNA genome, a poorly defined tegument protein layer, and a pleomorphic, glycoprotein-embedded envelope [3]. During infection, herpesviruses fuse their envelopes with cell membranes, resulting in the delivery of nucleocapsid into the cytoplasm of the host cells. This complex process requires a number of viral glycoproteins and host receptors functioning in a coordinated manner. Glycoproteins gB and gH/gL are conserved across all herpesviruses and are essential for virus entry into cells [4]. Receptor-binding to gH/gL-containing complexes—the composition of which differs among clinical and laboratory-adapted HCMV strains and across different herpesviruses [5]—triggers conformational changes of fusion protein gB, leading to fusion of the viral envelope with cell membrane [6]. This use of both a fusion protein and a receptor-binding complex for herpesvirus entry differs from many other enveloped viruses, which use a single protein for both receptor binding and membrane fusion.

Averaging up to tens of thousands of particle images by single-particle cryoEM method has resolved *in situ* structures of capsid proteins [7-9] and the capsid-associated tegument protein pp150 [10], up to atomic resolution [11]. However, such method is not applicable to the studies of herpesvirus gB and other glycoproteins due to their disorganized distribution on the pleomorphic viral envelope. Instead, the structures of gB ectodomains and various forms of gH/gL from herpes simplex virus (HSV) [12, 13], Epstein-Barr virus (EBV) [14, 15] and HCMV [16, 17] have been solved by x-ray crystallography. The gB ectodomain structures from these studies share structural similarities to other class III viral fusion proteins in their postfusion conformation [18-20]. Among these proteins, vesicular stomatitis virus (VSV) G is the only one whose ectodomain structure has

been solved for both prefusion [21] and postfusion [20] conformations, thanks to its pH-reversibility between the two conformations and amenability to crystallization at both high and low pH conditions. At pH 6.3-6.9 conditions, VSV G has also been observed to exist in monomeric forms both in solution and on virion envelope, possibly representing fusion intermediates [22, 23]. By contrast, the prefusion conformation of herpesvirus gB is metastable and its structure has been elusive (even the recent crystal structure of the full-length HSV-1 gB is also in the postfusion conformation [24]). While cryo electron tomography (cryoET) of HSV-1 virions has revealed glycoproteins on the native viral envelope, poor contrast of cryoET reconstructions makes it difficult to distinguish different glycoprotein structures and conformations [25]. Recent efforts resorted to the use of purified HSV-1 gB-decorated vesicles to visualize the prefusion gB, but its domain assignments have been controversial due to difficulties in interpreting cryoET structures with poor contrast and signal/noise ratio (SNR) [26, 27]. As a result, the mechanism underlying the complex process of receptor-triggered membrane fusion remains poorly understood for not only HCMV, but also for other herpesviruses.

Recently, the technologies of electron-counting [28, 29], energy filter and Volta phase plate (VPP) [30] have significantly improved contrast and SNR of cryoEM images and their combined use in cryoET has led to resolution of two functional states of 26S proteasome in neurons [31]. In this study, we first demonstrated the ability to distinguish prefusion and postfusion conformations of the VSV G trimer (171 kD) *in situ* by employing a combination of VPP, direct electron-counting, energy filtering and subtomographic averaging. Application of the same approach to HCMV virions has allowed us to identify different conformational states of HCMV gB (303 kD) in their native virion environments and to determine the *in situ* structure of prefusion gB at a resolution of ~21 Å. Moreover, we also observed prefusion gB forming a complex with gH/gL *in situ* for the first time. Integration of these structures and knowledge of class III viral fusion proteins has led to a working model of how conformational changes drive membrane fusion during HCMV entry into host cells.

2.4 Results

2.4.1 Establishing VPP cryoET for obtaining in situ prefusion and postfusion structures

We first established the validity of our cryoET method of combining VPP, direct electron detection, energy filtering, and subtomographic averaging by obtaining *in situ* structures of class III viral fusion proteins with known structures. Towards this end, we took advantage of the relative simplicity of VSV in having a single 57kD glycoprotein, G, on the viral envelope, with its trimeric structures known for both prefusion and postfusion conformations; and used VSV as a gold standard to validate our method. For VSV at pH=7.5, tomograms reconstructed from tilt series obtained by 300kV Titan Krios equipped with VPP, energy filter and direct electron detection show excellent contrast, enabling the visualizations of G projecting from viral envelope, the helical nucleocapsid, as well as the internal densities corresponding to polymerases L (Fig 1A and 1B). Two conformations of G are readily differentiable based on the height and shape of the ectodomain: the majority is long (12.5nm) and slim, while the minority is short (8.7nm) and fat (Fig 1B). Subtomographic averages of 337 long-form particles and 65 short-form particles from five tomograms both contain a prominent ectodomain, with the long one (~28 Å resolution) fit perfectly with the crystal structure of G ectodomain trimer in the postfusion conformation (Fig 1C~E) and the short one with that in the prefusion conformation (Fig 1G~I). Similar structures were observed in a previous electron tomography study performed on negatively stained sample [32].

Both crystal structures of G contain five domains, DI through DV, despite drastic domain arrangements (Fig 1E, 1F, 1I and 1J). The dramatically different appearances between the two conformations are primarily due to the refolding of the short loop (residue 273 to 275) in DIII, resulting in the elongation of the central helix and a taller postfusion trimer. DIII form the trimeric core in both conformations, buried in the center of the cryoET density map (Fig 1E and 1I). The other domains (DI, DII and DIV) undergo a rigid-body type rearrangement—only changing the relative orientations and locations while retaining their domain structures [21] (Fig 1F and 1J). This analysis demonstrates that our cryoET approach incorporating the three cutting-edge technologies

can distinguish the two forms of *in situ* structures of glycoprotein G and allows fitting existing domain structures of individual fusion protein into the density maps for functional interpretation.

2.4.2 Unprecedented structural details revealed by VPP cryoET of HCMV particles

Next, we applied the same strategy established above to obtain *in situ* structures of gB and its interaction with gH/gL complex. We imaged virions of the highly passaged laboratory HCMV strain AD169, taking advantage of its simplicity, as it has lost some glycoprotein genes and does not contain gH/gL/UL128/UL130/UL131A pentamers on its envelope [33]. We recorded cryoET tilt series of HCMV virions with and without VPP in a Titan Krios instrument equipped with an energy filter and a direct electron detector in electron-counting mode. Both the raw images in the tilt series and the reconstructed tomograms show significantly better contrast when VPP was used (S1 Fig, S1~4 Movies). Typical in virions obtained by high-speed centrifugation, the viral envelopes are pleomorphic and often exhibit membrane blebs likely due to mechanical stress during purification (Fig 2A). [As discussed below, such mechanical stress might also be responsible for triggering some of the "spring-loaded"/higher-energy (prefusion) gB to its lower energy ("postfusion") form, which were used as an internal control to validate our cryoET subtomographic averaging method.] In the tomograms reconstructed from the tilt series obtained with VPP (referred to as VPP tomograms) (Fig 2), three types of enveloped viral particles are readily recognized: virions with C-capsid containing densely-packed dsDNA genome (Fig 2B), non-infectious enveloped particles (NIEPs) with B-capsid containing a protein scaffold (red arrows in Fig 2A) or with empty A-capsid (cyan arrow in Fig 2A). Inside C-capsids, the dsDNA molecule occupies evenly throughout the entire interior of the capsid with the 20 Å-diameter dsDNA duplex resolved (Fig 2B)—the first time such detailed features ever observed directly by cryoET. In B-capsids, the scaffolding protein (pUL80, up to 1000 copies/capsid [34, 35]) is organized into a density sphere with an outer and inner diameter of ~700 and ~400Å, respectively. In capsids devoid of genome DNA, a portal complex for DNA translocation is visible at one of the 12 vertices of the capsid (Fig 2C). The viral

envelope is pleomorphic (Fig 2A) and its membrane resolved into two leaflets 40Å apart (Fig 2D), supporting sparsely and randomly located, and clearly identifiable glycoprotein spikes on the outer leaflet (Fig 2A and Fig 3A).

2.4.3 Identifications of gB trimers on the viral envelope

We used the following three pieces of evidence to establish the identifications of gB trimers on the viral envelope. First, among HCMV glycoproteins, gB is known to only exist as homotrimer with a combined mass of ~300 kD [36] and is the most abundant complex over 100 kD [37]. This mass is expected to occupy an estimated extracellular volume of ~300 nm^3. Among the density spikes decorating the outer leaflet of the viral membrane, only two differently shaped spikes with such volume were identified, suggesting that they might be gB trimer at different conformational states (Fig 3B). Second, the two distinctive side-view shapes—one triangular, Christmas-tree like (71%) and the other rectangular, columnar-tree like (29%) (Fig 3B)—are similar to the side-views of the cryoET reconstructions of HSV-1 gB trimers on purified vesicles in their putative prefusion and postfusion conformations, respectively [27]. Third, we performed subtomographic averaging to these two types of spikes, respectively, in order to examine them with a higher SNR. Both of the averaged models exhibit apparent three-fold symmetry with the symmetric axis perpendicular to the plane of viral membrane, despite slight distortion arising from the inherent "missing wedge problem" of electron tomography (S2 Fig). These three pieces of evidence all point to our tentative assignment of the Christmas tree-shaped and the columnar tree-shaped densities on the HCMV envelope as gB trimers in the prefusion and "postfusion" (quotation marks are used here since the conformation is not really *caused* by fusion but likely triggered by mechanical stress during virion purification with high-speed centrifugation) conformations, respectively. Indeed, as shown below, the available crystal structure of gB in the postfusion conformation matches perfectly with our final subtomographic average of the columnar tree-shaped density, further validating our assignments.

2.4.4 Subtomographic averages of the putative gB in prefusion and

"postfusion" conformations

As mentioned above, we performed subtomographic averaging to characterize the two putative gB conformations at a higher resolution. The significantly enhanced contrast afforded by imaging with VPP at a near-focus condition allowed the clear visualizations of different structures in the reconstructed tomograms. For direct comparison, we also obtained tilt series without using a VPP (referred to as non-VPP tomograms). For the latter data, we had to use a significantly larger defocus value (-4μm) to improve image contrast and record much more tilt series (28 total) in order to obtain a similar number of gB particles for subtomographic averaging due to greater difficulties in distinguishing different glycoprotein morphologies in the tomograms (S1 Fig). In addition, the use of large defocus has necessitated correction for contrast transfer function (CTF): the structure obtained without CTF correction contains phase-inverted, incorrect structure information beyond 25 Å (S4E Fig), as reflected by the broken connections between the ectodomain and the viral membrane in the absence of CTF correction (S3B, S4C and S4D Figs).

In total, 350 particles of the columnar tree-shaped and 1523 particles of the Christmas tree-shaped densities were included for subtomographic averaging. For the columnar tree-shaped structure, all particles were extracted from the VPP tomograms due to ambiguities in distinguishing its slender shape from background noise in the non-VPP tomograms. For the Christmas tree-shaped structure, 886 particles, which came from VPP tomograms, were first used and 637 particles from non-VPP tomograms eventually were also included to further improve resolution. Three-fold symmetry was imposed subsequently to improve SNR and the resolution of the averaged structures. Fourier shell correlation (FSC) analyses indicate that the resolutions for the symmetrized 3D subtomographic average of the columnar tree-shaped and Christmas tree-shaped spikes are 26 Å and 21 Å, respectively, based on the gold-standard criterion (S3A Fig).

The subtomographic average of the columnar tree-shaped spike resolves the two leaflets of the bilayer viral envelope and a prominent (161Å in height) ectodomain (Fig 3C~E, S5 Movie). The ectodomain density matches well with the crystal structure of the HCMV gB ectodomain trimer

[16] (Fig 3F~H), validating our initial assignment of the columnar tree-shaped density as gB

structure in its "postfusion" conformation and re-establishing the validity of our approach. The

subtomographic average of our putative prefusion gB densities reveals the two leaflets of the

bilayer viral envelope with prominent gB densities attached to both: a prominent ectodomain

attached to the outer leaflet (130Å in height) and a globular (about 35Å in height and 26Å in width)

endodomain to the inner leaflet (Fig 3I~K). The ectodomain in the putative prefusion gB is shorter

than that in the gB "postfusion" conformation and anchors to the membrane with three well-

separated densities, forming a tripod (Fig 3I~K, S6 Movie). Although no crystal structure of

prefusion gB is available to fit into our subtomographic average to directly confirm or refute this

prefusion gB assignment, it is believed that herpesvirus gB bears structural and mechanistic

similarities to other class III viral fusion proteins, which can be used to aid our assignment. Indeed,

the postfusion conformation of HCMV gB ectodomain is similar to the postfusion conformations of

all other class III viral fusion proteins [18], including the postfusion VSV G (Fig 1C-F). The lower

portion of the prefusion conformation of the VSV G trimer (Fig 1G-I) has a tripod shape similar to

the lower portion of the Christmas tree-shaped density (Fig 3I). The prefusion VSV G trimer is

shorter than—and undergoes drastic domain rearrangements towards—its postfusion

conformation [20, 21] (Fig 1); likewise, the Christmas tree-shaped density is shorter than the

columnar tree-shaped density. Taken together, these characteristic similarities to the prefusion

structure of VSV G corroborate our initial assignment of the Christmas tree-shaped density as the

in situ prefusion structure of HCMV gB trimer.

2.4.5 Domain assignments of the in situ gB structures in both conformations

Structure-guided sequence analysis (Fig 4A) indicates that each full-length gB protomer contains

an N-terminal ectodomain (residues 87-705), a membrane proximal region (MPR, residues 706-

750), a single transmembrane helix (residues 751-771) and a C-terminal endodomain (residues

772-906). For the "postfusion" gB trimer, the ectodomain in the subtomographic average can be

divided into a base in contact with the membrane, and two lobes—middle and crown—connected by a neck (Fig 3F). The crystal structure of the ectodomain trimer shows that each protomer consists of five domains: DI, DII, DIII, DIV and DV (Fig 3G and 3H) [16]. Except for DV, these domains can be located in our subtomographic average of the "postfusion" gB (Fig 3F). DI, each containing two fusion loops, is located at the base of the trimer; DII and DIV reside, respectively, in the middle and crown lobes, which are connected by DIII in the neck. DV contains a long loop connected by two short helices and is buried, thus is not resolved in our subtomographic average gB trimer due to the limited resolution.

As detailed in the Method, we employed a combination of manual rigid-body fitting of known domain structures from the existing HCMV gB postfusion structure [16], comparative modeling of DIII based on the homologous DIII from VSV G prefusion conformation [21], followed by optimization by the molecular dynamics flexible fitting (MDFF) method [38], to put forward a provisional domain arrangement model of the prefusion gB (Fig 5). DV was not considered in our domain modeling of HCMV gB prefusion conformation due to the lack of a template structure, since DV was truncated in the crystal structure of postfusion VSV G. MDFF not only optimized the chemical interactions among the fitted domains, but also improved overall model to map correlation coefficient from 0.83 to 0.94 (Fig 5H and 5I, S7 Movie).

The model from MDFF does not include the MPR (residues 706-750), which is proposed to lie between the ectodomain and the transmembrane helix (Fig 4A) and "mask" the fusion loops to prevent their premature (non-productive) association with lipid [39]. Helical wheel projection of the first 15 amino acids of the MPR shows an amphipathic helix (Fig 4B) whose hydrophobic side could interact with the fusion loops. This notion is consistent with our interpretation of DI in the subtomographic averages of both prefusion and "postfusion" conformations, with the fusion loops pointing to and in close proximity to the membrane.

2.4.6 Visualization of gB interacting with putative receptor-binding gH/gL complex

Among herpesviruses, gB and gH/gL are highly conserved and known to form a fusion machinery for virus entry [43]. Previous biochemical studies have indicated that gH/gL regulated fusion activity of gB [44] and might form a complex with gB in virions on the basis of co-immunoprecipitation experiments [45]. Besides gB trimer densities mentioned above, "L"-shaped spikes were also observed protruding outwards from the viral envelope, which we interpret as gH/gL complexes on the basis of size and shape similarities to the gH/gL crystal structure [13, 17]. Moreover, among such "L"-shaped spikes, ~7% were observed to be in contact with the Christmas tree-shaped, prefusion gB trimer, forming a gB-gH/gL complex (Fig 6B and 6C), while others were unbound. No "postfusion" gB trimer have been observed involving in gB-gH/gL complex.

A subtomographic average was obtained by aligning and averaging 49 such gB-gH/gL complexes to investigate the contact sites between prefusion gB and gH/gL (Fig 6C~F), with a resolution around 30Å reported by *calcFSC* in *PEET*. The HCMV gH/gL crystal structure [17] fits well in the "L"-shaped density in the subtomographic average (0.75 of the cross-correlation coefficient between the cryoET map and the model filtered to 30Å, Fig 6D and 6E). This fitting, together with the predicted domain arrangement in the prefusion gB structure (Fig 5H and 5I), reveals that DI of gB may contact the gH subunit of gH/gL (Fig 6D). The contact sites on gB and gH are consistent with the gH-binding site on HSV-1 gB suggested by blocking gH binding to gB with SS55 and SS56 antibodies (epitopes mapped to residues 153-363 of gB) (Fig 5H) [46] and the gB-binding sites on gH/gL suggested by anti-gH/gL antibody LP11 for HSV [13], respectively. Mutagenesis of gH cytotail has led to its proposed role of acting as a "wedge" to split the gB endodomain "clamp" to trigger gB ectodomain refolding [40]. Though the details of their interactions in the endodomain are yet to be resolved, this first observation of gH/gL complex making contact with prefusion gB *in situ* (Fig 6) supports the notion that receptor binding to gH/gL triggers transformation of gB from prefusion to postfusion conformation.

2.5 Discussion

Since the postfusion conformation of gB is energetically favorable and structurally more stable, it is

not surprising that purified recombinant gB so far have all adopted the "postfusion" conformation [12, 16, 47]. Therefore, imaging gB in its native, virion environment by cryoET seems to be the necessary approach to obtain the *in situ* structure in its metastable, prefusion conformation. However, a major challenge in interpreting *in situ* cryoET structures is the intrinsic poor contrast of tomographic reconstructions due to the use of low electron dose in order to avoid radiation damage to specimen. Poor contrast makes it difficult to identify different molecules or structures for subtomographic averaging. Normally for cellular tomography without phase plate, one could image with a large defocus value to achieve better contrast, aiding in distinguishing densities with different characteristics for subtomographic averaging. However, such approach only offers limited improvements in contrast (S1 Fig), and difficulties still exist in identifying the slender gB in postfusion conformation in our tomograms. This experience is consistent with two previous cryoET studies on HSV-1 gB structures, in which large defocus values were used to increase contrast to facilitate subsequent subtomographic averaging, yet the resulting structure either is at much lower resolution [26] than reported here or has led to controversial interpretations [27]. The greatly improved contrast afforded by VPP technology allowed the differentiation of various glycoprotein structures based on their characteristic appearances on the virion membrane (Fig 7A and 7B; S1 Fig). Therefore, cryoET with VPP offers a clear advantage in resolving structures of proteins in the native environments, enabling their identifications and subtomographic averaging to obtain structures of multi-functional states, as also demonstrated by the existence of two states of 26S proteasome inside neurons [31].

A vital step of herpesvirus infections is the fusion of viral and cell membranes, a complicated process involving at least three conserved proteins—gB, gH and gL. The *in situ* structures of gB at both prefusion and "postfusion" conformations reported here can shed lights on conformational changes of gB during membrane fusion and inform how herpesvirus entry into cell (Fig 7). Prior to fusion, gB needs to be maintained at its inactive, metastable prefusion conformation (Fig 7A). The maintenance of this metastable conformation possibly involves a

properly-folded endodomain of gB, since removal of the endodomain caused gB ectodomain to adopt the postfusion conformation [42]. In addition, the direct observation in our cryoET structure of gB-gH/gL complex (Fig 6) and its isolation by co-immunoprecipitation [45] both suggest that the metastable ectodomain of gB might also be stabilized through the interaction with the ectodomain of gH subunit (Fig 7C). Host receptor-binding to gH/gL complex would trigger a conformational change in gH/gL cytotail and its dissociation from, and the destabilization of, the endodomain of gB, which in turn triggers the massive conformational changes of gB ectodomain to expose its fusion loops (step 1). Subsequently, DIII central helix extends, unfurling other domains and swinging the fusion loops to engage with the host membrane (step 2). Facilitated by the intrinsic fluidity in the plane of the membrane, the refolding of gB domains to the lower-energy, postfusion conformation, in which its ectodomain C-terminal end and the fusion loops must come together, leads to fusion of the two membranes and the release of viral DNA-containing capsid into cytoplasm (step 3). In the absence of receptor binding as in the situation of this study, mechanical stress to the membrane caused by such means as high-speed centrifugation could also destabilize the membrane-associated endodomain, triggering metastable prefusion gB to undergo the cascade of transformation events, possibly accompanied by the exposure of the fusion loops (step 1). Lacking host cell membrane, these events, with exposed fusion loops eventually encountering and inserting its hydrophobic moieties into the viral membrane, will be followed by refolding of other domains into the stable, "postfusion" conformation (step 3). Notably, the topology of the conformational change during step 2 to step 3 would preclude transiting from prefusion to postfusion conformation without breaking the three-fold symmetry. Indeed, monomeric intermediates of VSV G have been observed both in solution and on the surface of virions at intermediate pH conditions [22, 23].

In our model, the fusion loops of prefusion gB point to and are in close proximity to the viral membrane, possibly buried within a hydrophobic "mask" of MPR, which is attached to the C-terminal end of the gB ectodomain crystal structure. This membrane-proximal location of the gB

fusion loops is the same as that based on the cryoET structure of the HSV-1 gB/anti-fusion loop 2-antibody at 5nm resolution [26] and is consistent with the fusion loop locations in all known atomic structures of classes I and III viral fusion proteins, including influenza HA [48], HIV env trimer [49], VSV G [21] and others [6]. Notably, our model is in stark contrast to the exposed fusion loops assigned to the membrane-distal tips of the "short-form" HSV-1 gB structures [27], which were obtained by cryoET of purified gB-containing vesicles. The ectodomain of the "short-form" vesicular HSV-1 gB structure is 15% shorter in height and 23% wider in diameter than that of our *in situ* HCMV gB structure, despite both sharing the Christmas tree shape (S5 Fig). Superposition of the domain assignment obtained by the hierarchical fitting approach [27] into the "short-form" HSV-1 gB structure shows that the densities projecting from the lower whorl of the Christmas tree-shaped trimer were unaccounted for (S5B Fig). Moreover, placing the same domain assignment into our *in situ* HCMV gB prefusion structure reveals that the fusion loops in this assignment are projecting out of the cryoET map, yet the leader density of the map is not accounted for (S5C Fig). When filtering the pseudoatomic model to 25Å, the cross-correlation coefficient is 0.74, as compared to 0.93 of our prefusion structure. We believe that an exposed fusion loop orientation of prefusion gB is unlikely for both chemical and biological reasons—exposed hydrophobic moieties are chemically unfavorable in solution and can lead to unproductive membrane insertion during infection. Indeed, the "short-form" HSV-1 gB structure was cautiously interpreted as an ambiguous "prefusion and/or intermediate" conformation [27], probably to reconcile these contradictory considerations.

Secondary structure prediction indicates that the endodomain is helix-rich (~50%) (Fig 4A). Our results suggest that gB endodomain undergoes significant conformational changes, from prominently visible/stable in the prefusion structure (Fig 3I and 3K), to invisible/flexible in the "postfusion" structure (Fig 3C and 3E). Proteolysis and circular dichroism analyses of the endodomain of the highly homologous HSV-1 gB posit that gB endodomain clamps the viral membrane and stabilizes gB in its prefusion conformation [40, 41]. This proposed model is

supported by studies on truncation and substitution mutations in endodomain [40, 42]. The

structured endodomain resolved in the recent crystal structure of full-length gB was thought to be

similar to that in prefusion gB [24]. Detergent solubilization of the membrane may be responsible

for the postfusion conformation of its ectodomain. Our observation of the endodomain structure of

HCMV gB changing from a stable, prefusion conformation (Fig 3I) to a flexible, postfusion

conformation (Fig 3E) is consistent with its proposed role in stabilization of gB prefusion

conformation on native viral membrane [24].

2.6 Materials and Methods

2.6.1 HCMV virion preparation

Human fibroblast MRC-5 cells (ATCC® CCL-171™) were cultured in Eagle's Minimum Essential

Medium (EMEM, ATCC® 30-2003™) with 10% fetal bovine serum (FBS, Omega scientific: FB-

11). Cells were grown in T-175 cm2 flasks to 90% confluence and infected with HCMV strain

AD169 (ATCC, Rockville, MD) at a multiplicity of infection (MOI) of 0.1–0.5, and incubated for

about 7 days. Once the cells showed 100% cytopathic effect, the media were collected and

centrifuged at 10,000 g for 15 min to remove cells and large cell debris. The clarified supernatant

was collected and centrifuged at 60, 000 g for 1 hour to pellet HCMV virions. Pellets were

resuspended in 20mM phosphate buffered saline (PBS, pH 7.4), loaded on a 15%–50% (w/w)

sucrose density gradient, and centrifuged at 60,000 g for 1 hr. After the density gradient

centrifugation, three light-scattering bands were observed in the density gradient: top, middle and

bottom. The middle band contained both HCMV virions and NIEPs (particles with intact viral

envelopes as judged by negative-staining EM) and was collected, diluted in PBS and then

centrifuged at 60,000 g for 1 hour. The final pellet was resuspended in PBS for further cryoET

sample preparation.

2.6.2 VSV virion preparation

VSV virion (Indiana serotype, San Juan strain) samples were produced as previously described [50]. Particularly, the inoculum was passaged multiple times in Hela cells with a very low multiplicity of infection (MOI), 0.001, to suppress the truncated defective-interference particles. The full VSV particles were isolated in a sucrose gradient and the final inoculum was also plaque-purified in Hela cells. We then pelleted the VSV virions at 30,000g for 2 hours and resuspended them in PBS. The stock was subjected to another low speed centrifugation at 12,000g for 5min in a desktop centrifuge to remove large aggregates. After resuspension, the pellets were banded on a 10ml density gradient containing 0-50% potassium tartrate and 30-0% glycerol. The virions-containing band was collected, diluted in PBS, pelleted at 30,000g for 2 hours, resuspended in PBS and kept in 4°C refrigerator for further cryoET sample preparation.

2.6.3 CryoET sample preparation and data collection

An aliquot of 2.5 μl of the sample mixed with 5-nm diameter gold beads were applied onto freshly glow-discharged Quantifoil Holey Carbon Grids. Grids were blotted and plunge-frozen in liquid ethane cooled by liquid nitrogen using an FEI Mark IV Vitrobot cryo-sample plunger and were stored in liquid nitrogen before subsequent usage. CryoEM imaging and cryoET tilt series acquisition were performed with *SerialEM* [51] on an FEI Titan Krios 300kV transmission electron microscope equipped with a Gatan imaging filter (GIF), a Gatan K2 Summit direct electron detector, and with or without a Volta phase plate (VPP). Tilt series were recorded by tilting the specimen covering the angular range of -66° to +60° (starting tilt from -48° to +60°, then from -50° to -66°) with 2° or 3° interval, with a nominal magnification of x53,000 (corresponding to a calibrated pixel size of 2.6 Å) and a cumulative electron dose of 100~110 e⁻/Å². Exposure time was multiplied by a factor of the square root of 1/cosα (in which α=tilt angle), and the exposure time at 0° was set at 1.2s for the tilt step-size of 2° or 1.6s for the tilt step-size of 3°. Movies were recorded with the frame rate of 0.2 frame/s on a Gatan K2 Summit direct electron detector operated in counting mode with the dose rate of 8-10 e⁻/pixel/s. An energy filter slit of 20 eV was

chosen for the GIF. For imaging with VPP, defocus value was targeted at -0.6µm. Note, one of the

benefits of using a phase plate is that the CTF is insensitive to the sign of the defocus value being

negative (underfocus) or positive (overfocus) [52]. VPP was advanced to a new position every tilt

series, followed by a 2 min waiting for stabilization, and pre-conditioned by electron illumination

with a total dose of 12 nC for 60s to achieve a phase shift of ~54° as previously described [53]. For

tilt series obtained without VPP, the defocus value was maintained at around -4µm while other

imaging parameters were kept the same as those for the tilt series with VPP.

2.6.4 3D reconstruction

Frames in each movie of the raw tilt series were aligned, drift-corrected and averaged with

Motioncorr [54] to produce a single image for each tilt angle. Both sets of tilt series, collected with

and without VPP, were reconstructed with *IMOD* 4.8 software package [55] in the following six

steps. All images in a tilt series were coarsely aligned by cross-correlation (step 1) and then finely

aligned by tracking selected gold fiducial beads (step 2). The positions of each bead in all images

of the tilt series were fitted into a specimen-movements mathematical model, resulting in a series

of predicted positions. The mean residual error (mean distance between the actual and predicted

positions) was recorded to facilitate bead tracking and poorly-modeled-bead fixing (step 3). With

the boundary box reset and the tilt axis readjusted (step 4), images were realigned (step 5).

Finally, two tomograms were generated by weighted back projection and simultaneous iterative

reconstruction technique (SIRT) method, respectively (step 6). For data collected without VPP,

contrast transfer function (CTF) was corrected with the *ctfphaseflip* program [56] of IMOD in step5.

The defocus value for each image in one tilt series was determined by *CTFTILT* [57], and the

estimated defocus value of each image was used as input for *ctfphaseflip*.

2.6.5 Subtomographic averaging

Subtomographic averaging was performed using *PEET* 1.11 [58, 59]. High contrast SIRT

tomograms were 4× binned by the *binvol* program of *IMOD* to facilitate particle picking. Particles

were picked manually in *IMOD* as follows. For distinct conformations of VSV G and HCMV gB on viral envelope, two points (*head* and *tail*) in one contour were used to define one particle (glycoprotein) —*head* is the membrane-proximal end of the protrusion density while *tail* is the membrane-distal end. An initial *motive* list file, a RotAxes file and three model files containing the coordinates of *head*, *centroid* and *tail* for each particle were generated by *stalkInit* in *PEET*. In total, we manually picked 337 long-form particles from 5 VPP tomograms of VSV, and 350 columnar tree-shaped particles and 886 Christmas tree-shaped particles from 11 VPP tomograms of HCMV. Besides, 637 Christmas tree-shaped particles were picked from 28 non-VPP tomograms, averaged either alone or together with those from the VPP tomograms for prefusion gB.

For the reconstruction of the long-form VSV G, subtomographic average was performed first with 4× binned SIRT tomograms using the sum of all particles as the initial reference. Through *stalkInit*, particle tilt orientations (axes normal to the membrane) were already coarsely aligned to Y axis, but twist orientations (angle around symmetry axis) were randomized. Therefore, in the first refinement cycle, we set the angular search range 180° max (-180°~180°) with 9° step in Phi (Y axis), and 5° (-5°~5°) max with 1° step in both Theta (Z axis) and Psi (X axis), and search distance 3 pixels along three axes. Due to the known symmetry of postfusion VSV G, the resulting averaged structure was then trimerized and used as the reference of the next refinement cycle. The trimerized structure was the sum of each refined particle and its two symmetrical copies—the two symmetrical copies have the same position and tilt orientation as the refined particle, but twist orientation differed by either 120° or 240°. For subsequent refinement cycles, reference was updated from the last refinement cycle (trimerized), with both angular and distance search ranges narrowing down gradually. After four refinement cycles, the averaged structure converged based on no further improvements in resolution. The following refinement cycles were performed with 2× binned tomograms reconstructed by weighted back projection, after up-sampling (generations of 2× binned model files and updates of corresponding motive list files from the latest refinement

cycle), with small search distance range (4 pixels) and narrow angular search range (-20°~20°). Same as the refinement cycles for the 4× binned tomograms, the reference was updated from the averaged structure of the last refinement cycle (trimerized). For particles with distance of <1 pixel and twist angle difference of <1°, the one representative with lower cross-correlation coefficient was treated as duplicate particle and removed during the refinement. The averaged structure, contributed by 330 particles, converged after eight refinement cycles and was filtered to the final resolution, calculated by *calcFSC* in *PEET* based on the 0.143 FSC criterion.

Reconstructions of columnar tree-shaped and Christmas tree-shaped particles on HCMV envelope followed the same refinement procedure as the reconstruction of long-form VSV G, except that the averaged structure was trimerized after the three-fold symmetry was strong. With the removal of duplicate particles, the final averaged structures of the two conformations were contributed by 350 particles and 1509 particles, respectively. Furthermore, gold-standard FSC calculations for both resulting structures were performed afterwards by splitting each original dataset into two independent groups, respectively. The same refinement procedure used above was applied to the two newly-generated groups independently. Upon the convergence of the both averaged structures, FSC were calculated by *calcUnbiasedFSC* in *PEET* (S3A Fig.).

For the reconstruction of the short-form VSV G, 65 particles were manually picked from five tomograms with single point to define the centroid position. Each particle was manually rotated around X, Y, Z axes to a similar orientation (both the tilt orientation and twist angle) in *IMOD* slicer window. By *slicer2MOTL* in *PEET*, the initial motive list files for subtomographic averaging were generated from the corresponding X, Y, Z rotation degrees. For the Angular Search Range, small search range was set during all seven refinement cycles. The final subtomographic average was Gaussian filtered with width 7 using the "volume filter" tool in UCSF *Chimera* [60]. Due to the limited number of particles (49 particles), HCMV gB-gH/gL complex was reconstructed with the same strategy above.

2.6.6 3D Visualization

We used *IMOD* [61] to visualize reconstructed tomograms and UCSF *Chimera* to visualize the

subtomographic averages in three dimensions. The crystal structures of prefusion VSV G (PDB:

5I2S) [21], postfusion VSV G (PDB: 5I2M) [20], HCMV postfusion gB (PDB: 5CXF) [16] and gH/gL

part from HCMV pentamer (PDB: 5VOB) [17] were fitted into subtomographic averages of

prefusion G, postfusion G, postfusion gB and gB-gH/gL complex, respectively, with the tool *fit in*

map in *Chimera*. Segmentation and surface rendering for the membrane and tegument proteins

were done by the tools *volume tracer* and *color zone* in *Chimera*. All membrane glycoproteins

were placed back on the viral membrane according to their locations in the original tomogram. A

published structure of HCMV capsid with inner tegument protein [11] was filtered to 10 Å and

placed back at the same position of the capsid in tomogram.

2.6.7 Domain modeling and structure prediction

As outlined below, we employed a combination of initial manual fitting of known domain structures,

followed by simulation with MDFF program [38] to generate a gB prefusion model based on our

cryoET prefusion gB trimer density map and the existing gB ectodomain postfusion crystal

structure (PDB: 5CXF) [16]. First, the ectodomain in the subtomographic averaged density map of

prefusion gB trimer was segmented out and its symmetric axis obtained with *Chimera*'s "volume

eraser" tool and "measure symmetry" command, respectively. Second, *Chimera*'s "fitmap"

command with "global search" and 15Å-resolution options was used to refine 1000 initial random

DI placements, resulting in 28 refined fitted positions, each with a correlation coefficient (between

the fitted model and the density map) and a "clash volume fraction" value (between symmetry-

related copies). We chose the fitted position with the largest fitting score, defined as the correlation

coefficient subtracted by the "clash volume fraction" penalty value (Fig. 5A). Third, we obtained our

initial DIII by computationally mutating the DIII model from the existing hypothetic model of EBV

prefusion gB [14], as it is known to differ substantially from its postfusion conformation for both

herpesvirus gB [14, 62] and homologous VSV G [21]. Compared to that in the postfusion gB, the

central helix α4 in DIII in the prefusion gB is bent in order to fit into the top of the Christmas tree-shaped density. This bent varies from only ~30° in our proposed HCMV gB prefusion structure to ~90° in VSV G (Fig. 5D) [21] and ~180° in influenza HA [48] and HIV env [49]. This DIII model, and the models of DII and DIV from the gB postfusion crystal structure were manually fitted as rigid bodies into our prefusion gB trimer cryoET density to produce a composite model with the above obtained DI trimer model by referencing the prefusion VSV G crystal structure. Connecting loops were then added to this composite model through the *Modloop* server [63]. Fourth, the resulting trimer model was used as the initial model for MDFF simulations [38] with grid force scale of 0.3. Secondary structure, *cis* peptide and chirality restraints were imposed during MDFF simulations. Simulations were performed with NAMD 2.12 [64], using the CHARMM36 force field with CMAP corrections [65].

Secondary structures for residues 707-906 of gB were predicted with *Phyre2* [66].

2.7 Acknowledgements

2.8 References

1.   Vancikova Z, Dvorak P. Cytomegalovirus infection in immunocompetent and immunocompromised individuals--a review. Curr Drug Targets Immune Endocr Metabol Disord. 2001;1(2):179-87.

2.   Adler SP. Congenital cytomegalovirus screening. Pediatr Infect Dis J. 2005;24(12):1105-6.

3.   Gibson W. Structure and Formation of the Cytomegalovirus Virion. In: Shenk TE, Stinski MF, editors. Human Cytomegalovirus. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 187-204.

4.   Vanarsdall AL, Johnson DC. Human cytomegalovirus entry into cells. Current opinion in virology. 2012;2(1):10.1016/j.coviro.2012.01.001.

5.   Heldwein EE. gH/gL supercomplexes at early stages of herpesvirus entry. Curr Opin Virol. 2016;18:1-8.

6.   Harrison SC. Viral membrane fusion. Virology. 2015;479-480:498-507.

7.   Chen DH, Jiang H, Lee M, Liu F, Zhou ZH. Three-Dimensional Visualization of Tegument/Capsid Interactions in the Intact Human Cytomegalovirus. Virology. 1999;260(1):10-6.

8.   Yu X, Trang P, Shah S, Atanasov I, Kim YH, Bai Y, et al. Dissecting human cytomegalovirus gene function and capsid maturation by ribozyme targeting and electron cryomicroscopy. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(20):7103-8.

9.   Trus BL, Gibson W, Cheng N, Steven AC. Capsid structure of simian cytomegalovirus from cryoelectron microscopy: evidence for tegument attachment sites. Journal of virology. 1999;73(3):2181-92.

10.   Dai X, Yu X, Gong H, Jiang X, Abenes G, Liu H, et al. The smallest capsid protein mediates binding of the essential tegument protein pp150 to stabilize DNA-containing capsids in human cytomegalovirus. PLoS pathogens. 2013;9(8):e1003525.

11.   Yu X, Jih J, Jiang J, Zhou ZH. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. Science (New York, NY). 2017;356(6345):10.1126/science.aam6892.

12.   Heldwein EE, Lou H, Bender FC, Cohen GH, Eisenberg RJ, Harrison SC. Crystal structure of glycoprotein B from herpes simplex virus 1. Science (New York, NY). 2006;313(5784):217-20.

13.   Chowdary TK, Cairns TM, Atanasiu D, Cohen GH, Eisenberg RJ, Heldwein EE. Crystal structure of the conserved herpesvirus fusion regulator complex gH-gL. Nature structural & molecular biology. 2010;17(7):882-8.

14.   Backovic M, Longnecker R, Jardetzky TS. Structure of a trimeric variant of the Epstein-Barr

virus glycoprotein B. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(8):2880-5.

15.   Matsuura H, Kirschner AN, Longnecker R, Jardetzky TS. Crystal structure of the Epstein-Barr virus (EBV) glycoprotein H/glycoprotein L (gH/gL) complex. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(52):22641-6.

16.   Burke HG, Heldwein EE. Crystal Structure of the Human Cytomegalovirus Glycoprotein B. PLoS pathogens. 2015;11(10):e1005227.

17.   Chandramouli S, Malito E, Nguyen T, Luisi K, Donnarumma D, Xing Y, et al. Structural basis for potent antibody-mediated neutralization of human cytomegalovirus. Science Immunology. 2017;2(12).

18.   Backovic M, Jardetzky TS. Class III viral membrane fusion proteins. Current opinion in structural biology. 2009;19(2):189-96.

19.   Kadlec J, Loureiro S, Abrescia NGA, Stuart DI, Jones IM. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. Nature structural & molecular biology. 2008;15(10):1024-30.

20.   Roche S, Bressanelli S, Rey FA, Gaudin Y. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science (New York, NY). 2006;313(5784):187-91.

21.   Roche S, Rey FA, Gaudin Y, Bressanelli S. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science (New York, NY). 2007;315(5813):843-8.

22.   Albertini AA, Merigoux C, Libersou S, Madiona K, Bressanelli S, Roche S, et al. Characterization of monomeric intermediates during VSV glycoprotein structural transition. PLoS pathogens. 2012;8(2):e1002556.

23.   Baquero E, Albertini AA, Raux H, Abou-Hamdan A, Boeri-Erba E, Ouldali M, et al. Structural intermediates in the fusion-associated transition of vesiculovirus glycoprotein. The EMBO journal. 2017;36(5):679-92.

24.   Cooper RS, Georgieva ER, Borbat PP, Freed JH, Heldwein EE. Structural basis for

membrane anchoring and fusion regulation of the herpes simplex virus fusogen gB. Nature

structural & molecular biology. 2018;25(5):416-24.

25.	Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, et al. Three-

dimensional structure of herpes simplex virus from cryo-electron tomography. Science (New York,

NY). 2003;302(5649):1396-8.

26.	Fontana J, Atanasiu D, Saw WT, Gallagher JR, Cox RG, Whitbeck JC, et al. The Fusion

Loops of the Initial Prefusion Conformation of Herpes Simplex Virus 1 Fusion Protein Point Toward

the Membrane. mBio. 2017;8(4).

27.	Zeev-Ben-Mordehai T, Vasishtan D, Hernandez Duran A, Vollmer B, White P, Prasad

Pandurangan A, et al. Two distinct trimeric conformations of natively membrane-anchored full-

length herpes simplex virus 1 glycoprotein B. Proceedings of the National Academy of Sciences of

the United States of America. 2016;113(15):4176-81.

28.	Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by

electron cryo-microscopy. Nature. 2013;504(7478):107-12.

29.	Jiang J, Pentelute BL, Collier RJ, Zhou ZH. Atomic structure of anthrax protective antigen

pore elucidates toxin translocation. Nature. 2015;521(7553):545-9.

30.	Khoshouei M, Radjainia M, Baumeister W, Danev R. Cryo-EM structure of haemoglobin at

3.2 A determined with the Volta phase plate. bioRxiv. 2016.

31.	Asano S, Fukuda Y, Beck F, Aufderheide A, Forster F, Danev R, et al. Proteasomes. A

molecular census of 26S proteasomes in intact neurons. Science (New York, NY).

2015;347(6220):439-42.

32.	Libersou S, Albertini AA, Ouldali M, Maury V, Maheu C, Raux H, et al. Distinct structural

rearrangements of the VSV glycoprotein drive membrane fusion. The Journal of cell biology.

2010;191(1):199-210.

33.	Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR. Human

cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. Journal of

virology. 1996;70(1):78-83.

34.   Gibson W. Structure and formation of the cytomegalovirus virion. Current topics in microbiology and immunology. 2008;325:187-204.

35.   Newcomb WW, Trus BL, Booy FP, Steven AC, Wall JS, Brown JC. Structure of the herpes simplex virus capsid. Molecular composition of the pentons and the triplexes. J Mol Biol. 1993;232(2):499-511.

36.   Sharma S, Wisner TW, Johnson DC, Heldwein EE. HCMV gB shares structural and functional properties with gB proteins from other herpesviruses. Virology. 2013;435(2):239-49.

37.   Varnum SM, Streblow DN, Monroe ME, Smith P, Auberry KJ, Pasa-Tolic L, et al. Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. Journal of virology. 2004;78(20):10960-6.

38.   Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods (San Diego, Calif). 2009;49(2):174-80.

39.   Shelly SS, Cairns TM, Whitbeck JC, Lou H, Krummenacher C, Cohen GH, et al. The membrane-proximal region (MPR) of herpes simplex virus gB regulates association of the fusion loops with lipid membranes. mBio. 2012;3(6).

40.   Rogalin HB, Heldwein EE. Interplay between the Herpes Simplex Virus 1 gB Cytodomain and the gH Cytotail during Cell-Cell Fusion. Journal of virology. 2015;89(24):12262-72.

41.   Silverman JL, Greene NG, King DS, Heldwein EE. Membrane requirement for folding of the herpes simplex virus 1 gB cytodomain suggests a unique mechanism of fusion regulation. Journal of virology. 2012;86(15):8171-84.

42.   Vitu E, Sharma S, Stampfer SD, Heldwein EE. Extensive mutagenesis of the HSV-1 gB ectodomain reveals remarkable stability of its postfusion form. J Mol Biol. 2013;425(11):2056-71.

43.   Connolly SA, Jackson JO, Jardetzky TS, Longnecker R. Fusing structure and function: a structural view of the herpesvirus entry machinery. Nature reviews Microbiology. 2011;9(5):369-81.

44.   Heldwein EE, Krummenacher C. Entry of herpesviruses into mammalian cells. Cellular and molecular life sciences : CMLS. 2008;65(11):1653-68.

45.   Vanarsdall AL, Howard PW, Wisner TW, Johnson DC. Human Cytomegalovirus gH/gL Forms a Stable Complex with the Fusion Protein gB in Virions. PLoS pathogens. 2016;12(4):e1005564.

46.   Atanasiu D, Whitbeck JC, de Leon MP, Lou H, Hannah BP, Cohen GH, et al. Bimolecular complementation defines functional regions of Herpes simplex virus gB that are involved with gH/gL as a necessary step leading to cell fusion. Journal of virology. 2010;84(8):3825-34.

47.   Cooper RS, Georgieva ER, Rogalin HB, Borbat PP, Freed JH, Heldwein EE. Fusion on a Pedestal: The Structure of the Full-Length HSV-1 Fusogen gB. Biophysical Journal.112(3):188a.

48.   Lin YP, Xiong X, Wharton SA, Martin SR, Coombs PJ, Vachieri SG, et al. Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. Proceedings of the National Academy of Sciences. 2012;109(52):21474-9.

49.   Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science (New York, NY). 2013;342(6165):1477-83.

50.   Holland JJ, Villarreal LP, Breindl M. Factors involved in the generation and replication of rhabdovirus defective T particles. Journal of virology. 1976;17(3):805-15.

51.   Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. Journal of structural biology. 2005;152(1):36-51.

52.   Fan X, Zhao L, Liu C, Zhang JC, Fan K, Yan X, et al. Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM. Structure. 2017;25(10):1623-30 e3.

53.   Fukuda Y, Laugks U, Lucic V, Baumeister W, Danev R. Electron cryotomography of vitrified cells with a Volta phase plate. Journal of structural biology. 2015;190(2):143-54.

54.   Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.

Nature methods. 2013;10(6):584-90.

55.   Mastronarde DN. Dual-Axis Tomography: An Approach with Alignment Methods That Preserve Resolution. Journal of structural biology. 1997;120(3):343-52.

56.   Xiong Q, Morphew MK, Schwartz CL, Hoenger AH, Mastronarde DN. CTF Determination and Correction for Low Dose Tomographic Tilt Series. Journal of structural biology. 2009;168(3):378-87.

57.   Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology. 2003;142(3):334-47.

58.   Nicastro D, Schwartz C, Pierson J, Gaudette R, Porter ME, McIntosh JR. The molecular architecture of axonemes revealed by cryoelectron tomography. Science (New York, NY). 2006;313(5789):944-8.

59.   Heumann JM, Hoenger A, Mastronarde DN. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. Journal of structural biology. 2011;175(3):288-99.

60.   Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004;25(13):1605-12.

61.   Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. Journal of structural biology. 1996;116(1):71-6.

62.   Gallagher JR, Atanasiu D, Saw WT, Paradisgarten MJ, Whitbeck JC, Eisenberg RJ, et al. Functional fluorescent protein insertions in herpes simplex virus gB report on gB conformation before and after execution of membrane fusion. PLoS pathogens. 2014;10(9):e1004373.

63.   Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. Bioinformatics (Oxford, England). 2003;19(18):2500-1.

64.   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry. 2005;26(16):1781-802.

65.   Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. Journal of computational chemistry. 2013;34(25):2135-45.

66.   Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocols. 2015;10(6):845-58.

2.9 Figures
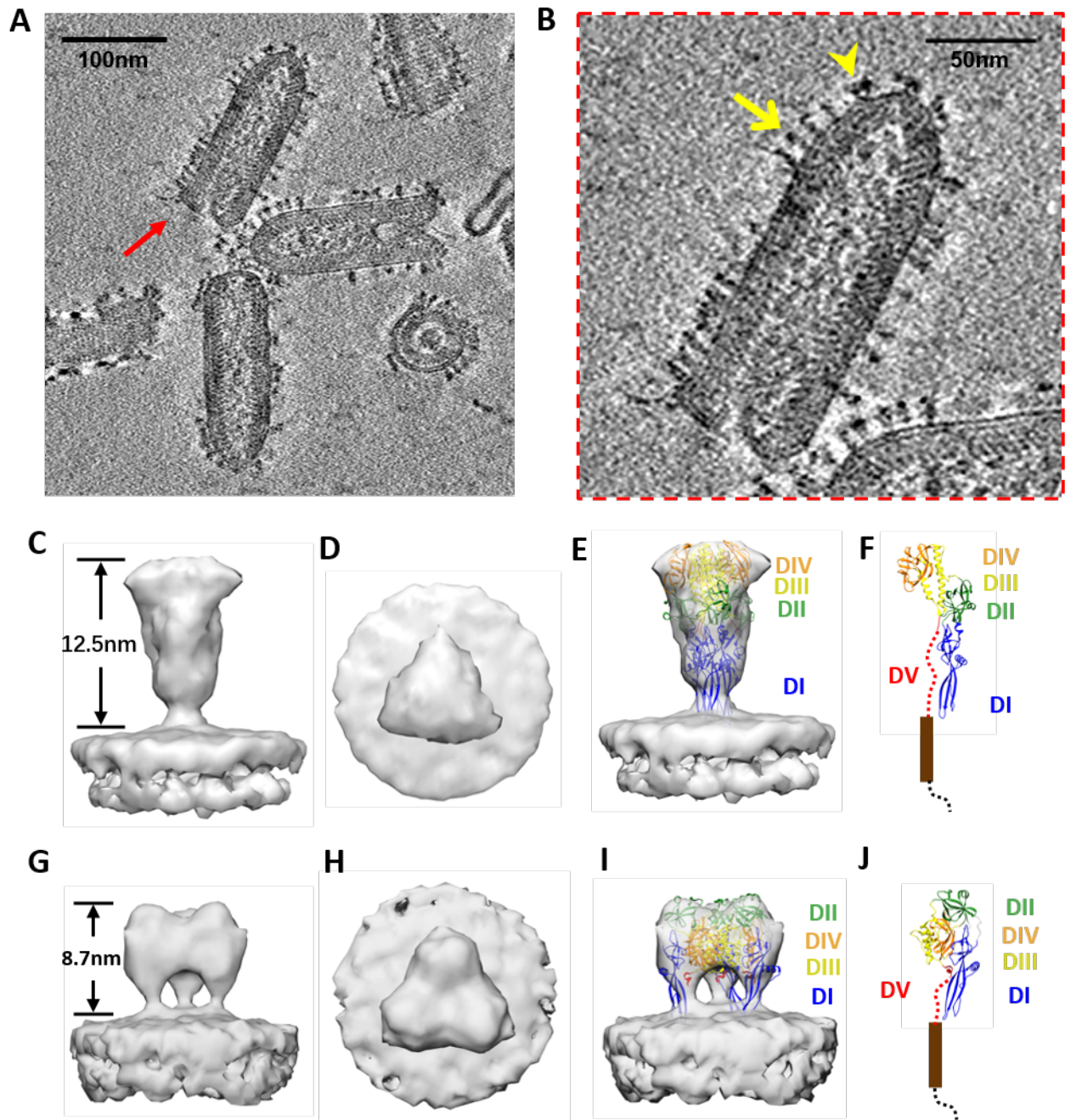
Figure 2- 1. In situ structures of two distinct conformations of VSV G.

(A) A 10Å-thick slab from the tomogram showing bullet-shaped VSV virions.

(B) The VSV particle indicated by red arrow in (A), with a yellow arrow pointing to long-form G and a yellow arrow head pointing to short-form G.

(C~F) Subtomographic average of the long-form G densities, whose ectodomain matches the crystal structure of G in postfusion conformation. The subtomographic average of the long-form G densities is shown either as shaded surfaces viewed from side (C) and top (D), or as semi-transparent gray (E) fitted with crystal structure of the G trimer (ribbon) in the postfusion conformation (PDB: 5I2M) [20]. For clarity, one of the subunits is shown alone in (F) with five domains (DI~DV) indicated.

(G~J) Subtomographic average of the short-form G densities, whose ectodomain matches the crystal structure of G in prefusion conformation. The subtomographic average of the short-form G densities is shown either as shaded surfaces viewed from side (G) and top (H), or as semi-transparent gray (I) fitted with crystal structure of the G trimer (ribbon) in the prefusion conformation (PDB: 5I2S) [21]. For clarity, one of the subunits is shown alone in (J) with five domains (DI~DV) indicated.
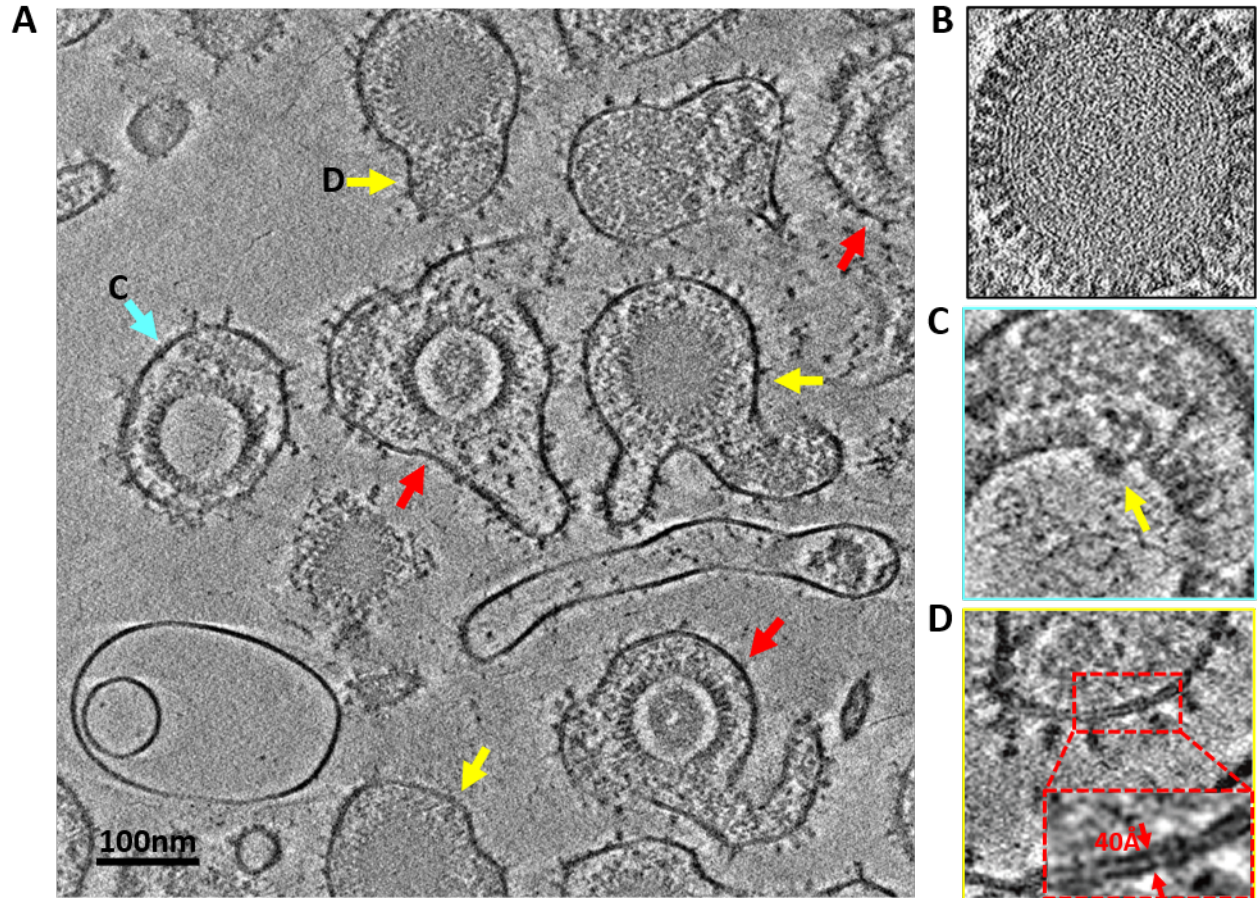
Figure 2- 2. CryoET of HCMV

(A) A 10Å-thick slab from the tomogram of Movie S2 showing a digital slice of various HCMV particles, each comprising an outer envelope layer, an intermediate tegument compartment, and an icosahedral capsid inside. Though all wrapped a pleomorphic glycoprotein-decorated envelope, these particles differ inside: each virion (yellow arrows) containing a C capsid (with DNA genome) and non-infectious enveloped particle (NIEP) either an A capsid (empty, blue arrow), or a B capsid (containing scaffolding protein but no DNA, red arrows).

(B) A zoom-in slice of a C capsid, with dsDNA duplexes resolved among the fingerprint-like pattern of the genome.

(C) A slice from the particle indicated by the cyan arrow in (A) showing the unique portal complex (arrow) at one of the 12 vertices.

(D) A zoom-in envelope region of the particle indicated by a yellow arrow in (A) showing the two resolved leaflets of the lipid bilayer envelope (inset: the enlarged boxed region). The side of the boxes in B-D is 120 nm.
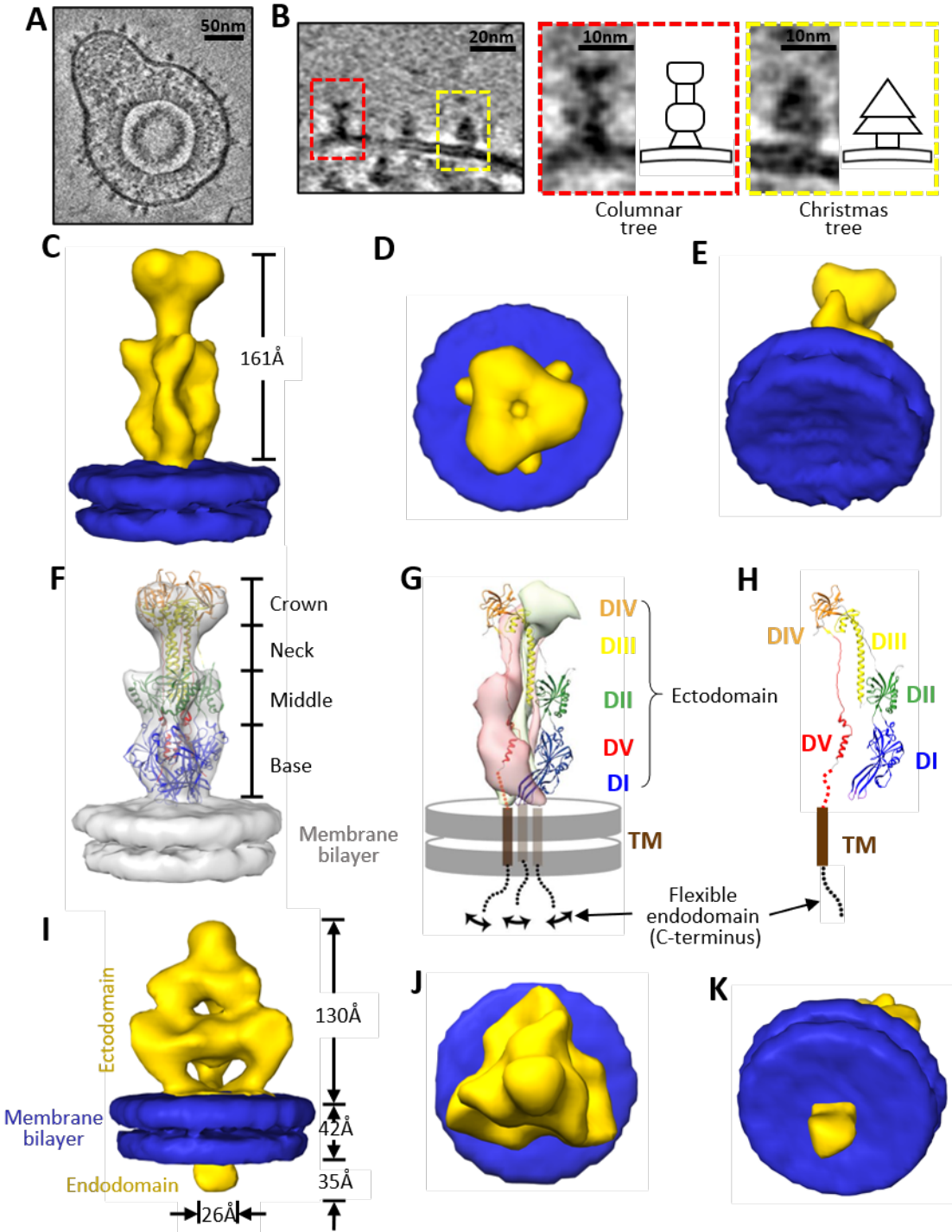
Figure 2- 3. In situ structures of gB in the "postfusion" and prefusion conformations.

(A) A representative virion showing various glycoprotein densities on its envelope.

(B) Identifications of the columnar tree-shaped (red box) and Christmas tree-shaped (yellow box) glycoprotein densities that both match the expected volume of gB (see main text). Insets are the enlargements of the two forms with their corresponding shape schematic.

(C~H) Sub-tomographic average of the columnar tree-shaped glycoprotein densities, whose ectodomain matches the crystal structure of gB ectodomain in the postfusion conformation (PDB: 5CXF) [16]. The subtomographic average of the columnar-shaped densities (yellow) and associated membrane bilayer (blue) are shown either as shaded surfaces viewed from side (C), top (D) and slanted bottom (E), or as semi-transparent gray fitted with the gB ectodomain trimer crystal structure (ribbon) at the postfusion conformation (PDB: 5CXF) [16] (F). Two subunits of the gB trimer crystal structure are shown as pink and gray surfaces, while the third subunit as ribbons with its domains colored as in [16] and its transmembrane helix as brown cylinder and the C-terminal flexible endodomain as a swinging dotted lines (G). For clarity, the third subunit is shown alone in (H) with five domains (DI~DV) indicated.

(I~K) Sub-tomographic average of the Christmas tree-shaped densities (yellow) and associated membrane bilayer (blue) viewed from side (I), top (J) and slanted bottom (K).
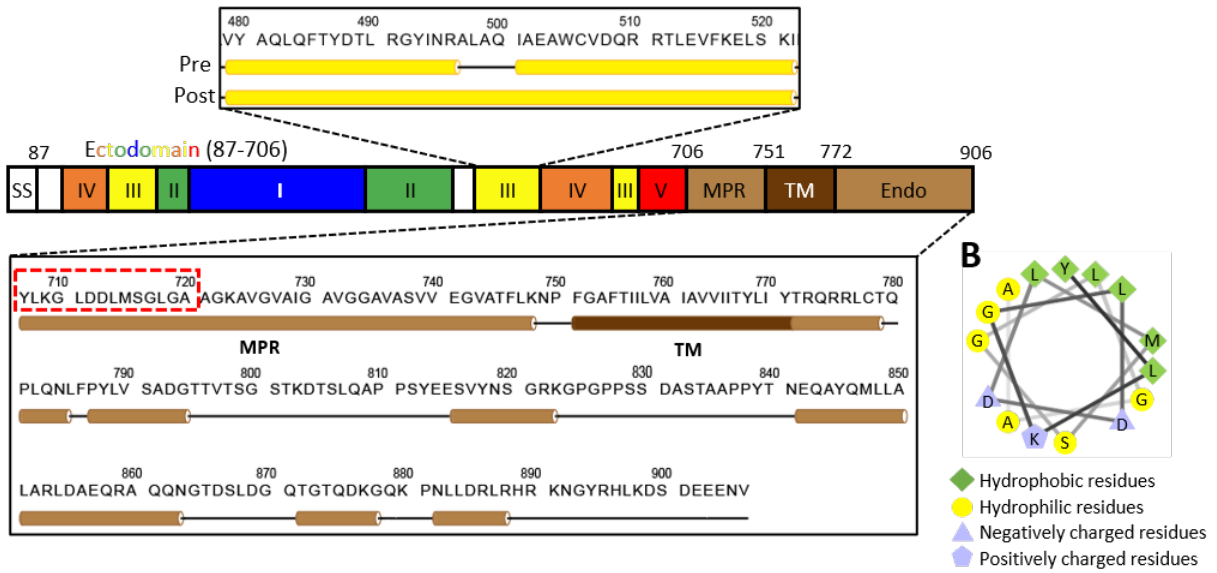
Figure 2- 4. Schematic illustration of the full-length HCMV gB

(A) Mapping of domains to the full-length HCMV gB primary sequence (SS = signal sequence, MPR = membrane proximal region, TM = transmembrane domain, Endo = endodomain). Upper inset: in the prefusion (Pre) conformation, the sequence of the central helix in DIII resolved in the postfusion gB crystal structure (Post) is predicted to fold into two helices joined by a short loop around residues 498-500. Lower inset: predicted secondary structures of the sequence encompassing the MPR, TM and endodomain of gB in the prefusion conformation.

 (B) Helical wheel diagram of the first 15 amino acids of MPR (sequence in red dashed box in (A)), showing one side with a cluster of hydrophobic amino acids.
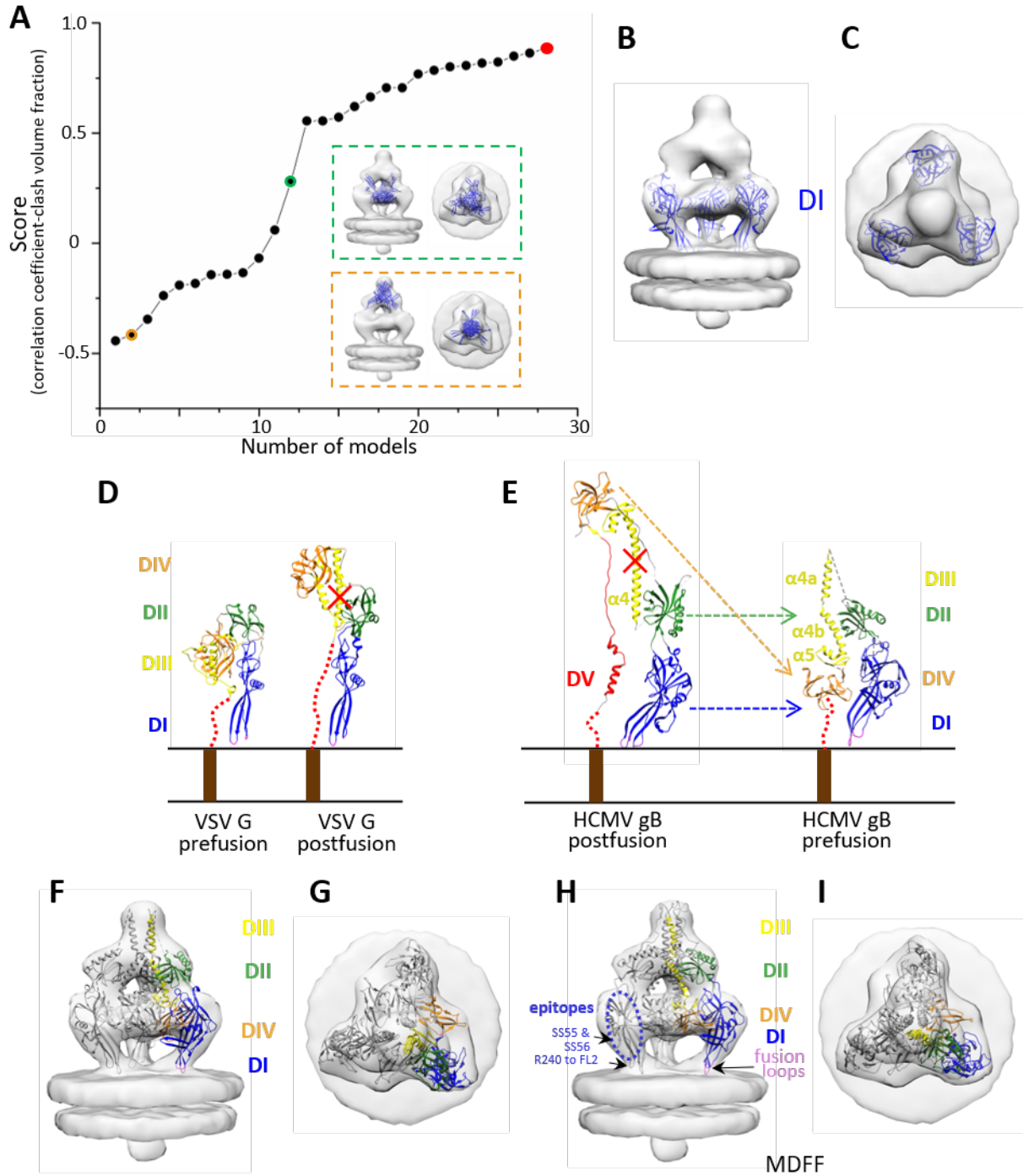
Figure 2- 5. Domain fitting for prefusion gB

(A~C) Domain fitting for DI. (A) Scores of 35 models for DI. Models are ranked in descending

order of scores. Red dot indicates the model of highest score. (B, C) DI structure, indicated as red

dot in (A), is superimposed in the subtomographic average of the Christmas tree-shaped density

(semitransparent gray), viewed from side (B) and top (C).

(D) VSV G domain rearrangement of crystal monomer structure from prefusion (left, PDB: 2J6J) [21] to postfusion (right, PDB: 2CMZ) [21]. The red dotted lines represent the unresolved domain DV between transmembrane helix and the ectodomain.

(E) the crystal structure of one protomer of the postfusion HCMV gB ectodomain (left, PDB: 5CXF) [16] is shown as ribbon next to the predicted prefusion gB structure (right) with domains arranged according to those in the prefusion VSV G. α4 and α5 represent the long central helix and the following short helix in DIII in postfusion gB structure. α4a and α4b represent the two helix breaking from α4.

(F, G) The predicted prefusion gB structure shown in (E, right) is superposed with two other symmetric copies (gray ribbon) in the subtomographic average of the Christmas tree-shaped density (semi-transparent gray), viewed from side (F) and top (G).

(H, I) The MDFF-simulated prefusion gB structure is superimposed with two other symmetric copies (gray ribbon) in the subtomographic average of the Christmas tree-shaped density (semi-transparent gray), viewed from side (H) and top (I). The epitopes of HSV-1 antibodies SS55/SS56 and R240 are DI and fusion loop 2 of HSV-1 gB, respectively ([26]); the corresponding locations of these two epitopes in our domain model of HCMV gB are indicated.
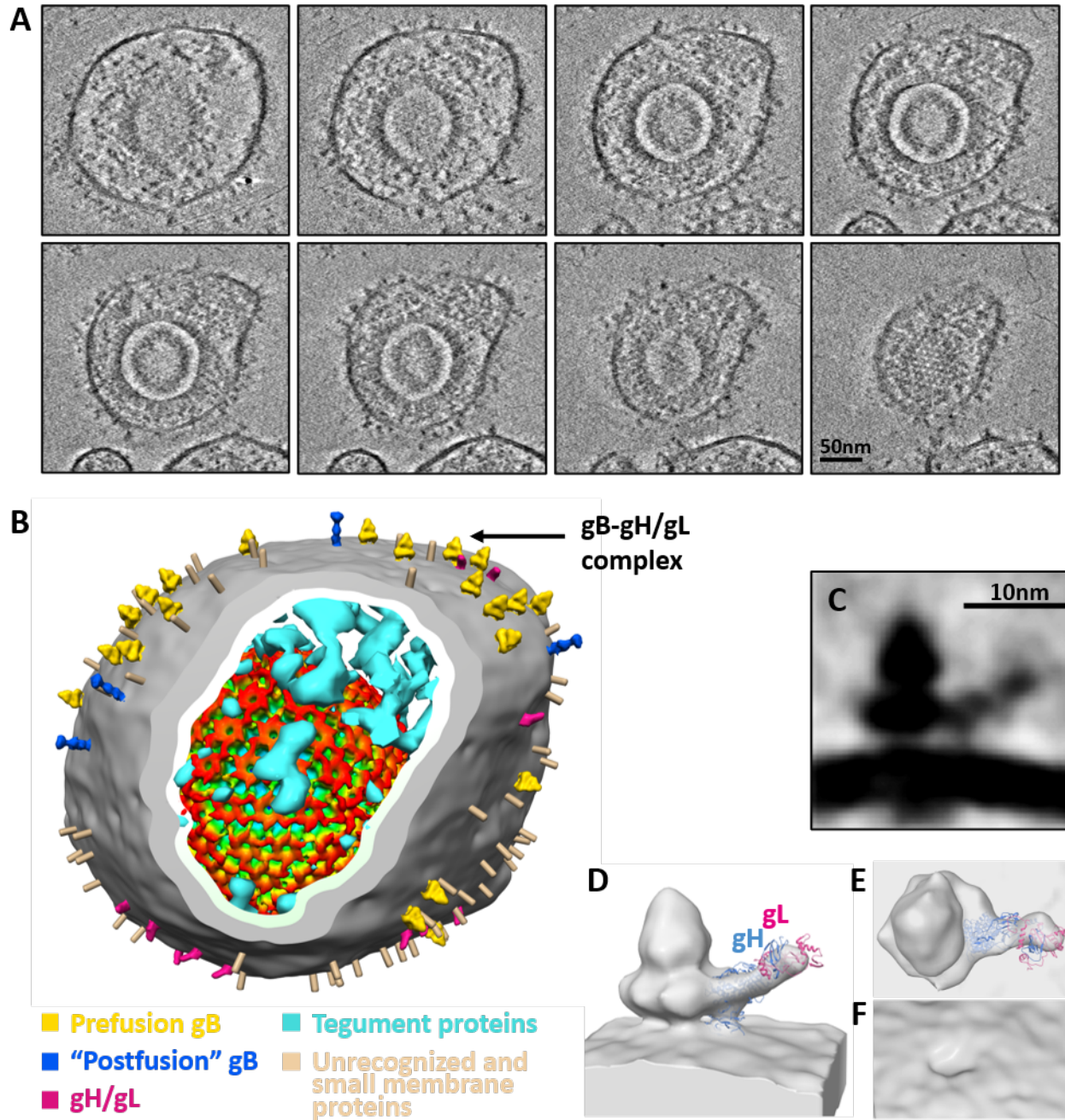
Figure 2- 6. In situ structure of gH/gL complex adjacent to prefusion gB.

(A) A series of slices in a tomogram showing a HCMV particle at different sections.

(B) A 3D surface view from the HCMV particle in (A) with subtomographic average of prefusion

and "postfusion" gB and gH/gL placed back on the viral envelope segmented from the tomogram.

Black arrow indicates the gB/gH/gL complex. Unidentified glycoprotein densities are indicated as

rods. Tegument proteins are shown as cyan densities. The recently published icosahedral

reconstruction of capsid [11] was low-passed to 10Å, radially colored and placed back in its location.

(C~F) The subtomographic average (C) showing a putative gH/gL complex adjacent to prefusion gB. The subtomographic average is also shown fitted with crystal structure of gH/gL (ribbon) (PDB: 5VOB) [17] in semitransparent surface viewed from side (D), top (E) and bottom (F).
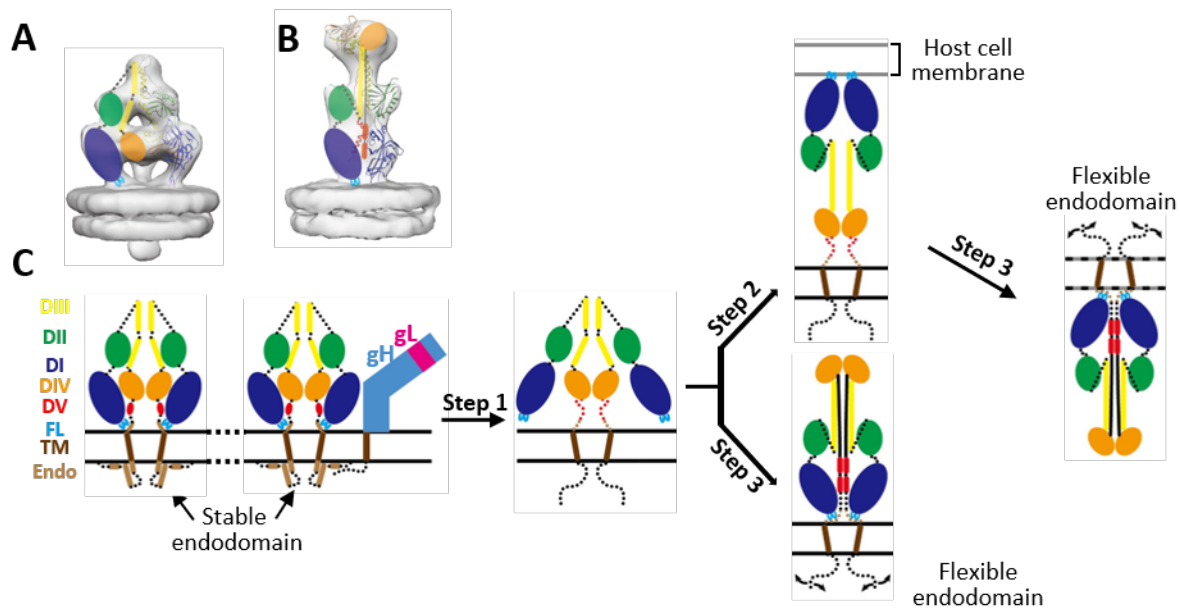
Figure 2- 7. Schematic illustration of conformation changes of gB during membrane fusion

(A, B) Subtomographic averages of prefusion gB (A) with domains illustrated as in Figure 3 and "postfusion" gB (B) with domains colored as in [16].

(C) A working model of gB conformational change during membrane fusion. In step 1, destabilization of the endodomain of prefusion gB either by cytotail conformational changes following gH/gL receptor-binding or by other means (*e.g.*, mechanical stress such as high-speed centrifugation during viral purification) triggers DI and DII to reorient, exposing the fusion loops on DI. Subsequently, the exposed fusion loops could make contact either with cell membrane in close proximity (in the case of receptor binding) (step 2) or with viral membrane. Finally (step 3), DV

refolds into an extended form, transforming gB into its "postfusion" conformation: in the presence of cell membrane, the C-terminal part and the fusion loops come together and the membranes fuse; in the absence of cell membrane, the exposed fusion loops insert into the viral membrane.

2.10 Supplementary Figures
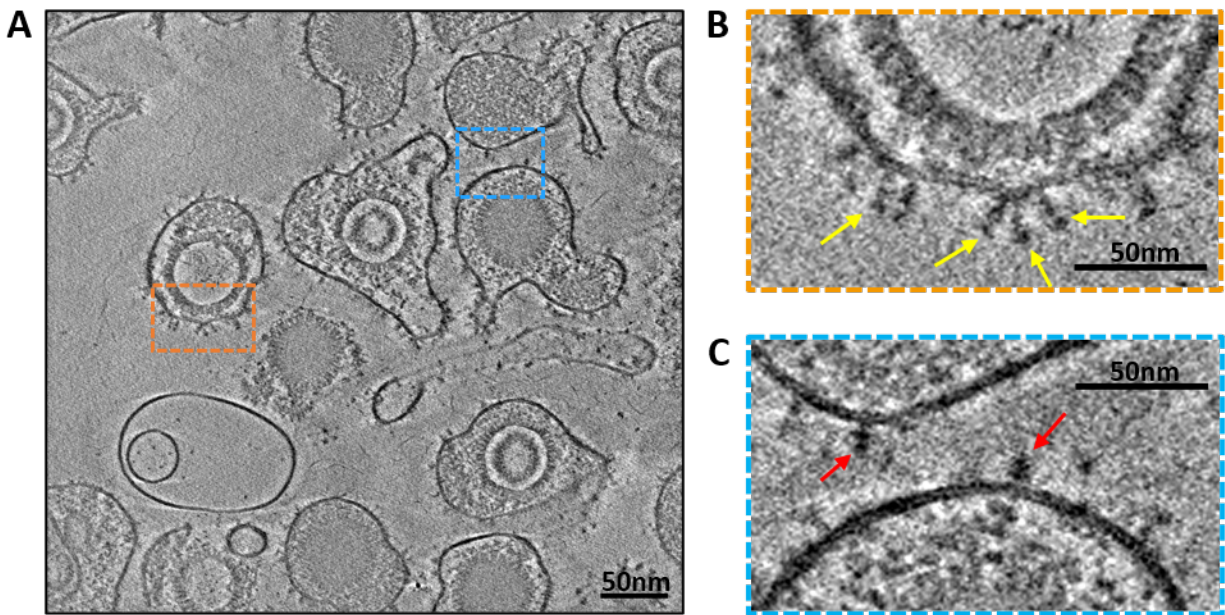
With Volta Phase Plate
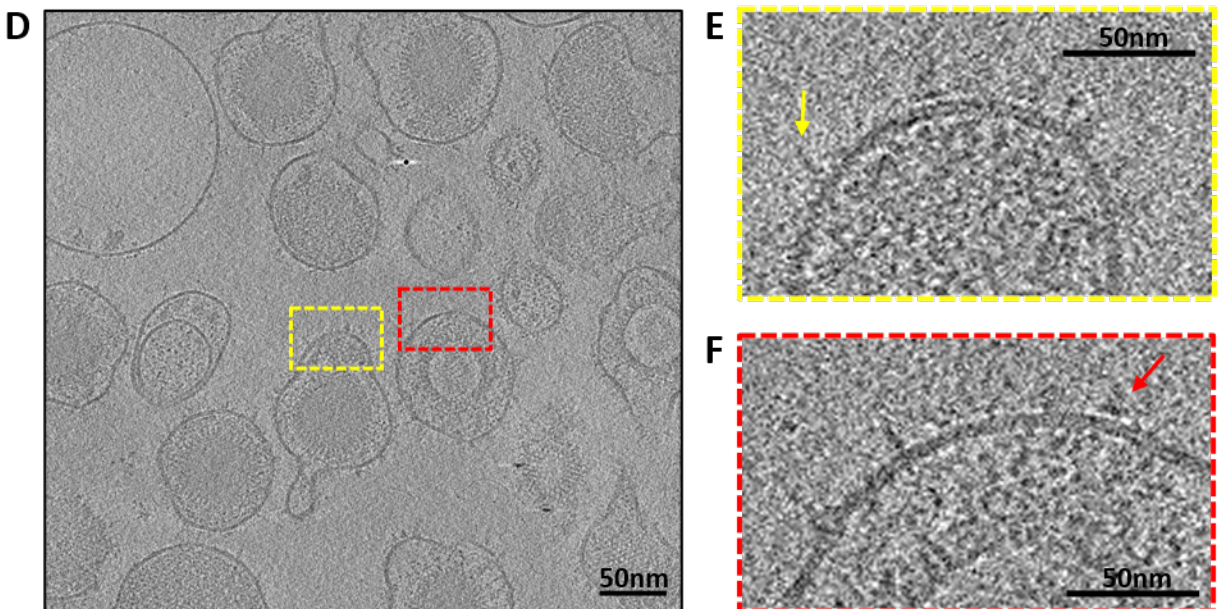
Without Volta Phase Plate

Supplemental Figure 2- 1. Comparison of tomograms obtained with and without VPP.

(A~C) A slice (A) and zoom-in envelope regions (B, C) of a tomogram reconstructed from tilt series obtained with VPP, showing greatly improved contrast that is sufficient to distinguish columnar tree-shaped ("postfusion") gB (yellow arrows in B) from the Christmas tree-shaped (prefusion) gB (red arrows in C).

(D~F) A slice (D) and zoom-in envelope regions (E, F) of a tomogram reconstructed from tilt series obtained without VPP, showing the relatively poor contrast and great ambiguity to distinguish columnar tree-shaped ("postfusion") gB (yellow arrow in E) from the Christmas tree-shaped (prefusion) gB (red arrow in F). Consequently, significantly more tilt series without VPP than with VPP had to been recorded in order to obtain similar number of particles for subtomographic averaging.
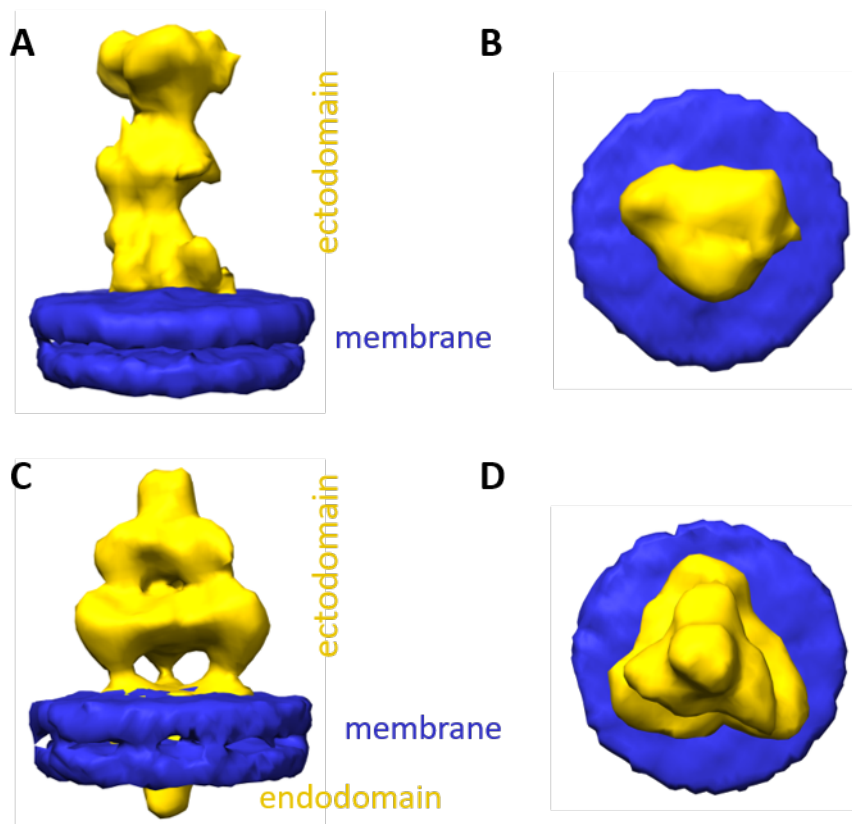
Supplemental Figure 2- 2. Subtomographic averages of gB without imposing symmetry

(A, B) Subtomographic averages of gB in its postfusion conformation without imposing symmetry viewed from side (A) and top (B).

(C, D) Subtomographic averages of gB in its prefusion conformation without imposing symmetry viewed from side (C) and top (D).
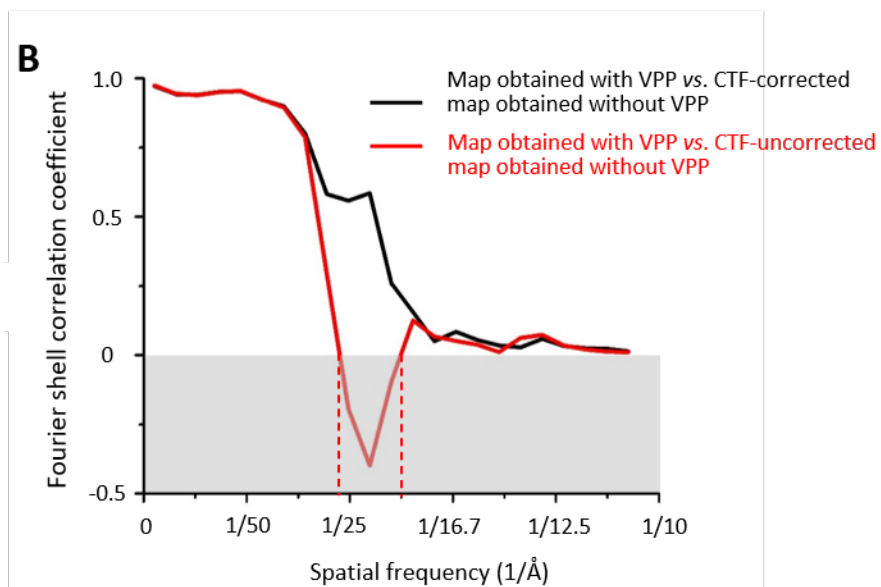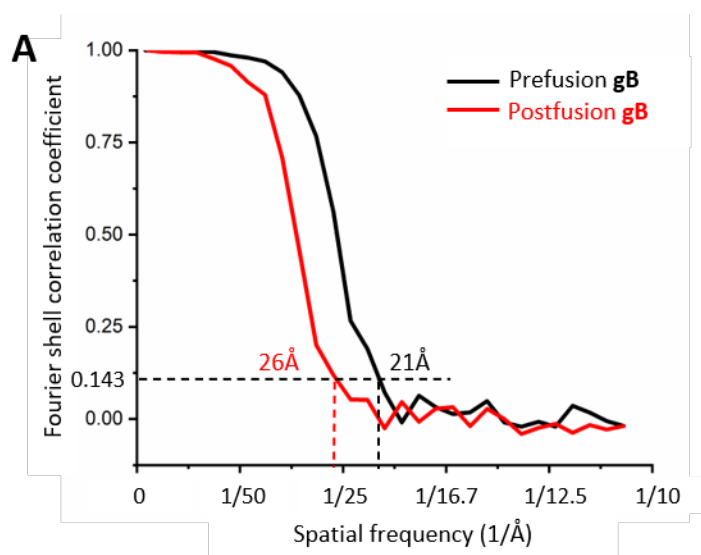
Supplemental Figure 2- 3. Fourier shell correlation (FSC) analyses and resolution

comparisons.

(A) FSC coefficients as a function of spatial frequency for the gold-standard resolution determined for final subtomographic averages of prefusion (black) and "postfusion" (red) gB trimers. .

(B) FSC coefficients as a function of spatial frequency between subtomographic averages of prefusion gB trimers obtained with VPP and without VPP. For the average obtained without VPP, CTF correction is necessary as indicated by the negative correlation coefficients in the range from 1/26 Å^{-1} to 1/20 Å^{-1} spatial frequencies.
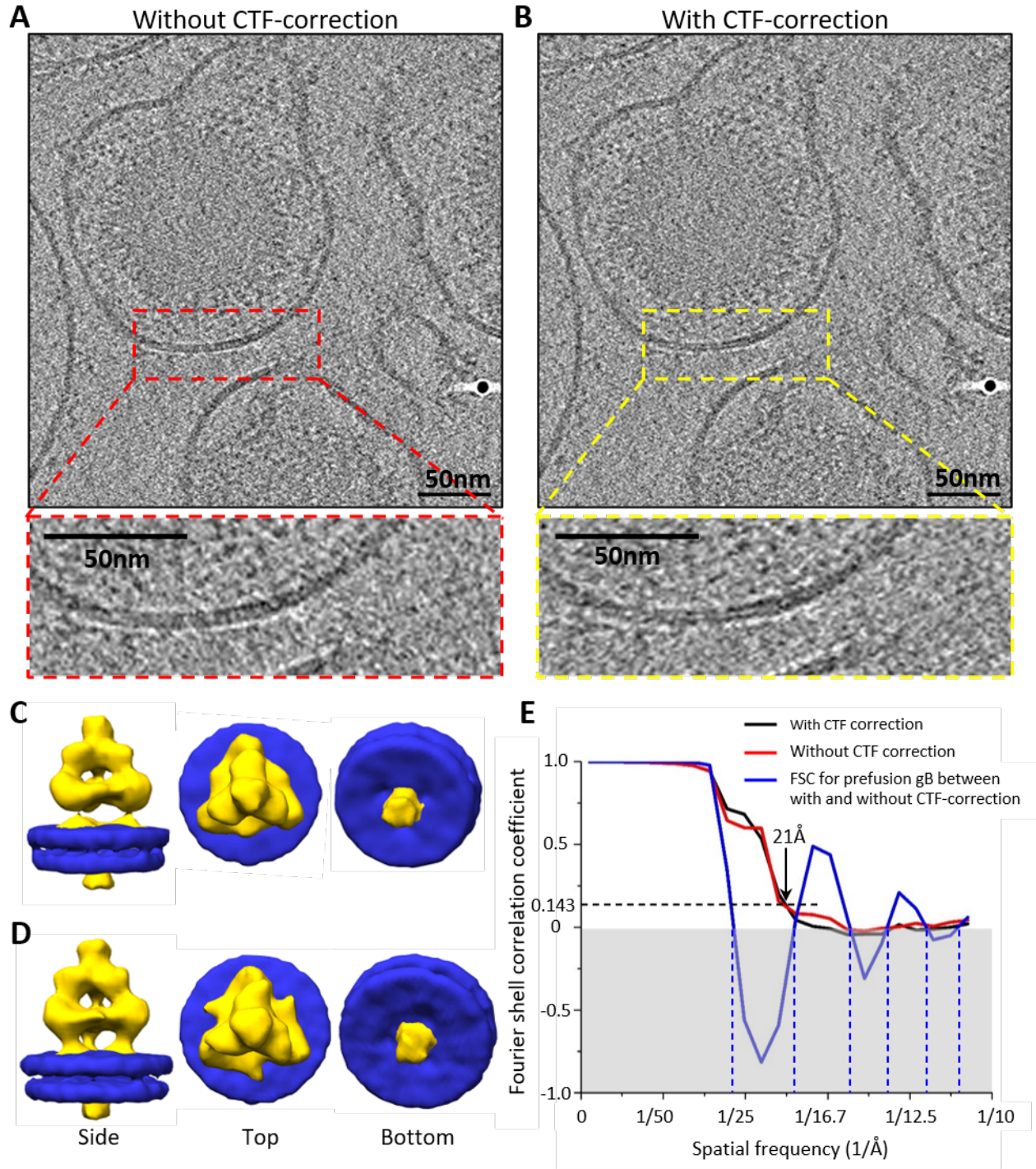
Supplemental Figure 2- 4. Tomograms and subtomographic averages from tilt series obtained without VPP

(A, B) Comparison of corresponding slices from CTF-uncorrected (A) and CTF-corrected (B) tomograms. The viral envelope region of the particle indicated by the dashed boxes in (A, red) and

(B, yellow) are enlarged, showing that the membrane bilayer is better resolved after CTF correction in the yellow zoom-in inset.

(C, D) Subtomographic average of the Christmas tree-shaped densities (yellow) and associated membrane bilayer (blue) viewed from side, top and slanted bottom. (C) is obtained without CTF-correction and (D) is with CTF-correction.

(E) FSC coefficients as a function of spatial frequency between subtomographic averages of prefusion gB trimers obtained with CTF correction and without CTF correction. For the subtomographic average obtained without VPP, CTF correction is necessary as indicated by the negative correlation coefficients (grey zone) for some spatial frequencies.

A. Christmas tree-shaped prefusion gB on HCMV virion (this study), reviewed from its side and top, as in Figures 3 and 5.

13nm

10.5nm

DIII
DII
DIV
DI
Fusion loops

B. "Short-form" gB of HSV-1 [27] showing as in A, colored surface (upper panel) and superposition (lower panel) of semitransparent density (gray) and fitted domains (ribbons). Red dotted ellipses indicate the densities that were unaccounted for by the fitted psedoatomic model.

11nm

13nm

Fusion loops
DI
DII

C. Superposition of the proposed HSV-1 gB domain (DI and DII) arrangement from [27] into the density of the prefusion gB trimer on HCMV virion (this study) showing *mismatch*, which is indicated by red dotted ellipses.

Fusion loops
DI
DII

Supplemental Figure 2- 5. Direct comparisons of the averaged map and fitted pseudoatomic model between prefusion HCMV gB and previous "short-form" HSV-1 gB

(A) Christmas tree-shaped prefusion gB on HCMV virion (this study), reviewed from its side and top, as in Figures 3 and 5.

(B) "Short-form" gB of HSV-1 [27] showing as in (A), colored surface (upper panel) and superposition (lower panel) of semitransparent density (gray) and fitted domains (ribbons). Red dotted ellipses indicate the densities that were unaccounted for by the fitted psedoatomic model.

(C) Superposition of the proposed HSV-1 gB domain (DI and DII) arrangement from [27] into the density of the prefusion gB trimer on HCMV virion (this study) showing mismatch, which is indicated by red dotted ellipses

Chapter 3: Structural basis for capsid recruitment and coat formation during HSV-1 nuclear egress

Elizabeth B. Draganova[1], Jiayan Zhang[2,3,4], Z. Hong Zhou[2,3,4], and Ekaterina E. Heldwein[1,#]

[1]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, 02111, USA

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

[3]Molecular Biology Institute, UCLA, Los Angeles, California, 90095, USA

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

[#]Correspondence: Ekaterina Heldwein, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111; 617-636-0858; katya.heldwein@tufts.edu

3.1 Abstract

During herpesvirus infection, nascent viral capsids egress the nucleus into the cytoplasm by an unusual mechanism whereby capsids bud at the inner nuclear membrane. This process is mediated by the conserved heterodimeric nuclear egress complex (NEC), anchored to the inner nuclear membrane, that deforms the membrane around the capsid by forming a hexagonal array. However, how the NEC coat interacts with the capsid and how proper curvature of the coat is achieved to enable budding are yet unclear. Here, we show that binding of a capsid protein, UL25, promotes the formation of a pentagonal rather than hexagonal NEC arrangement. Our results suggest that during nuclear budding, interactions between the UL25 bound to the pentagonal capsid vertices and the NEC introduce pentagonal insertions into the hexagonal NEC array to yield an NEC coat of the appropriate size and curvature, leading to productive budding and egress of UL25-decorated capsids.

3.2 Introduction

To replicate, all viruses must assemble their progeny virions and release them from the cell while overcoming many obstacles, including cellular compartmentalization. Viruses are thus experts at hijacking, manipulating, and, sometimes, even remodeling cellular architecture during viral morphogenesis and egress. Identifying and understanding the unique aspects of virus-induced cellular remodeling could unveil targets for therapeutic intervention; yet, we are only beginning to understand the mechanisms behind many of these processes.

One prominent example of virus-induced remodeling of cellular architecture can be observed during egress of herpesviruses – enveloped, double-stranded DNA viruses that infect a wide range of hosts, from mollusks to humans. All herpesviruses can establish lifelong, latent infections within the host, from which they can periodically reactivate, spreading to uninfected tissues and hosts

and causing a number of ailments. During reactivation of a latent infection, the progeny virions are assembled and released from the cell in a process termed egress whereby herpesvirus capsids traverse cellular membranes twice [reviewed in[1-3]]. First, nuclear capsids bud at the inner nuclear membrane (INM) forming enveloped vesicles that pinch off into the perinuclear space. These perinuclear viral particles fuse with the outer nuclear membrane (ONM), which releases the capsids into the cytosol. Cytoplasmic capsids then bud again at vesicles derived from the *trans*-Golgi network and early endosomes [reviewed in [1]] to form mature, infectious virions that are released from the cell by exocytosis. Whereas many enveloped viruses acquire their lipid envelopes by budding at the cytoplasmic membranes or the plasma membrane, herpesviruses are unusual among vertebrate viruses in their ability to bud at the nuclear envelope[2].

Capsid budding at the nuclear envelope requires the nuclear egress complex (NEC), composed of two conserved herpesviral proteins that, in herpes simplex virus type 1 (HSV-1), are named UL31 and UL34 [reviewed in [2,4,5]]. The NEC heterodimer is anchored at the INM through the single C-terminal transmembrane helix of UL34 and faces the nucleoplasm[6]. UL31 is a nuclear phosphoprotein that colocalizes with UL34[7,8] and interacts with the capsid during nuclear egress[9,10]. Both UL31 and UL34 are necessary for efficient nuclear egress, and in the absence of either protein, capsids accumulate in the nucleus and viral replication is reduced by several orders of magnitude[11,12].

Previously, we discovered that HSV-1 NEC has an intrinsic ability to deform and bud membranes by demonstrating that purified recombinant NEC vesiculates synthetic lipid bilayers *in vitro* without any additional factors or chemical energy[13]. Similar findings were reported with the NEC homolog from a closely related pseudorabies virus (PRV)[14]. Using cryogenic electron microscopy and tomography (cryoEM/ET), we then showed that the NEC forms hexagonal "honeycomb" coats on the inner surface of budded vesicles *in vitro*[13]. Hexagonal coats of similar dimensions were also observed in capsidless perinuclear vesicles formed in uninfected cells expressing PRV NEC[15]. Furthermore, HSV-1 NEC formed a hexagonal lattice with dimensions

similar to the hexagonal coats observed in budded vesicles in the crystals[16]. The resulting crystal structure illuminated interactions forming the NEC lattice at the atomic level[16]. Mutations that disrupt oligomeric interfaces reduce budding *in vitro*[13,16] and *in vivo*[17,18]. Collectively, these findings demonstrated that the NEC is a budding machine that generates negative membrane curvature by oligomerizing into a hexagonal coat on the surface of the membrane.

What remains unclear, however, is how the NEC achieves appropriate coat geometry compatible with negative membrane curvature formation during budding. A purely hexagonal arrangement is flat, so curvature is typically achieved either by insertions, such as pentagons found at 12 vertices of an icosahedron[19], or by inclusion of irregular defects, as observed in several viral coats[20-23]. Pentagons have not yet been observed in NEC coats formed either *in vitro*[24], in uninfected cells expressing UL31 and UL34[15], or in perinuclear viral particles purified from cells infected with a US3 kinase-null HSV[25]. Therefore, it is unknown whether interaction with a mature capsid could promote pentagonal formation within NEC coats and if so, how the NEC coat would be arranged around the capsid. In perinuclear viral particles visualized in infected cells, the NEC coats appear to be tightly associated with the capsid[15]. The capsid is thought to engage the NEC during nuclear budding directly by using the capsid-associated tegument protein UL25 to bind UL31[9,26]. Moreover, herpesviral capsids have icosahedral geometry[27,28], so it is tempting to speculate that the capsid geometry may influence the geometry of the NEC coat.

A fortuitous observation that UL25 co-localizes with synthetic liposomes *in vitro* in the presence of the NEC prompted us to investigate the interactions of UL25 and NEC from HSV-1 and the effect of UL25 on the NEC-mediated budding *in vitro*. Here, by confocal microscopy, we show that free UL25 (*i.e.*, not on capsid vertices) inhibits NEC-mediated budding *in vitro*. 3D visualization of the molecular architecture by cryoET further reveals that free UL25 forms a net of interconnected five-pointed stars that coat the membrane-bound NEC layer. We hypothesize that this UL25 net blocks budding by preventing membrane-bound NEC coats from undergoing conformational changes necessary to form negative membrane curvature required for budding.

We have also uncovered the ability of the NEC to form an alternative pentagonal, rather than hexagonal, arrangement when bound to the UL25. We propose that during nuclear budding, incorporation of pentagonal insertions within the hexagonal NEC coat above the vertex positions of an incoming capsid with pentagonal UL25 vertices enables the formation of the NEC coat of appropriate size and curvature leading to successful budding and egress of the viral capsid.

3.3 RESULTS

3.3.1 Generation of UL25 variants.

HSV-1 UL25 can be expressed in soluble form in *E. coli* only when residues 1-44 are deleted[29]. In the cryoEM structure of the HSV-1 capsid, these residues lack defined secondary structure (Fig. 1a) and appear to fold due to extensive interactions with another capsid protein, UL17[27], which suggests that these residues are disordered in isolated UL25, potentially leading to aggregation and poor solubility. Therefore, we generated and expressed an HSV-1 UL25Δ44 construct, which lacks residues 1-44. UL25Δ44 was soluble and could be purified, in agreement with the previous report[29], but was proteolytically cleaved during purification despite the presence of protease inhibitors (Fig. 1b). N-terminal sequencing (data not shown) of the cleavage product revealed that UL25Δ44 was cut between residues Q72 and A73. To prevent heterogeneity due to cleavage, we generated two constructs. The first construct, UL25Δ44 Q72A, has a single point mutation that should eliminate the cleavage site whereas the second construct, UL25Δ73, corresponds to the cleavage product. Both constructs yielded a single UL25 species after purification (Fig. 1b).

3.3.2 UL25Δ44 Q72A inhibits NEC-mediated budding.

To assess the effect of UL25 on NEC-mediated budding, we used an established *in vitro* budding assay utilizing fluorescently labelled giant unilamellar vesicles (GUVs) and membrane-impermeable fluorescent dye, Cascade Blue[13]. NEC220 and UL25Δ44 Q72A were added to the GUVs in 1:1, 1:6, or 1:10 molar ratios, and budding events were quantified. UL25Δ44 Q72A

inhibited NEC-mediated budding in a dose-dependent manner, and at 1:10 NEC:UL25 molar ratio, few budding events, if any, were observed (Fig. 1c).

UL25Δ44 consists of a long N-terminal α-helix (residues 48-94), followed by a flexible linker unresolved in the cryoEM structure, and the globular core (residues 134-580) (Fig. 1a). To identify the region responsible for budding inhibition, we tested the UL25Δ133 variant that contains the globular core alone. UL25Δ133 did not inhibit budding even at a 1:10 ratio of NEC:UL25 (Fig. 1c), and thus the globular core of UL25 was insufficient for budding inhibition. UL25Δ73 likewise did not inhibit budding (Fig. 1c), indicating that the N-terminal half of the long α-helix was necessary for budding inhibition. To further narrow down the inhibitory region within UL25, we analyzed its sequence conservation. Sequence alignment of UL25 homologs from five alphaherpesviruses revealed a divergent N terminus followed by the highly conserved alanine-rich region, residues 61-69 (Fig. 1a). To assess the contribution of the divergent N terminus to inhibition, we generated the UL25Δ58 Q72A construct. UL25Δ58 Q72A did not inhibit NEC220 budding (Fig. 1c). We also generated a UL25Δ50 Q72A construct (as a control for studies using eGFP-UL25 chimera described below), which inhibited budding at a 1:10 NEC:UL25 ratio (Fig. 2a). We conclude that while residues 45-50 are dispensable for inhibition, the remainder of the N terminus is essential.

3.3.3 UL25 does not bind synthetic membranes.

UL25 could, in principle, inhibit NEC-mediated budding by competing with the NEC for binding to membranes. To test this, we used an established co-sedimentation assay utilizing multilamellar vesicles (MLVs) of the same composition as the GUVs used in the budding assay[13]. Unlike NEC220[13], UL25Δ44 did not bind acidic lipid vesicles (Fig. 1d). Therefore, competition for binding to membranes does not explain inhibition of the NEC-mediated budding by UL25.

3.3.4 UL25Δ44 Q72A, UL25Δ73, and UL25Δ133 bind immobilized NEC.

UL25 does not bind membranes on its own; therefore, to inhibit NEC-mediated budding, UL25 must instead bind to the NEC. Indeed, using surface plasmon resonance (SPR), we detected

binding of UL25Δ44 Q72A to NEC220-His immobilized on an Ni-NTA chip (Supplementary Fig.

S1). Unexpectedly, UL25Δ73 and UL25Δ133 could also bind the immobilized NEC220-His

(Supplementary Fig. S1). Thus, binding of UL25 to NEC is not sufficient for inhibition of budding.

Unfortunately, SPR binding curves could not be reliably fitted and accurate binding constants

could not be derived due to the lack of saturation even at high concentrations of UL25,

presumably, due to the presence of non-specific binding in the system.

No interaction between UL25 and NEC was detected in solution, either between UL25Δ44

and NEC220, as measured by isothermal titration calorimetry (ITC) (Supplementary Fig. S1), or

between UL25Δ44 and NEC185Δ50 (a truncated construct that was crystallized previously[16]) by

size-exclusion chromatography (Supplementary Fig. S1). Therefore, UL25 binds NEC only when

the latter is immobilized.

3.3.5 Binding of UL25Δ44 Q72A to the NEC inhibits budding whereas binding of UL25Δ73 or UL25Δ133 does not.

To visualize UL25 localization during NEC-mediated budding by confocal microscopy, both the

inhibitory (eGFP-UL25Δ50 Q72A) and the non-inhibitory (eGFP-UL25Δ73 and eGFP-UL25Δ133)

variants of UL25 N-terminally tagged with eGFP were produced. eGFP-UL25Δ50 Q72A was used

instead of eGFP-UL25Δ44 Q72A because the latter was prone to degradation. eGFP-UL25

constructs were imaged in the presence of GUVs labelled with red fluorescent dye, either alone or

the presence of the NEC220, at a 1:10 molar ratio of NEC to UL25. With eGFP-UL25 constructs

alone, no eGFP signal was detected on the GUV membranes (data not shown). In the presence of

eGFP-UL25Δ50 Q72A and the NEC220, the eGFP signal colocalized with the GUV membranes

(Fig. 2b) and very little budding was detected (Fig. 2a). We conclude that binding of eGFP-

UL25Δ50 Q72A to the NEC220 on the surface of the GUVs inhibits budding.

By contrast, in the presence of eGFP-UL25Δ73 and NEC220, the eGFP signal sometimes

colocalized with the membranes of the intraluminal vesicles (ILVs) inside the GUVs (Fig. 2c). To

image ILVs with Cascade Blue, a separate budding reaction was performed with a 1:5 molar ratio

of NEC to UL25 used to reduce aggregation as described below. Multiple ILVs filled with Cascade Blue or with eGFP signal on their membranes were observed (Supplementary Fig. S2). These observations confirmed that binding of eGFP-UL25Δ73 to the NEC220 did not block budding (Fig. 2a) and demonstrated that eGFP-UL25Δ73 remained associated with the membrane-bound NEC220 throughout budding. However, in other cases, the eGFP signal was clustered around the unbudded GUVs, which could be due to the aggregation of eGFP-UL25Δ73 (Fig. 2d). Such aggregation was not observed for either eGFP-UL25Δ50 Q72A (Fig. 2b) or eGFP-UL25Δ133 (Supplementary Fig. S3), and we conclude that the N-terminal truncation destabilizes the long helix in UL25 (Fig. 1a), which causes much of eGFP-UL25Δ73 to aggregate on NEC-coated GUVs. Although such aggregation inhibits budding locally (Fig. 2d), bulk measurements show that NEC-mediated budding remains efficient in the presence of eGFP-UL25Δ73 (Fig. 2a). We hypothesize that sequestration of large amounts of eGFP-UL25Δ73 on a few NEC-coated GUVs reduces its concentration throughout the sample, allowing budding to proceed.

eGFP-UL25Δ133 colocalized minimally with the GUVs in the presence of the NEC220, and none of the ILVs had any eGFP signal (Supplementary Fig. S3). This contradicted the ability of UL25Δ133 to bind NEC220 as shown by SPR (Supplementary Fig. S1). We hypothesize that the eGFP tag may interfere with binding of eGFP-UL25Δ133 to the NEC. Nevertheless, given that untagged UL25Δ133 binds the NEC but does not inhibit NEC-mediated budding (Fig. 1c), we conclude that the UL25 core alone is insufficient for budding inhibition and requires a full-length UL25 N-terminal helix.

3.3.6 The structure rather than the sequence of the N-terminal helix in UL25 is important for budding inhibition.

N-terminal truncations pinpointed the intact N-terminal helix of UL25 as essential for inhibition. At the same time, the UL25 core alone is unable to inhibit budding, ruling out its involvement in budding inhibition. Therefore, we tested whether the peptide corresponding to the N-terminal helix of UL25, UL25$^{45-94}_{native}$ (Fig. 3a), was sufficient for inhibition. At a 1:10 molar ratio of NEC to

peptide, the UL25$^{45-94}_{native}$ peptide inhibited NEC budding by ~80%, nearly as efficiently as the UL25Δ44 Q72A construct (~90% inhibition) (Fig. 3b). Next, the peptide sequence was scrambled, generating a scrambled peptide UL25$^{45-94}_{scr}$ (Fig. 3a), that inhibited budding by ~80%, similarly to the UL25$^{45-94}_{native}$ peptide (Fig. 3b), which suggested that the native sequence was unimportant for the inhibitory activity. To test the importance of the 45-94 sequence in the context of the UL25Δ44 Q72A construct, the native 45-94 sequence was replaced with the scrambled sequence to generate the UL25$^{45-94}_{scr}$-UL25$^{95-580}$ construct (Fig. 3a). The UL25$^{45-94}_{scr}$-UL25$^{95-580}$ construct inhibited NEC budding by ~70% (Fig. 3b), which confirmed that the native sequence of the 45-94 region was of limited importance for inhibition and suggesting that the secondary structure and, potentially, amino acid composition was more important.

Both the UL25$^{45-94}_{native}$ and UL25$^{45-94}_{scr}$ peptides were only ~20% α-helical in solution (Fig. 3c,d) according to the circular dichroism (CD) estimates using DichroWeb[30] (Supplementary Table S1). However, within capsid-bound UL25, amino acids 48-94 form an α-helix (Fig. 1a). To assess the propensity of the UL25$^{45-94}_{native}$ and UL25$^{45-94}_{scr}$ peptides to form an α-helix in solution, CD was done in the presence of 30% trifluoroethanol (TFE) known to stabilize secondary structure[31]. Both peptides adopted α-helical structure (~50% and 65%, respectively) in the presence of TFE (Fig. 3c,d) suggesting that each peptide could, in principle, adopt α-helical structure when bound to a binding partner, *e.g.*, the NEC.

The shorter UL25$^{74-94}$ peptide (Fig. 3a) did not inhibit NEC-mediated budding (Fig. 3b). CD estimates showed that the UL25$^{74-94}$ peptide adopted a random-coil conformation in solution whether alone or in the presence of 30% TFE (Fig. 3e). Therefore, the UL25$^{74-94}$ peptide may be unable to inhibit NEC budding because it cannot form an α-helix. We hypothesize that within UL25Δ73, residues UL25$^{74-94}$ may also be disordered. We conclude that inhibitory activity of UL25 requires an intact N-terminal helix.

To determine if any α-helical peptide of similar length could inhibit NEC-mediated budding, we generated a heterologous 53-amino-acid peptide containing a series of well-characterized α-

helical EAEKAAK repeats ($E_{16}A_{22}K_{15}$)[32]. The $E_{16}A_{22}K_{15}$ peptide inhibited NEC-mediated budding

less efficiently than either the UL25$^{45-94}$~native~ or the UL25$^{45-94}$~scr~ peptides, only by ~50% (Fig. 3b).

The reduced inhibitory activity of this heterologous $E_{16}A_{22}K_{15}$ peptide could be due to differences in

the amino-acid composition. Additionally, the $E_{16}A_{22}K_{15}$ peptide was 60% α-helical in solution even

in the absence of TFE (Fig. 3f), in contrast to both the UL25$^{45-94}$~native~ and the UL25$^{45-94}$~scr~ peptides

that became mostly α-helical only in the presence of TFE (Fig. 3c,d). We hypothesize that a

stable, preformed α-helix may not be able to form appropriate inhibitory contacts with the NEC and

that for efficient inhibition, an α-helix of proper length and amino-acid content should form in the

presence of the NEC, in an induced-fit manner.

3.3.7 Mutations within the putative capsid-binding site obviate UL25 inhibition.

Charged residues within the membrane-distal region of UL31 (D275, K279, and D282) have been

implicated in capsid binding in PRV[33]. We generated the UL31 mutant in which D275, C278, K279,

and D282 were replaced with alanines. The corresponding mutant NEC220, termed capsid-binding

mutant (NEC220-CBM), mediated budding at levels similar to the WT NEC220 (Fig. 4a) but could

not be inhibited by UL25Δ44 Q72A (Fig. 4a). Moreover, eGFP-UL25Δ50 Q72A did not co-localize

with the GUV membranes in the presence of the NEC220-CBM (Fig. 4b). These results suggest

that binding of UL25Δ44 Q72A to the putative capsid-binding site within UL31 is essential for its

inhibitory activity.

3.3.8 UL25 interacts with membrane-bound NEC.

To determine the molecular mechanism by which UL25 inhibits NEC-mediated budding, we turned

to cryoEM. UL25Δ44 Q72A and NEC220 were incubated with large unilamellar vesicles (LUVs) of

the same composition as the GUVs used in the budding assay at a 1:10 molar ratio of NEC to

UL25 and visualized by cryoEM. Previously, we showed that NEC-mediated budding of synthetic

LUVs results in the formation of vesicles containing internal NEC coats[13]. By contrast, LUVs

incubated with UL25Δ44 Q72A and NEC220 appeared as mostly spherical vesicles with ~17-nm

thick, arrays of spikes on the outer surface of the LUVs (Fig. 5a). Given that the thickness of the

NEC coats is ~11 nm[13], we conclude that these external arrays are composed of a UL25Δ44 Q72A

positioned on top of the NEC220 layer at the membrane surface (Fig. 5a). Very few budded

vesicles were observed under these conditions, which is consistent with the inefficient budding

observed by confocal microscopy (Fig. 1c). Thus, binding of UL25Δ44 Q72A to the external

NEC220 coats on the surface of the lipid vesicles inhibits NEC-mediating budding, likely by

preventing the conformational rearrangements required for membrane deformation and budding

(see cryoET results below).

LUVs incubated with UL25Δ73 and NEC220 yielded some budded vesicles (Fig. 5b) as

well as unbudded LUVs containing protein deposits on the outer surface (Fig. 5c). This

observation is consistent with the confocal microscopy results (Fig. 2c, d). Some of the budded

vesicles contained ~17-nm thick internal coats (Fig. 5b), consistent with UL25 bound to the NEC

coat whereas other budded vesicles contained only the internal NEC coats and no UL25 (data not

shown). Unbudded vesicles contained heterogeneous protein deposits >25 nm in thickness (Fig.

5c), reminiscent of UL25Δ73 aggregates on unbudded vesicles observed by confocal microscopy

(Fig. 2d).

3.3.9 UL25Δ44 Q72A forms a star-shaped net and induces a pentagonal arrangement within the NEC array.

Interactions between UL25Δ44 Q72A and NEC220 were visualized in three dimensions by cryoET

(Fig. 6). Sub-tomographic averaging of the 3D reconstructions of unbudded LUVs coated with

NEC220 and UL25Δ44 Q72A (Fig. 6a) revealed that UL25Δ44 Q72A formed a "net" of five-pointed

stars on top of the membrane-bound NEC220 layer (Fig. 6b). Unexpectedly, UL25Δ44 Q72A-

bound NEC220 was arranged into a pentagonal lattice (Fig. 6c). This is in contrast to the

hexagonal lattice of NEC220 observed within budded vesicles both *in vitro,* as visualized by

cryoET both in our earlier work[13] and confirmed here (Fig. 6b), and *in vivo* within thin sections of

uninfected cells expressing the NEC[15]. The NEC pentagon is ~10.5 nm wide (Fig. 6d), similar to the ~11 nm-wide NEC hexagons observed *in vitro*[13] (Fig. 6b). The side of the NEC pentagon is ~6.5 nm (Fig. 6d), only slightly longer than the side of the NEC hexagon, ~6.3 nm (Fig. 6b). The similar dimensions of the pentagons and the hexagons suggest that each NEC pentagon is composed of five copies of NEC. A side view of the NEC pentagon on the unbudded vesicles shows that the NEC density is slightly tilted relative to the plane of the membrane (Fig. 6a); by contrast, the NEC density within the budded vesicle is perpendicular to the membrane (Fig. 6b). The ability of the NEC to adopt a range of orientations may be important for inducing negative membrane curvature. Five-pointed stars of UL25 are positioned directly on top of the NEC pentagons (Fig. 6c, d), suggesting that formation of this UL25 "net" prevents the NEC from undergoing conformational changes necessary to generate negative membrane curvature and, thus, budding.

3.4 DISCUSSION

The intrinsic membrane-deforming ability of the NEC, which allows the capsid to bud at the INM during nuclear egress, has been well established. Oligomerization of the NEC into a hexagonal coat is a fundamental property of this process[16,17]. However, how the hexagonal NEC coat forms a curved coat around the capsid is unclear, both in terms of achieving a curvature and anchoring to the capsid. Here, by following up on the observation that an outer capsid protein, UL25, inhibits NEC-mediated budding *in vitro*, we have uncovered the ability of the NEC to form an alternative, pentagonal, arrangement when bound to the UL25. This arrangement may reflect how a mature capsid interacts with the NEC coat to trigger budding at the nuclear membrane. Binding of NEC to UL25 arranged around the 5-fold vertices of the capsid would not only anchor the coat to the capsid but would also promote formation of NEC pentagons within the hexagonal coat necessary to achieve appropriate coat curvature.

3.4.1 UL25/NEC interactions in vitro are characterized by both avidity and

specificity.

UL25 and the NEC – a monomer and a heterodimer, respectively – do not interact in solution but do interact when the NEC is immobilized, either on the membrane surface or on an SPR chip. This suggests the importance of avidity in binding. Although the oligomeric state of the NEC prior to the onset of budding is yet unknown, it could form isolated hexamers or patches of hexagonal lattice that could serve as binding sites for UL25. Likewise, NEC on an SPR chip may form an oligomeric array conducive to UL25 binding. The interaction is specific because mutations in UL31, within a helix at the membrane-distal end of the NEC, abolish inhibition, implicating this region in pUL25 interactions. This is in line with previous reports demonstrating that capsid-bound pUL25 is capable of binding pUL31[9,26]. Charged residues within the membrane-distal region of UL31 have also been implicated in capsid interactions in PRV[33].

3.4.2 UL25 inhibits NEC-mediated budding in vitro by forming a star-shaped "net" over the membrane-bound NEC layer.

We observed that a N-terminally truncated UL25, UL25Δ44 Q72A, which is composed of a globular core and a long N-terminal helix, inhibited NEC budding *in vitro* by interacting with the NEC220 bound to the outer surface of lipid vesicles. UL25 forms five-pointed stars positioned directly on top of the NEC pentagons and linked into a net. Formation of this UL25 net would prevent the NEC from undergoing conformational changes necessary to generate negative membrane curvature and, thus, budding.

  The five-pointed stars formed by UL25 in our reconstructions resemble the five-pointed stars formed by the hetero-pentameric capsid-associated tegument complex (CATC) on the surface of HSV-1 capsids, resolved to a 3.5-Å resolution[27]. Five copies of the CATC surround each 5-fold capsid vertex (Fig. 7a). The CATC is composed of two copies of UL25, one copy of UL17, and two copies of the C-terminal portion of the tegument protein UL36[27]. The central feature of the CATC is a coiled coil composed of two parallel helices of UL25 and two parallel helices of UL36,

the latter arranged in an antiparallel fashion relative to UL25 (Fig. 7a). The two helices of UL25 have minimal interactions such that the coiled coil requires all four helices – two UL25 and two UL36 – for stability. Five copies of CATC form a five-pointed star at each capsid vertex, with 10 globular cores of UL25, arranged as loosely associated dimers, located at the center with 5 coiled-coil structures radiating out (Fig. 7a)[27]. We propose that UL25Δ44 Q72A forms a similar arrangement when bound to the NEC220 on the membrane surface: each UL25 "star" consists of 10 copies of UL25, with cores arranged in the center and 5 pairs of the N-terminal helices radiating out (Fig. 7b). Two pairs of the N-terminal helices from adjacent stars would form a four-stranded, antiparallel coiled coil (Fig. 7b) reminiscent of the UL25/UL36 coiled coil in CATC (Fig. 7a). The UL25 stars formed *in vitro* are smaller than the CATC stars on the capsid, and we hypothesize that this compaction is achieved both by a closer packing of the UL25 cores within the star center and by a closer positioning of the N-terminal helix relative to the UL25 core due to the flexible linker connecting them.

Regardless of the oligomeric nature of UL25, the central feature of our model is that the interactions between the N-terminal helices bridge the neighboring UL25 copies into a "net" over the NEC layer. Our results indicate that the length of the N-terminal helix is important for forming stable connections between the neighboring UL25 stars. The sequence of the N-terminal helix is less important than the length because mutant UL25 construct containing a helix with a scrambled sequence (UL25$^{45-94}_{scr}$-UL25$^{95-580}$) could still inhibit budding even if less efficiently than the WT UL25Δ44 Q72A. Thus, UL25/NEC interactions are likely mediated by the UL25 core whereas UL25/UL25 interactions are mediated by the N-terminal helix. UL25 constructs with the truncated N-terminal helix, e.g., UL25Δ73, can bind the NEC but do not inhibit budding because they cannot form stable UL25/UL25 interactions. UL25 constructs with a truncated helix were also prone to aggregation, so truncations may destabilize the helix.

3.4.3 UL25-derived peptides block NEC-mediated budding by an unknown mechanism.

We found that UL25-derived peptides efficiently inhibited NEC-mediated budding *in vitro*. Their inhibitory mechanism is yet uncharacterized, and the inhibitory peptides likely interact with the NEC differently from UL25. We showed that the length, rather than the sequence, of UL25-derived peptides was important for inhibition because only the 50-amino-acid peptides, be they native or scrambled, efficiently inhibited budding. Interestingly, an engineered ~50-amino-acid helical peptide poorly inhibited budding. CD analysis showed that the inhibitory UL25-derived peptides were largely unstructured in solution and became α-helical in the presence of TFE. By contrast, the engineered peptide was α-helical in solution even in the absence of the TFE. It is plausible that to be an effective inhibitor, the peptide needs to be unstructured until it binds the NEC, forming a helix in an induced-fit manner. Nevertheless, the discovery that UL25-derived peptides can inhibit NEC-mediated budding opens opportunities for developing peptide and small-molecule inhibitors of the nuclear egress process.

3.4.4 Inhibition of NEC-mediated budding by UL25 is likely an in vitro phenomenon.

Although the NEC has a robust budding ability *in vitro* and in uninfected cells expressing the NEC, to prevent premature, non-productive budding in infected cells, this budding ability must be inhibited, presumably by a viral protein, until the capsid comes along. It is thus tempting to speculate that free UL25 found in the nucleus may contribute to inhibition during infection. In this scenario, free UL25 would bind the NEC and inhibit its budding activity until being displaced by the capsid-bound UL25. If free UL25 inhibited the NEC activity, it would be expected to accumulate at the nuclear rim, but this has not yet been observed. Therefore, inhibition of NEC-mediated budding by UL25 could, instead, be purely an *in vitro* phenomenon.

3.4.5 UL25 binding changes the oligomeric state of the NEC.

The ability of the NEC to oligomerize into a hexagonal lattice *in vitro* and *in vivo* is well documented[13,15,16] and is an important feature of its membrane deformation mechanism[13,16,17]. But,

a strictly hexagonal arrangement is flat, so curvature must be achieved through the inclusion of either regular insertions of a different geometry, for example, pentagons – as observed in icosahedral capsids – or irregular insertions. Neither have yet been visualized in NEC coats[15,24,25], likely, due to technical limitations. In perinuclear viral particles visualized in PRV-infected cells, the NEC coats appear tightly associated with the capsid[15]. Capsidless perinuclear vesicles formed in uninfected cells expressing PRV NEC[15] are relatively uniform in size (~115 nm in diameter) but smaller than the capsid (~125 nm in diameter[27,34]) or the perinuclear vesicles isolated from cells infected with the HSV-1 US3-null mutant virus (~160 nm in diameter[25]) (Fig. 7c). The capsid thus appears to define the size of NEC-budded vesicles during infection, and it is plausible that the capsid geometry may influence the geometry of the NEC coat.

We found that when bound to UL25, the NEC formed a pentagonal, rather than a hexagonal, lattice on the membrane surface. This established that the NEC can form different kinds of oligomers and that the oligomeric state of the NEC is influenced by binding to UL25. We believe that this finding can explain both how the curvature of the coat is achieved and how the NEC coat is anchored to the capsid. We propose that the NEC binds the UL25 core (within the CATC) at the 5-fold vertices and that this promotes formation of NEC pentagons, which attach the coat to the capsid vertices through increased avidity (Fig. 7c). In the HSV-1 capsid structure, the cores of 5 neighboring copies of UL25 at the vertex are too far apart to bind an NEC pentagon (Fig. 7a). However, unlike the N-terminal helices of UL25 that are anchored on the capsid as CATC components, the cores are dynamic – being connected to the N-terminal helices by long, flexible linkers – and could move closer together to interact properly with the NEC pentagons. The incorporation of pentagons into the hexagonal lattice would also facilitate the formation of the curvature in the coat growing around the capsid. An alternative strategy for achieving appropriate coat curvature is through the inclusion of irregular defects within the NEC hexagonal coat. This has been observed in the immature HIV-1 capsid formed by the Gag protein[20,23] and by the D13 protein that forms the early poxvirus envelope[21,22]. Although neither irregular defects nor

pentagons have yet been observed in any NEC coats analyzed thus far, based on our observation that the NEC can form pentagons when bound to the capsid protein UL25, we propose that during nuclear egress, NEC coat of appropriate curvature is formed around the capsid through the inclusion of pentagons.

We do not yet understand how the NEC coat curvature is achieved in the absence of capsid. Given that the hexagonal NEC coats formed in *in vitro* or in NEC-expressing cells have a smaller diameter than those formed around the capsid (Fig. 7c), they may achieve coat curvature by other means, for example, by having irregular lattice defects.

3.4.6 A model of NEC-mediated budding in the presence or absence of UL25, in vitro and during infection.

Our observations of NEC/UL25 interactions and their effect on NEC-mediated budding has led to the following model of NEC-mediated budding in the presence or absence of UL25, *in vitro* and during infection (Fig. 8). *In vitro*, the NEC-mediated membrane budding leads to the formation of negative membrane curvature and the internal NEC coats on the budded vesicles that are mostly hexagonal but contain yet uncharacterized defects (Fig. 8a). UL25Δ44, present in excess, binds the NEC on the membrane surface and forms a net of five-pointed stars that restricting conformational changes needed to induce negative curvature leading to the overall effect of budding inhibition (Fig. 8b). Binding of UL25Δ44 to the NEC remodels the latter into a pentagonal arrangement (Fig. 8b). Inhibition requires the intact N-terminal helix of UL25, and UL25 construct with truncated N-terminal helix (UL25Δ73) binds the NEC but cannot inhibit budding (Fig. 8c). Truncated UL25 constructs are also prone to aggregation (Fig. 8c). During capsid assembly within the nucleus of infected cells, UL25 is recruited to form pentagonal CATC structures on the 12 vertices [35,36]. When the capsid reaches the INM, binding of the NEC to UL25 at the capsid vertices would anchor the forming NEC coat to the capsid and promote formation of NEC pentagons (Fig. 8d). The incorporation of pentagons into the hexagonal lattice would also facilitate the formation of

the curved NEC coat around the capsid.

In both HSV-1 and PRV, removal of UL25 results in an accumulation of capsids at the INM, unable to undergo egress[37,38]. Our results suggest that UL25 not only anchors the NEC coat to the capsid but also contributes to formation of a curved coat. Additionally, our results may explain why mostly mature, DNA-containing C-capsids undergo budding at the INM[39,40]. A- and B- capsids have fewer UL25 copies on the capsid surface[41], and we hypothesize that only C-capsid, which contain UL25 at a full occupancy can generate pentagonal NEC insertions necessary for the formation of an NEC coat around the capsid, thereby acting as a checkpoint during nuclear egress.

3.5 ACKNOWLEDGMENTS

3.6 AUTHOR CONTRIBUTIONS

E.B.D. and E.E.H. designed and coordinated the project; E.B.D. performed the experiments (with the exception of the ITC and size-exclusion experiments) under the guidance of E.E.H; E.B.D. and J.Z. collected cryoET data under the guidance of Z.H.Z; J.Z. processed the cryoET data; all authors analyzed the data, interpreted the results and wrote the manuscript.

3.7 COMPETING INTERESTS

The authors declare no competing interests.

3.8 DATA AVAILABILITY STATEMENT

The EM datasets generated in this study will be deposited into the Electron Microscopy Data Bank and will be immediately available upon publication.

3.9 METHODS

3.9.1 Cloning.

All primers used in cloning are listed in Supplementary Table S2. Codon-optimized UL25 gene from HSV-1 strain KOS was synthesized by GeneArt. Digested PCR fragments encoding UL25Δ44 were subcloned by restriction digest into the pJP4 plasmid, which contains a His_6-SUMO-PreScission tag in frame with the BamHI restriction site of the multiple-cloning site in a pET24b vector, creating the pJB104 plasmid. DNA fragments encoding UL25Δ50, UL25Δ73, or UL25Δ133 were amplified by PCR from pJB104 (UL25Δ44) and subcloned into pJP4 by restriction digest using BamHI and XhoI, creating the UL25Δ50 (pED13), UL25Δ73 (pJB123), and UL25Δ133 (pED31) plasmids. A UL25Δ133 construct with an N-terminal GSGS linker immediately after the BamHI restriction site and before the UL25Δ133 DNA sequence was also generated (pED32), as described above, for eGFP cloning (described below). Site-directed mutagenesis of pJB104 yielded the UL25Δ44 Q72A mutant plasmid (pED03).

DNA encoding the eGFP sequence was PCR amplified out of the eGFP-N2 plasmid (Clontech) and subcloned via single-cut restriction digest into the corresponding UL25 plasmid harboring the cleavable His_6-SUMO tag [(either UL25Δ50 (pED13), UL25Δ73 (pJB123), or GSGS-UL25Δ133 (pED32)] creating either the eGFP-UL25Δ50 (pED14), eGFP-UL25Δ73 (pED05), or eGFP-UL25Δ133 (pED33) constructs. For eGFP-UL25Δ133 (pED33), the GSGS-UL25Δ133 construct was used to create a linker space between the eGFP and UL25Δ133 proteins.

DNA sequences encoding each peptide were subcloned into the prokaryotic expression vector pGEX-6P-1 that encodes a N-terminal GST-tag followed by a PreScission Protease cleavage site in frame with the BamHI restriction site within the multiple cloning site. Both the $UL25^{45-94}_{native}$ and $UL25^{74-94}$ peptide DNA sequences were amplified by PCR from pED03 and subcloned into the pGEX-6P-1 vector using restriction digest via BamHI and XhoI to create the pED23 ($UL25^{45-94}_{native}$) and pED38 ($UL25^{74-94}$) plasmids. Both the scrambled ($UL25^{45-94}_{scr}$) and α-helix ($E_{16}A_{22}K_{15}$) peptide DNA sequences were obtained as a gBlock gene fragment (IDT) and

subjected to restriction digest cloning with BamHI and XhoI into the pGEX-6P-1 vector (gene fragment sequences are given in Supplementary Table 2) to create the pED34 (UL25$^{45-94}_{scr}$ peptide) and pED39 ($E_{16}A_{22}K_{15}$ peptide) plasmids.

For the UL25$^{45-94}_{scr}$ -UL25 95-580 fusion construct, the UL25 95-580 DNA sequence was PCR amplified from pED03 and subcloned with restriction digest into pJP4 to create pED36 (UL25 95-580). The scrambled peptide-UL25 95-580 fusion DNA insert was then cloned out of pED34 (UL25$^{45-94}_{scr}$ plasmid) using primers designed for Gibson assembly (NEB) (Supplementary Table S2). The fusion construct was created by Gibson assembly using a BamHI-digested pED36 vector and the subcloned fusion insert following the manufacturer's protocol.

Site-directed mutagenesis of pKH90 (UL31 1-306) using a splicing by overlap extension protocol[42] followed by restriction digest into the pJP4 vector was used to create the UL31 D275A/C278A/K279A/D282A mutant (pJB118) in the capsid binding mutant construct (NEC-CBM).

3.9.2 Expression and purification of NEC constructs.

Plasmids encoding HSV-1 UL31 1-306 (pKH90) and either UL34 1-220 (pJB02) or UL34 1-220-8x-His (pJB57) were co-transformed into *Escherichia coli* BL21(DE3) LoBSTr cells (Kerafast) to generate either NEC220 or NEC220-8x-His, respectively[13]. Plasmids encoding HSV-1 UL31 1-306 D275A/C278A/K279A/D282A (pJB118) and UL34 1-220 (pJB02) were co-transformed into *Escherichia coli* BL21(DE3) LoBSTr cells (Kerafast) to generate NEC-CBM. All constructs were expressed using autoinduction at 37 °C in Terrific Broth (TB) supplemented with 100 µg/mL kanamycin, 100 µg/mL ampicillin, 0.2% lactose and 2 mM MgSO$_4$ for 4 h. The temperature was then reduced to 25 °C for 16 h. Cells were harvested at 5,000 x g for 30 min. NEC proteins were purified as previously described[13] with slight modifications. The NEC220 and NEC-CBM constructs were passed over 2 x 1 mL HiTrap Talon columns (GE Healthcare), rather than ion exchange as previously described, to remove excess cleaved His$_6$-SUMO before injection onto size-exclusion chromatography (as previously described). The NEC220-8x-His construct was injected

immediately onto size-exclusion chromatography (as previously described[13], without going over

ion-exchange, after overnight tag cleavage.

3.9.3 Expression and purification of UL25 constructs and peptides.

Plasmids encoding either HSV-1 UL25 or eGFP-UL25 constructs were transformed into *E. coli*

BL21(DE3) LoBSTr cells and expressed using autoinduction at 37 °C in TB supplemented with

100 μg/mL kanamycin, 0.2% lactose, and 2 mM MgSO$_4$ for 4 h. The temperature was then

reduced to 25 °C for 16 h. Cells were harvested at 5,000 x g for 30 min. All purification steps were

performed at 4 °C. UL25 constructs were purified in lysis buffer (50 mM Na HEPES pH 7.5, 500

mM NaCl, 1 mM TCEP, and 10% glycerol). Cells were resuspended in lysis buffer supplemented

with Complete protease inhibitor (Roche) and lysed with a microfluidizer (Microfluidics). The cell

lysate was clarified by centrifugation at 13,000 x g for 35 min and was passed over Ni-NTA

sepharose (GE Healthcare) column. The column was subsequently washed with 20 mM and 40

mM imidazole lysis buffer and bound proteins were eluted with 250 mM imidazole lysis buffer. The

His$_6$-SUMO tag was cleaved for 16 h using PreScission Protease produced in-house from a GST-

PreScission fusion protein expression plasmid. As a final purification step, UL25 constructs were

purified with size-exclusion chromatography using either a Superdex 75 or 200 column (GE

Healthcare) equilibrated with 20 mM Na HEPES, pH 7.0, 100 mM NaCl, and 1 mM TCEP. The

UL25 constructs were purified to homogeneity as assessed by 12% SDS-PAGE and Coomassie

staining. Fractions containing UL25 were concentrated up to ~30 mg/mL and stored at -80 °C to

prevent degradation observed at 4 °C. Protein concentration was determined by absorbance

measurements at 280 nm. The typical yield was 35 mg/L of TB culture.

Plasmids encoding the peptide sequences were transformed into *E. coli* BL21(DE3) LoBSTr

cells and were expressed as described above for the UL25 protein constructs with the exception

that 100 μg/mL ampicillin was used. Peptides were purified in lysis buffer (50 mM Na HEPES pH

7.5, 500 mM NaCl, 1 mM TCEP, and 10% glycerol). Cells were resuspended in lysis buffer

supplemented with Complete protease inhibitor (Roche) and lysed with a microfluidizer. The cell

lysate was clarified by centrifugation at 13,000 x g for 35 min, passed over a Glutathione

sepharose 4B (GE Healthcare) column, and the column was subsequently washed with lysis

buffer. The GST-tag was cleaved on the glutathione sepharose column for 16 h using the

PreScission Protease described above. Peptides were eluted off the column with lysis buffer. As a

final purification step, peptides were purified with size-exclusion chromatography using a Superdex

75 column (GE Healthcare) equilibrated with 20 mM Na HEPES, pH 7.0, 100 mM NaCl, and 1 mM

TCEP (for budding assays) or with 10 mM sodium phosphate, pH 7.4, and 100 mM NaF (for CD

studies). Absorbance was monitored at 214 nm because the peptides lack aromatic residues. The

peptides were purified to homogeneity as assessed by both 16.5% Tris-Tricine SDS-PAGE and

12% SDS-PAGE with Coomassie staining. Fractions containing peptide were concentrated and

sent off for amino acid analysis (University of Colorado Denver Anschutz Medical Campus) for

accurate concentration determination.

3.9.4 Circular dichroism (CD) studies.

Far-UV CD spectra of UL25 proteins and peptides (0.1 mg/mL) were recorded in 10 mM Na

phosphate, pH 7.4, and 100 mM NaF buffer using a Jasco 815 CD Spectropolarimeter at the

Center for Macromolecular Interactions at Harvard Medical School. Peptides spectra were also

measured in the presence of trifluoroethanol (TFE). Data were collected at ambient temperature

with a scan speed of 50 nm/min and 5 accumulations of each sample was averaged. The raw data

was blank subtracted and converted to mean residue ellipticity (θ). Helical content was estimated

using DichroWeb[30] and the values are provided in Supplementary Table S1.

3.9.5 Co-sedimentation assay.

Co-sedimentation of UL25Δ44 to acidic multilamellar vesicles (MLVs) was performed as previously

described[13]. MLVs were prepared in a 3:1:1 ratio of POPC:POPS:POPA. Background signal in the

absence of liposomes is due to protein aggregation during centrifugation.

3.9.6 In vitro GUV budding assays.

Giant unilamellar vesicles (GUVs) were prepared as previously described[13]. For NEC220 only budding quantification, a total of 10 µL of GUVs with a 3:1:1 ratio of POPC:POPS:POPA containing ATTO-594 DOPE at a concentration of 0.2 µg/µL was mixed with 1 µM NEC220 (final concentration), and 0.2 mg/mL (final concentration) Cascade Blue Hydrazide (ThermoFisher Scientific). For the NEC and UL25 titration experiments, 10 µL of GUVs and either 1, 6, or 10 µM of UL25Δ44 Q72A, UL25Δ73, or UL25Δ133 (final concentration) were incubated with 1 µM of NEC220 (final concentration) along with Cascade Blue. For NEC-CBM and UL25 titration experiments, 10 µL of GUVs and either 1, 6, or 10 µM of UL25Δ44 Q72A (final concentration) were incubated with 1 µM of NEC-CBM (final concentration) along with Cascade Blue. For the NEC and peptide experiments, 10 µL of GUVs and 10 µM (final concentration) of either the $UL25^{45-94}_{native}$, $UL25^{74-94}$, $UL25^{45-94}_{scr}$, or $E_{16}A_{22}K_{15}$ peptides were incubated with 1 µM of NEC220 (final concentration) along with Cascade blue. The total volume of each sample during imaging for all experiments was brought to 100 µL with gel filtration buffer and the reaction was incubated for 5 min at 20 °C. Samples were imaged in a 96-well chambered cover-glass. Images were acquired using a Nikon A1R Confocal Microscope with a 60x oil immersion lens at the Tufts Imaging Facility in the Center for Neuroscience Research at Tufts University School of Medicine. Images of NEC budding in the presence of eGFP-UL25 constructs were recorded after incubation of 10 µL of GUVs with either 5 or 10 µM (final concentration) of either eGFP-UL25Δ50 Q72A or eGFP-UL25Δ73 and 1 µM of NEC220 (final concentration). Quantification was performed by counting vesicles in 15 different frames of the sample (~300 vesicles total). Each condition was tested in at least two biological replicates. Prior to analysis, the background was subtracted from the raw values. The reported values represent the average budding activity relative to NEC220 (100%). The standard error of the mean is reported for each measurement. Significance compared to NEC220 was calculated using an unpaired one-tailed *t*-test against NEC220.

3.9.7 Isothermal titration calorimetry (ITC).

ITC measurements were recorded using a Microcal ITC200 (Malvern Panalytical) at the Center for Macromolecular Interactions at Harvard Medical School. A solution of UL25Δ44 (200 μM) was titrated into a solution of NEC220 (20 μM) in 20 mM Na HEPES, pH 7.0, 150 mM NaCl, 1 mM TCEP. Control experiments were performed by injecting UL25Δ44 into buffer. Thermograms were plotted by subtracting heats of the control experiments from the sample experiments. The data were not fit due to no detectable binding.

3.9.8 Surface plasmon resonance (SPR).

SPR measurements were recorded using a Biacore T100 with a four-channel system (GE Healthcare) in the Study Center on the Immunogenetics of Infectious Disease at the Tufts University School of Medicine. NEC220-His was immobilized onto an NTA S Series chip (GE Healthcare) in flow cell 4 (~20 RU). Flow cell 3 was used for the reference cell. The experimental buffer was 20 mM Na HEPES, pH 7.0, 150 mM NaCl, 1 mM TCEP, and 0.05% Tween-20. UL25 constructs were injected onto the chip at 30 μL/min (to avoid mass transfer) for a total of 180 s. Reference-subtracted sensorgrams were analyzed for NEC:UL25 binding.

3.9.9 Cryoelectron microscopy and tomography.

A volume of 10 μL of a 1:1 mixture of 400-nm and 800-nm large unilamellar vesicles (LUVs) made of 3:1:1 POPC:POPS:POPA [prepared as previously described[13]] were mixed on ice with a 30 μL solution of NEC220 and either UL25Δ44 Q72A or UL25Δ73, yielding an NEC:UL25 ratio of 1:10 (NEC concentration was at 1 mg/mL). After 30 min, 3 μL of sample was applied to glow-discharged (30 s) Quantifoil copper grids (R2/2, 200 mesh, Electron Microscopy Sciences), blotted on both sides for 4 s, and vitrified by rapid freezing in liquid ethane (Vitrobot). Grids were stored in liquid nitrogen until loaded into a Tecnai F20 transmission electron microscope (FEI) via a cryo holder (Gatan). The microscope was operated in low dose mode at 200 keV using SerialEM[43] and

images were recorded with a 4k x 4k charge coupled device camera (Ultrascan, Gatan) at 29,000-fold magnification (pixel size: 0.632 nm). 2D cryo-EM images were recorded at defocus values of -4 to -8 μm and an electron dose ~15 e/$Å^2$. Images are displayed using ImageJ[44].

For single-axis cryoET data used to generate 3D EM data, samples were incubated on ice for 30 min, and 0.8 μL of 10 nm colloidal gold coated with protein A (Cell Microscopy Core, University Medical Center Utrecht, Department of Cell Biology) was added to the solution and mixed. The mixture (2.5 μL) was applied to freshly glow-discharged (30 s) Quantifoil R 3.5/1 grids (Electron Microscopy Sciences) and manually blotted before being flash-frozen in liquid ethane. Grids were loaded into a FEI Titan Krios electron microscope equipped with a Gatan imaging filter (GIF) and a Gatan K2 summit direct electron detection camera (Roper Technologies, Inc.), operated at 300 kV. The acquisition for automated cryoET tilt series collection was performed using SerialEM[43]. A tilt series was collected in which the sample was tilted from 0° to +60° degrees and then from 0° to -60°, each in a stepwise fashion with 2° increments. Tilt series were acquired at a magnification of x53,000 (corresponding to a calibrated pixel size of 2.6 Å) with a maintained defocus value of -3 to -4 μm. The total electron dose was ~100 e/$Å^2$.

3.9.10 3D reconstruction and subtomographic averaging.

The detailed steps of the 3D reconstruction and subtomographic averaging were previously described[45]. Briefly, frames from each recorded tilt series were drift-corrected and averaged with *Motioncorr*[43] and was further reconstructed with contrast transfer function (CTF) correction using the IMOD software package[46]. Two resulting tomograms were produced by the weighted back projection and simultaneous iterative reconstruction technique (SIRT) methods. A total of 1200 particles were picked for tomograms containing LUVs, NEC220 and UL25Δ44 Q72A. 3D sub-tomographic averaging was completed as described[45] using the PEET (particle estimation for electron tomography) software[47]. Five-fold symmetry was only applied after five-fold symmetry was apparent in the averaged structure. The original dataset was split into two separate groups, even

group and odd group, and averaged independently. Gold standard Fourier Shell Correlation (FSC) analysis for the averaged structure was performed by *calcUnbiasedFSC* in PEET when the two averaged structures converged. The reported resolution is 29 Å based on the 0.143 gold-standard FSC criterion. EM maps will be deposited into the Electron Microscopy Data Bank (EMDB) for immediate access upon publication.

3.10 References

1.   Johnson, D.C. & Baines, J.D. Herpesviruses remodel host membranes for virus egress. *Nat. Rev. Microbiol.* **9**, 382-394 (2011).

2.   Bigalke, J.M. & Heldwein, E.E. Nuclear exodus: Herpesviruses lead the way. *Annu. Rev. Virol.* **3**, 387-409 (2016).

3.   Roller, R.J. & Baines, J.D. Herpesvirus nuclear egress. *Adv. Anat. Embryol. Cell. Biol.* **223**, 143-169 (2017).

4.   Bigalke, J.M. & Heldwein, E.E. Have NEC coat, will travel: Structural basis of membrane budding during nuclear egress in herpesviruses. *Adv. Virus. Res.* **97**, 107-141 (2017).

5.   Mettenleiter, T.C., Muller, F., Granzow, H. & Klupp, B.G. The way out: what we know and do not know about herpesvirus nuclear egress. *Cell. Microbiol.* **15**, 170-8 (2013).

6.   Shiba, C. et al. The UL34 gene product of herpes simplex virus type 2 is a tail-anchored type II membrane protein that is significant for virus envelopment. *J. Gen. Virol.* **81**, 2397-405 (2000).

7.   Chang, Y.E. & Roizman, B. The product of the UL31 gene of herpes simplex virus 1 is a nuclear phosphoprotein which partitions with the nuclear matrix. *J. Virol.* **67**, 6348-56 (1993).

8.   Reynolds, A.E. et al. U(L)31 and U(L)34 proteins of herpes simplex virus type 1

        form a complex that accumulates at the nuclear rim and is required for

        envelopment of nucleocapsids. *J. Virol.* **75**, 8803-17 (2001).

9.   Yang, K. & Baines, J.D. Selection of HSV capsids for envelopment involves

        interaction between capsid surface components pUL31, pUL17, and pUL25. *Proc.*

        Natl Acad. Sci. USA **108**, 14276-81 (2011).

10.   Trus, B.L. et al. Allosteric signaling and a nuclear exit strategy: binding of

        UL25/UL17 heterodimers to DNA-Filled HSV-1 capsids. *Molecular cell* **26**, 479-89

        (2007).

11.   Roller, R.J., Zhou, Y., Schnetzer, R., Ferguson, J. & DeSalvo, D. Herpes simplex

        virus type 1 U(L)34 gene product is required for viral envelopment. *J. Virol.* **74**, 117-

        29 (2000).

12.   Fuchs, W., Klupp, B.G., Granzow, H., Osterrieder, N. & Mettenleiter, T.C. The

        interacting UL31 and UL34 gene products of pseudorabies virus are involved in

        egress from the host-cell nucleus and represent components of primary enveloped

        but not mature virions. *J. Virol.* **76**, 364-78 (2002).

13.   Bigalke, J.M., Heuser, T., Nicastro, D. & Heldwein, E.E. Membrane deformation

        and scission by the HSV-1 nuclear egress complex. *Nat. Commun.* **5**, 4131 (2014).

14.   Lorenz, M. et al. A single herpesvirus protein can mediate vesicle formation in the

        nuclear envelope. *J. Biol. Chem.* **290**, 6962-6974 (2015).

15. Hagen, C. et al. Structural basis of vesicle formation at the inner nuclear membrane. *Cell* **163**, 1692-701 (2015).

16. Bigalke, J.M. & Heldwein, E.E. Structural basis of membrane budding by the nuclear egress complex of herpesviruses. *The EMBO journal* **34**, 2921-36 (2015).

17. Roller, R.J., Bjerke, S.L., Haugo, A.C. & Hanson, S. Analysis of a charge cluster mutation of herpes simplex virus type 1 UL34 and its extragenic suppressor suggests a novel interaction between pUL34 and pUL31 that is necessary for membrane curvature around capsids. *J Virol* **84**, 3921-34 (2010).

18. Arii, J. et al. Roles of the Interhexamer Contact Site for Hexagonal Lattice Formation of the Herpes Simplex Virus 1 Nuclear Egress Complex in Viral Primary Envelopment and Replication. *J Virol* **93**(2019).

19. Zandi, R., Reguera, D., Bruinsma, R.F., Gelbart, W.M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proc Natl Acad Sci U S A* **101**, 15556-60 (2004).

20. Briggs, J.A. et al. Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A* **106**, 11090-5 (2009).

21. Heuser, J. Deep-etch EM reveals that the early poxvirus envelope is a single membrane bilayer stabilized by a geodetic "honeycomb" surface coat. *The Journal of cell biology* **169**, 269-83 (2005).

22. Hyun, J.K. et al. Membrane remodeling by the double-barrel scaffolding protein of poxvirus. *PLoS Pathog* **7**, e1002239 (2011).

23.   Schur, F.K. et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. *Nature* **517**, 505-8 (2015).

24.   Bigalke, J.M., Heuser, T., Nicastro, D. & Heldwein, E.E. Membrane deformation and scission by the HSV-1 nuclear egress complex. *Nat Commun* **5**, 4131 (2014).

25.   Newcomb, W.W. et al. The primary enveloped virion of herpes simplex virus 1: Its role in nuclear egress. *mBio* **8**(2017).

26.   Yang, K., Wills, E., Lim, H.Y., Zhou, Z.H. & Baines, J.D. Association of herpes simplex virus pUL31 with capsid vertices and components of the capsid vertex-specific complex. *J Virol* **88**, 3815-25 (2014).

27.   Dai, X. & Zhou, Z.H. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. *Science (New York, N.Y.)* **360**(2018).

28.   Furlong, D. Direct evidence for 6-fold symmetry of the herpesvirus hexon capsomere. *Proc Natl Acad Sci U S A* **75**, 2764-6 (1978).

29.   Bowman, B.R. et al. Structural characterization of the UL25 DNA-packaging protein from herpes simplex virus type 1. *J Virol* **80**, 2309-17 (2006).

30.   Whitmore, L. & Wallace, B.A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392-400 (2008).

31.   Luo, P. & Baldwin, R.L. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **36**, 8413-21 (1997).

32. Zhou, N.E., Kay, C.M., Sykes, B.D. & Hodges, R.S. A single-stranded amphipathic alpha-helix in aqueous solution: design, structural characterization, and its application for determining alpha-helical propensities of amino acids. *Biochemistry* **32**, 6190-7 (1993).

33. Ronfeldt, S., Klupp, B.G., Franzke, K. & Mettenleiter, T.C. Lysine 242 within helix 10 of the pseudorabies virus nuclear egress complex pUL31 component is critical for primary envelopment of nucleocapsids. *J Virol* (2017).

34. Liu, Y.T. et al. A pUL25 dimer interfaces the pseudorabies virus capsid and tegument. *J Gen Virol* **98**, 2837-2849 (2017).

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

36. McElwee, M., Vijayakrishnan, S., Rixon, F. & Bhella, D. Structure of the herpes simplex virus portal-vertex. *PLoS biology* **16**, e2006191 (2018).

37. Klupp, B.G., Granzow, H., Keil, G.M. & Mettenleiter, T.C. The capsid-associated UL25 protein of the alphaherpesvirus pseudorabies virus is nonessential for cleavage and encapsidation of genomic DNA but is required for nuclear egress of capsids. *J Virol* **80**, 6235-46 (2006).

38. Kuhn, J. et al. Partial functional complementation of a pseudorabies virus UL25 deletion mutant by herpes simplex virus type 1 pUL25 indicates overlapping functions of alphaherpesvirus pUL25 proteins. *J Virol* **82**, 5725-34 (2008).

39.   Klupp, B.G., Granzow, H. & Mettenleiter, T.C. Nuclear envelope breakdown can substitute for primary envelopment-mediated nuclear egress of herpesviruses. *J Virol* **85**, 8285-92 (2011).

40.   Roizman, B. & Furlong, D. The Replication of Herpesviruses. in *Comprehensive Virology* (eds. Fraenkel-Conrat, H. & Wagner, R.R.) 229-403 (Springer, Boston, MA, 1974).

41.   Newcomb, W.W., Homa, F.L. & Brown, J.C. Herpes simplex virus capsid structure: DNA packaging protein UL25 is located on the external surface of the capsid near the vertices. *J Virol* **80**, 6286-94 (2006).

42.   Heckman, K.L. & Pease, L.R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature protocols* **2**, 924-32 (2007).

43.   Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* **152**, 36-51 (2005).

44.   Schindelin, J., Rueden, C.T., Hiner, M.C. & Eliceiri, K.W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular reproduction and development* **82**, 518-29 (2015).

45.   Si, Z. et al. Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLoS Pathog* **14**, e1007452 (2018).

46.   Kremer, J.R., Mastronarde, D.N. & McIntosh, J.R. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology* **116**, 71-6 (1996).

47.   Nicastro, D. et al. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944-8 (2006).

48.   Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **47**, W636-w641 (2019).

49.   Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-4 (2014).

3.11 Figures
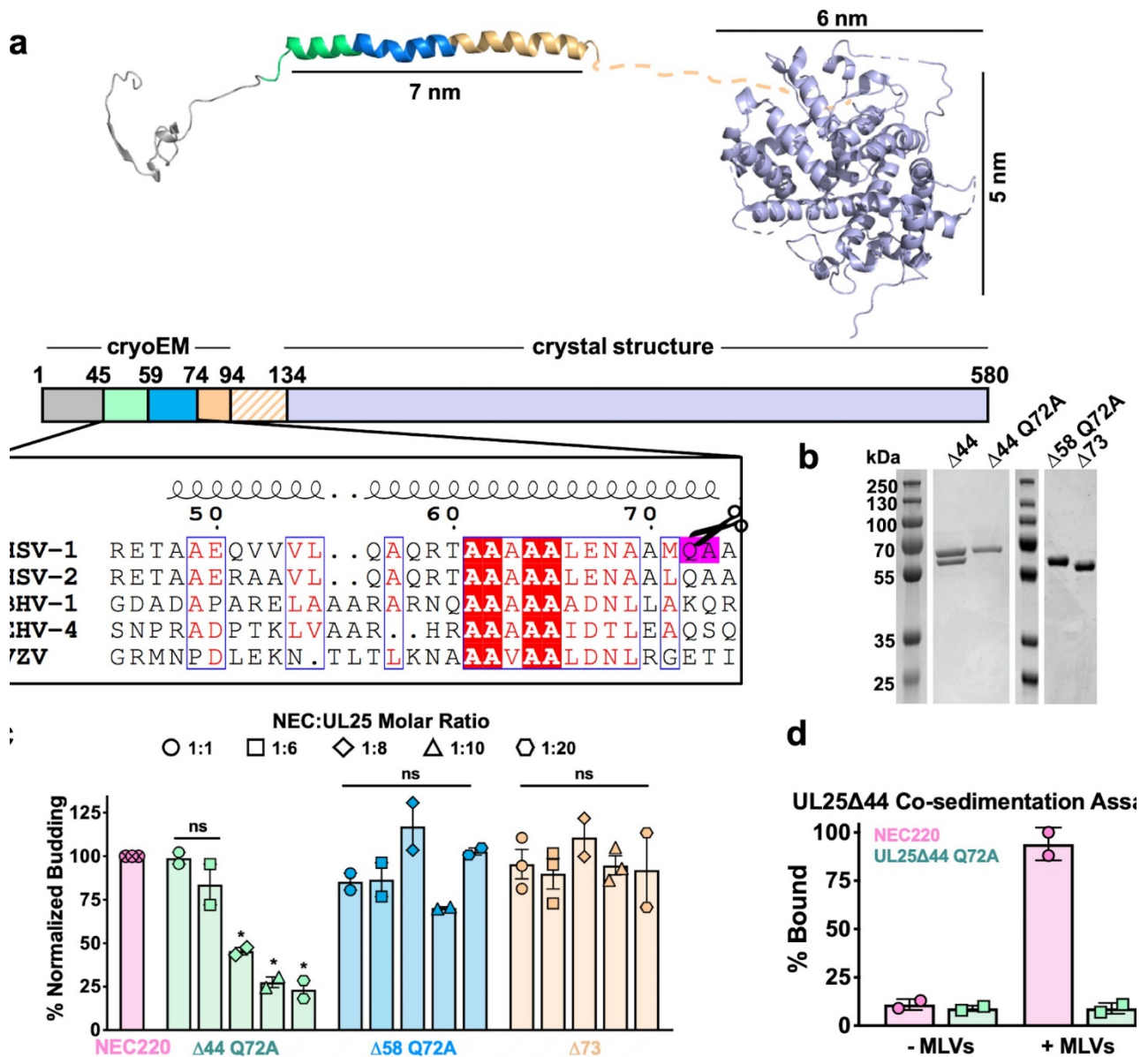
Figure 3- 1. Inhibition of NEC-mediated budding by UL25 constructs.

(a) The UL25 structure and a diagram of domain organization is shown along with a multiple sequence alignment of UL25 residues 45–74 from five alphaherpesviruses. Sequence alignment was generated using Clustal Omega45 and displayed using ESPript 3.046. Identical residues are shown as white letters on a red background. Similar residues are shown as red letters in a blue box. Secondary structure derived from the cryoEM reconstruction of capsid-bound HSV-1 UL25 is shown above the alignment. The following herpesvirus sequences were used (GenBank GeneID numbers in parentheses): HSV-1, herpes simplex virus type 1, strain 17 (2703377); HSV-2, herpes

simplex virus type 2, strain HG52 (1487309); BHV-1, bovine herpesvirus-1 (4783418); EHV-4, equine herpesvirus-4, strain NS80567 (1487602); and VZV, varicella-zoster virus, strain Dumas (1487687).

(b) SDS-PAGE of purified UL25 constructs: UL25Δ44 (cleaved product; 57 kDa), UL25Δ44 Q72A (single product; 57 kDa), UL25Δ58 Q72A (56 kDa) and UL25Δ73 (54 kDa).

(c) UL25Δ44 Q72A inhibits NEC budding, whereas other UL25 constructs do not. For each condition, NEC-mediated budding was tested at 1:1, 1:6, 1:8, 1:10, and 1:20 NEC:UL25 molar ratios. Each construct was tested in at least two biological replicates, consisting of three technical replicates. Symbols show average budding efficiency of each biological replicate relative to NEC220 (100%; pink). Error bars represent the standard error of measurement for at least two individual experiments. Significance compared to NEC220 was calculated using an unpaired t-test against NEC220. *p-value<0.1. The source file with all raw data values is provided in Figure 1—source data 1.

(d) UL25Δ44 Q72A does not bind to acidic lipid membranes.
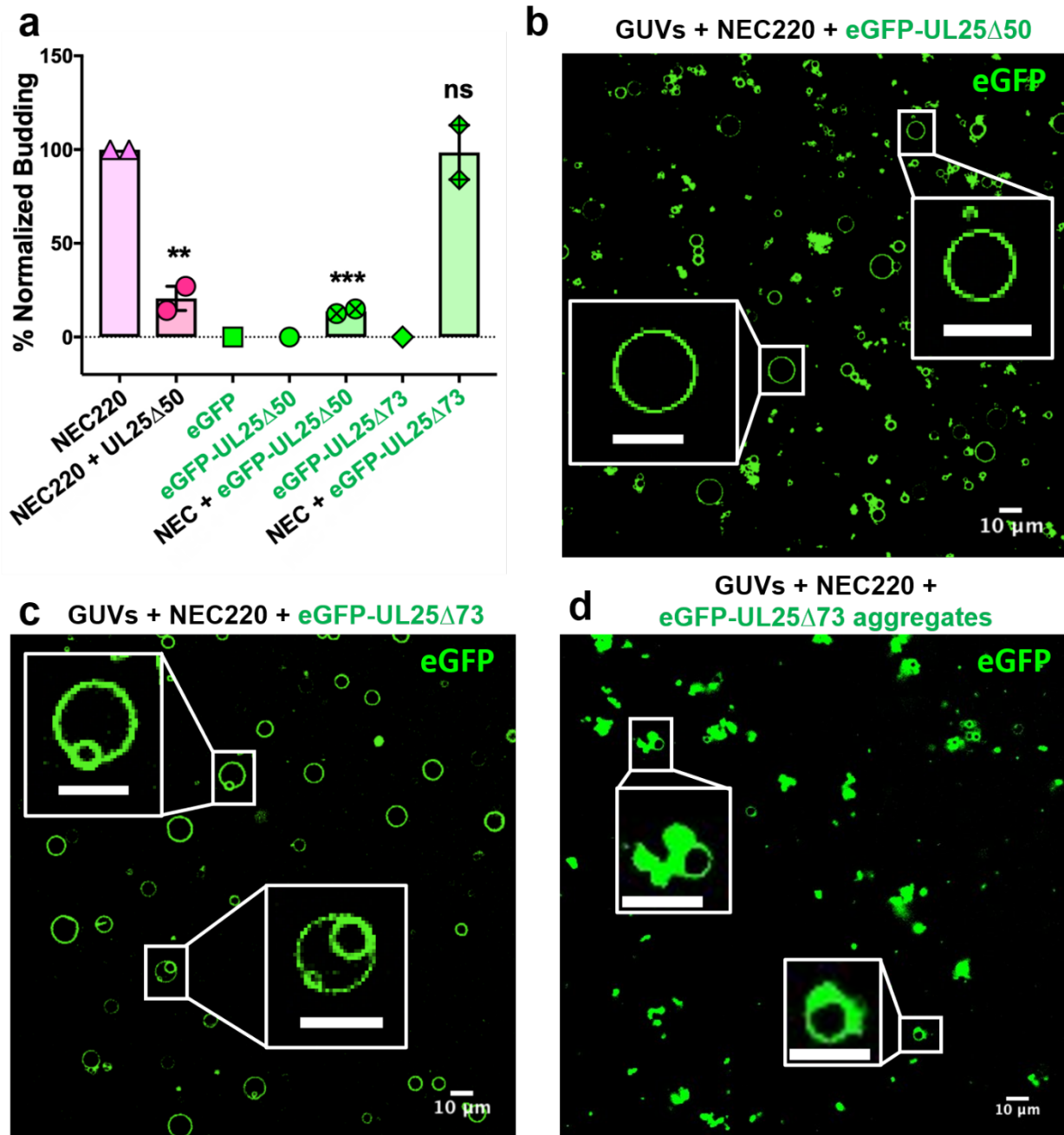
Figure 3- 2. eGFP-UL25Δ50 inhibits NEC budding while eGFP-UL25Δ73 does not.

(a) Quantification of NEC budding in the presence of either eGFP-UL25Δ50 or eGFP-UL25Δ73. Each construct (except in the absence of NEC220) was tested in at least two biological replicates, each consisting of three technical replicates. Symbols show the average budding efficiency of each biological replicate relative to NEC220 (100%). Error bars represent the standard error of measurement for at least two individual experiments. Significance compared to NEC220 was

calculated using an unpaired t-test against NEC220. **p-value<0.01 and ***p-value<0.001. The

source file with all raw data values is provided in Figure 2—source data 1.

(b) Confocal image of eGFP-UL25Δ50 bound to NEC-coated vesicles. No budding is observed.

(c) Confocal image of eGFP-UL25Δ73 either bound to or budded into vesicles with the NEC.

(d) Confocal image of eGFP-UL25Δ73 aggregating on the surface of NEC-coated vesicles. All

scale bars = 10 μm.
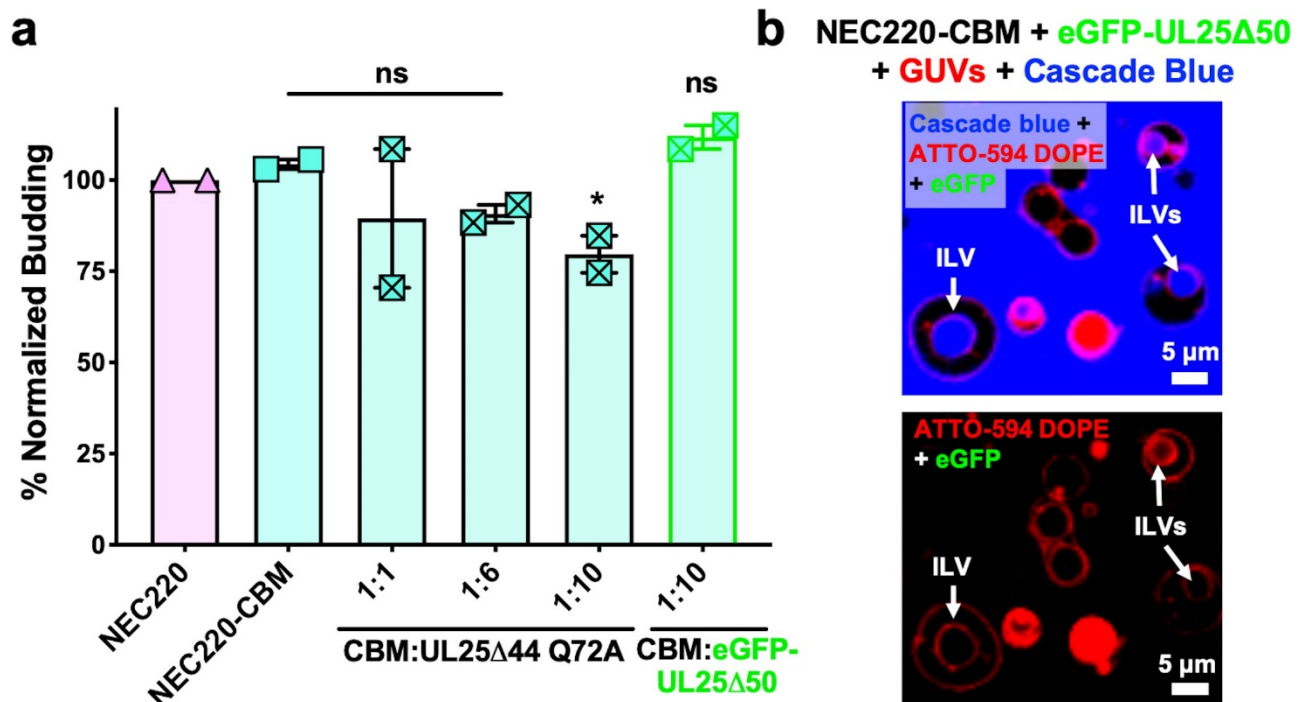
Figure 3- 3. UL25 inhibits NEC220-CBM budding to a lesser extent.

(a) NEC220-CBM budding is not inhibited to the same extent as NEC220 budding by either

UL25Δ44 Q72A or eGFP-UL25Δ50 Q72A. Budding was tested at 1:1, 1:6 and 1:10 NEC220-

CBM:UL25 molar ratios for UL25Δ44 Q72A and at a 1:10 NEC-CBM:UL25 molar ratio for eGFP-

UL25Δ50 Q72A. Each condition was tested in at least two biological replicates, each consisting of

three technical replicates. Symbols represent average budding efficiency of each biological

replicate relative to NEC220 (100%). Error bars represent the standard error of measurement for

at least two individual experiments. Significance compared to NEC220 was calculated using an

unpaired t-test against NEC220. *p-value<0.1. The source file with all raw data values is provided

in Figure 3—source data 1.

(b) Confocal microscopy image showing eGFP-UL25Δ50 Q72A does not bind to NEC220-CBM coated GUVs as indicated by the lack of green signal on the membranes of intraluminal vesicles (ILVs) formed by NEC220-CBM budding (indicated by white arrows). Top panel shows red (ATTO-594 DOPE), green (eGFP), and blue (Cascade Blue) channels. Bottom panel shows red (ATTO-594 DOPE) and green (eGFP) channels only. Scale bars = 5 μm.
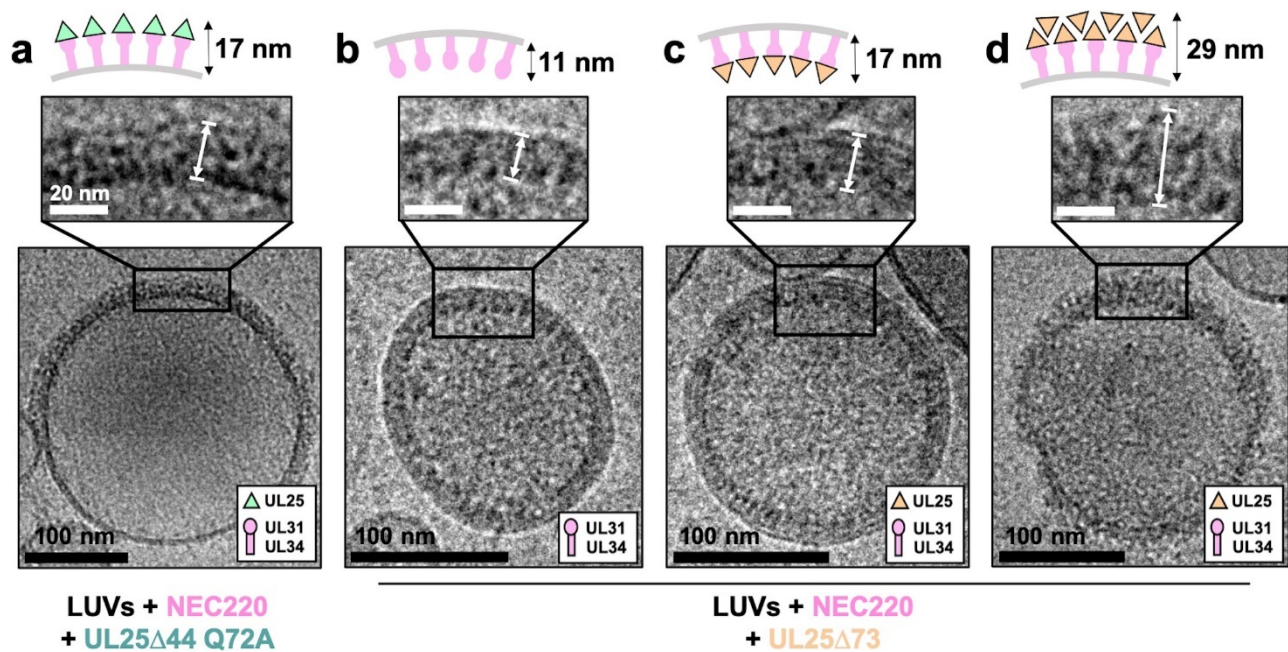
Figure 3- 4. CryoEM shows UL25Δ44 Q72A inhibits NEC220 budding while UL25Δ73 does not.

(a) UL25Δ44 Q72A bound to the NEC220 on the outside of the unbudded lipid vesicles, forming a fence-like array (~17 nm). In the presence of UL25Δ73, three scenarios have been observed: (b) NEC220 alone bound to the inner surface of the budded lipid vesicles (~11 nm); (c) UL25Δ73 bound to the NEC220, which is itself bound to the inner surface of the budded lipid vesicles (~17 nm), and (d) UL25Δ73 aggregates bound to the NEC on the outside of the unbudded lipid vesicles (>29 nm). Budded lipid vesicles in panels b and c are no longer contained within a 'mother' lipid vesicle and represent the end-product of budding. Scale bars = 100 nm. Inset scale bars = 20 nm. All inset panels are shown on the same scale. White arrows in insets define measurement

boundaries of vesicle-bound proteins displayed in the corresponding cartoon models.
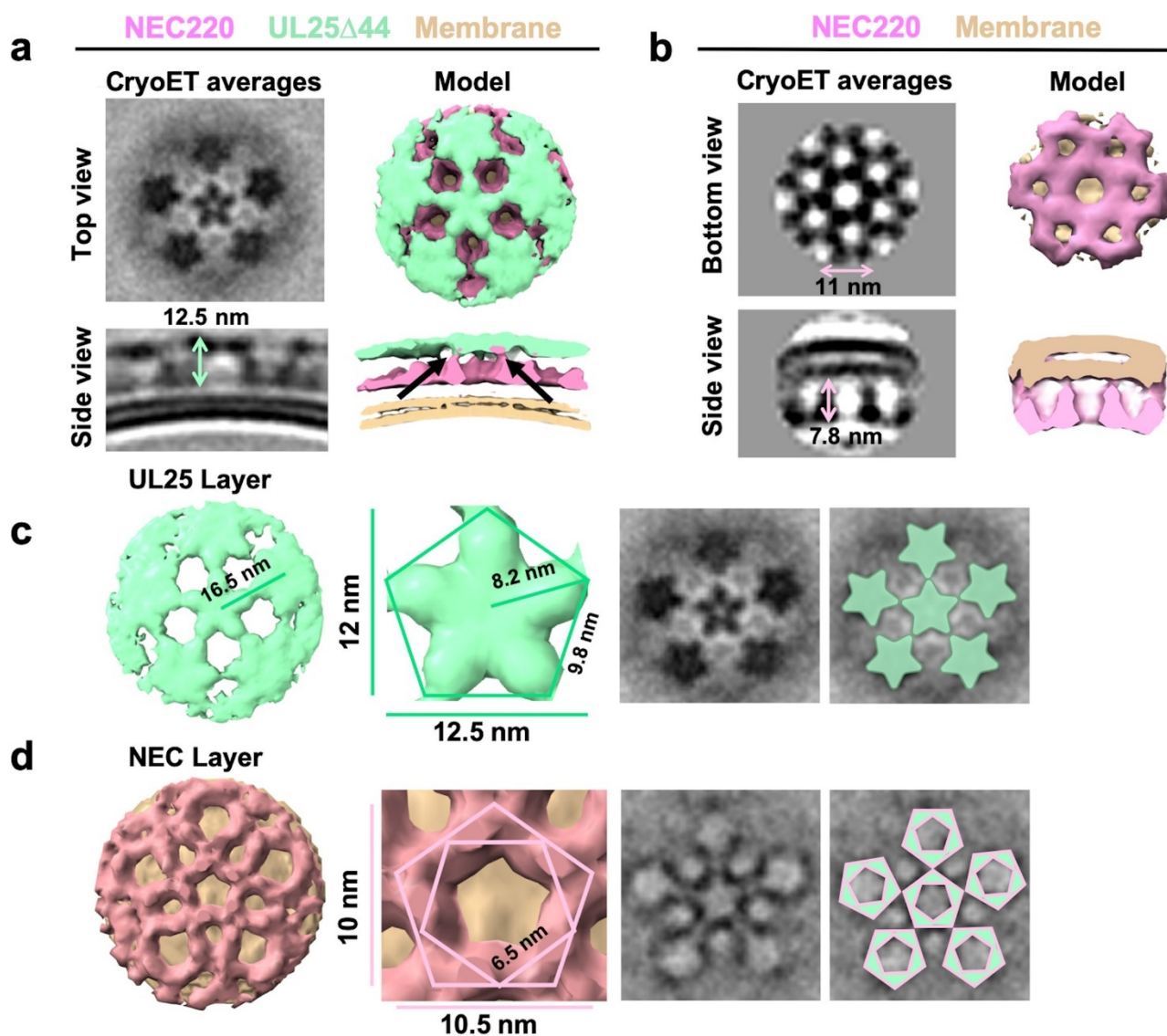
Figure 3- 5. CryoET of UL25-mediated inhibition of NEC budding..

(a) CryoET averages of NEC220 in the presence of UL25Δ44 Q72A (top and side views).

Corresponding 3D models are shown with NEC220 (pink) and UL25Δ44 Q72A (green). The

vesicle bilayer is shown in beige. The models show the UL25 layer coating the NEC layer in five-

pointed stars on the outside of the vesicles. The length of the NEC-UL25 spikes is 12.5 nm. Black

arrows indicate the point of tilt within the NEC layer. (b) CryoET averages of NEC220 forming

hexameric lattices in the presence of membranes (bottom and side views). Corresponding 3D

models are shown with NEC (pink) and the vesicle bilayer (beige). The diameter of the hexameric

rings is ~11 nm, while the length of the spikes is 7.8 nm. (c) CryoET model and averages of the UL25 layer (green) highlighting the five-pointed star formation of UL25 (represented here as a pentamer of dimers) in the presence of NEC. (d) CryoET model and averages of the NEC layer showing NEC220 forming a pentagonal lattice (pink pentagons), rather than hexagonal (as seen for wild-type in panel b). Green triangles indicate location of UL25 binding to the NEC.
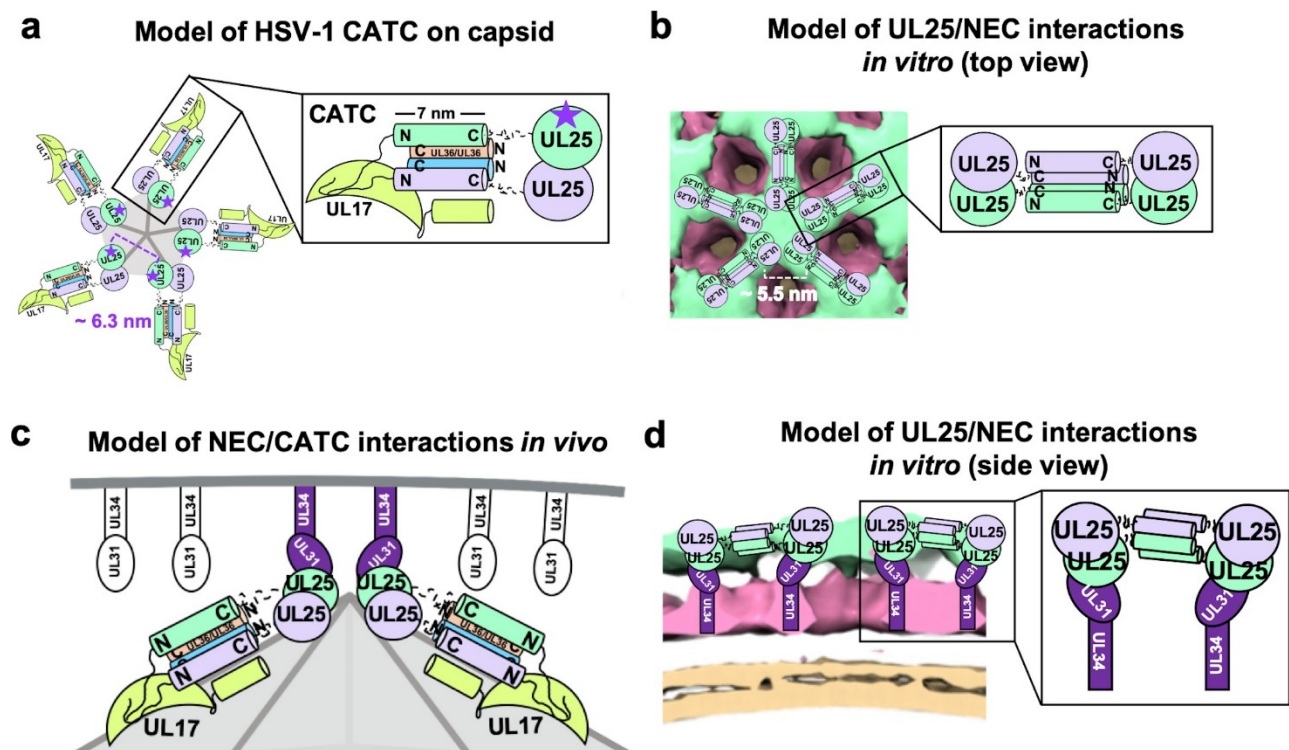
Figure 3- 6. Models of UL25/UL25 and UL25/NEC interactions in vitro and in vivo.

(a) A schematic representation of the pentagonal HSV-1 CATC [two copies of UL25 (green and purple), two copies of C-terminal UL36 (peach and blue) and one copy of UL17 (lime green)] arrangement at the capsid vertex. Inset shows a close-up view of the characteristic antiparallel four-helix bundle composed of two UL25 helices and two UL36 helices. Purple stars indicate the proposed UL25 copies that bind to the NEC upon capsid docking. The distance between the centers of two adjacent inner UL25 cores (green) in the capsid (Dai and Zhou, 2018) is ~6.3 nm. (b) Proposed model of the UL25 stars formed in vitro. The distance between the centers of two adjacent UL25 dimers is ~5.5 nm. Inset shows a close-up view of the proposed antiparallel four-

helix bundle composed of two pairs of UL25 helices from adjacent stars. We hypothesize that four-

helix bundles link the neighboring UL25 stars into a net. (c) Proposed side-view model of the NEC

(purple) interacting with the most surface exposed capsid-bound UL25 (green), resulting in a

pentameric NEC (indicated by dark purple coloring). NEC molecules prior to capsid binding are

shown in an unknown oligomeric state (white). (d) Side view of the proposed NEC/UL25

interactions in vitro.
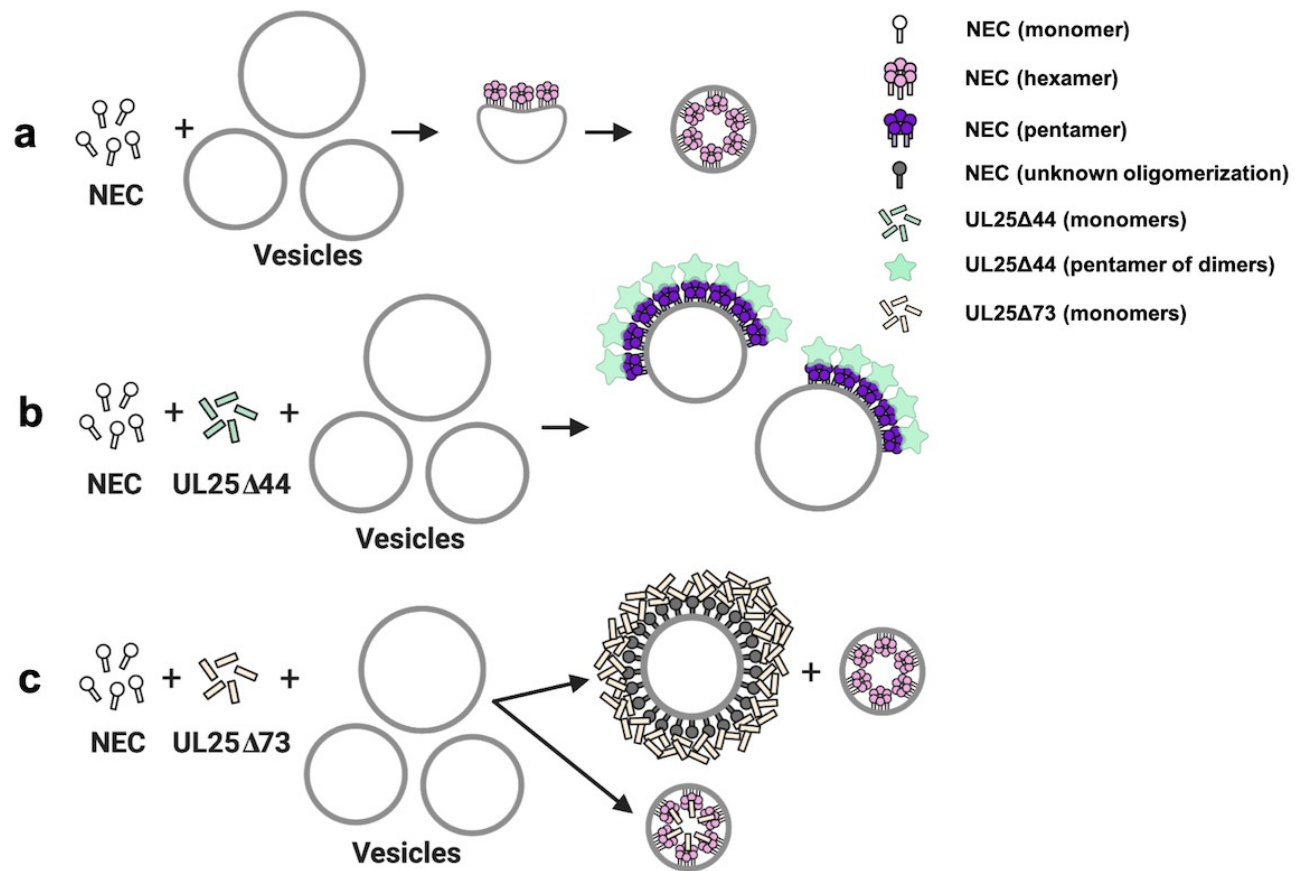
Figure 3- 7. A model of NEC-mediated budding in the absence and presence of UL25, in vitro.

(a) NEC-mediated budding requires only the NEC, which vesiculates membranes by forming

hexagonal coats (pink) that, potentially, contain irregular defects to achieve curvature. (b)

UL25Δ44 Q72A (green) inhibits NEC-mediated budding by inducing the formation of a pentagonal

NEC coat (purple) suboptimal for budding. (c) UL25Δ73 (peach) aggregates around some NEC-

coated vesicles, which blocks budding. Sequestration of UL25Δ73 at a few locations reduces its

concentration elsewhere and enables budding. Binding of UL25Δ73 to NEC in the absence of

aggregation does not interfere with budding, and bound UL25Δ73 buds into vesicles with the NEC.

This figure was created with Biorender.com.
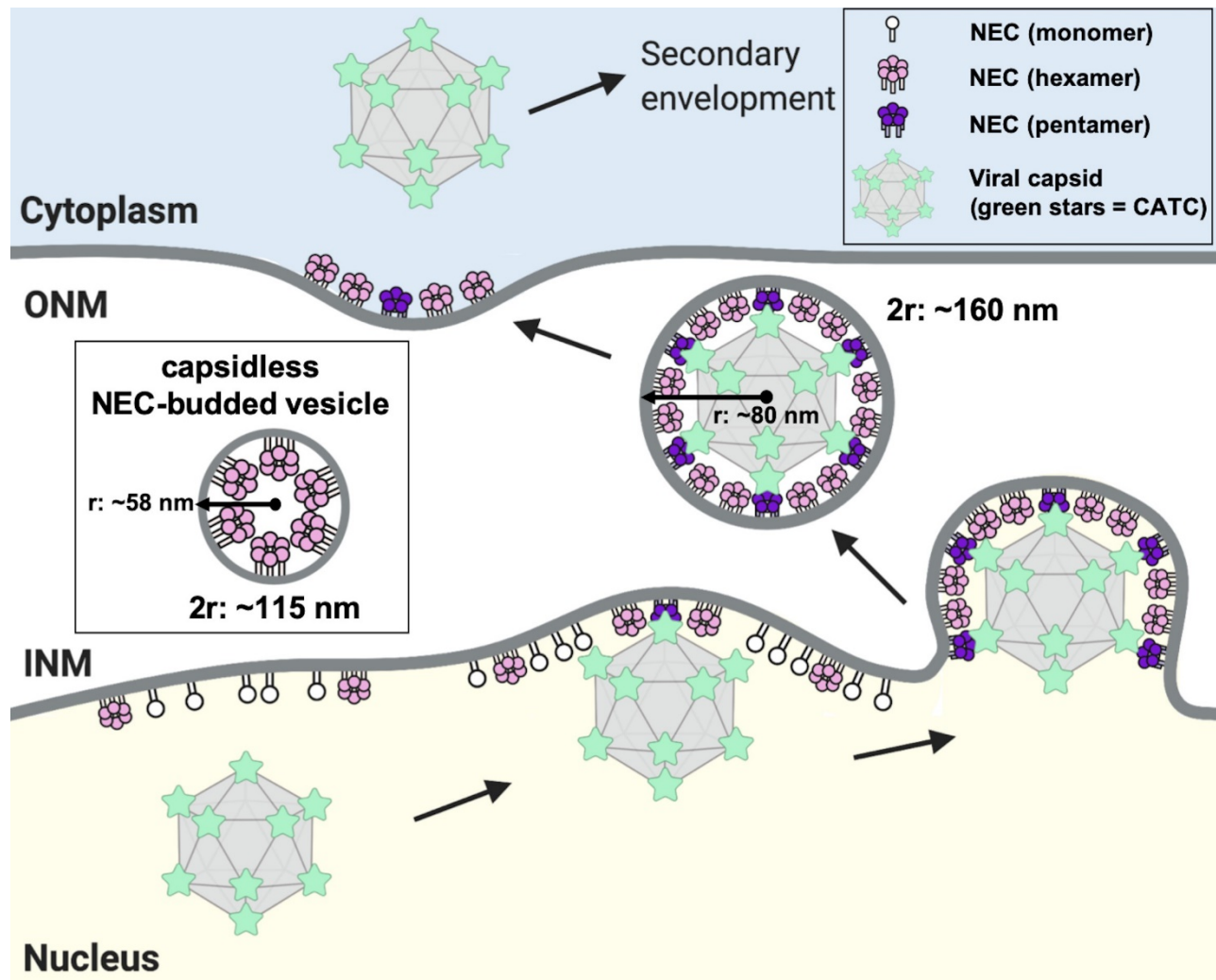
Figure 3- 8. A model of NEC-mediated budding in HSV-1 infected cells.

Capsid-bound UL25 induces the formation of pentagonal insertions (purple pentamers) within the

NEC coat (pink hexamers and white monomers) as it is forming, which enables the formation of an

NEC coat of appropriate size and curvature around the capsid. Inset shows a capsidless

perinuclear vesicle formed in NEC-expressing uninfected cells that forms a hexagonal coat with

presumably irregular defects, similar to the NEC coat formed in vitro. This figure was created with

3.12 Supplemental Figures

a UL25Δ44 to NEC220-His

2500 nM
1250 nM
625 nM
313 nM
156 nM
78 nM
39 nM

b UL25Δ73 to NEC220-His

2500 nM
1250 nM
625 nM
313 nM
156 nM
78 nM
39 nM

c UL25Δ133 to NEC220-His

2500 nM
1250 nM
625 nM
313 nM
156 nM
78 nM
39 nM

d

NEC220/UL25Δ44

e

void 670   158   44   17   1.35   (kDa)

UL25Δ44
NEC185Δ50
UL25 + NEC
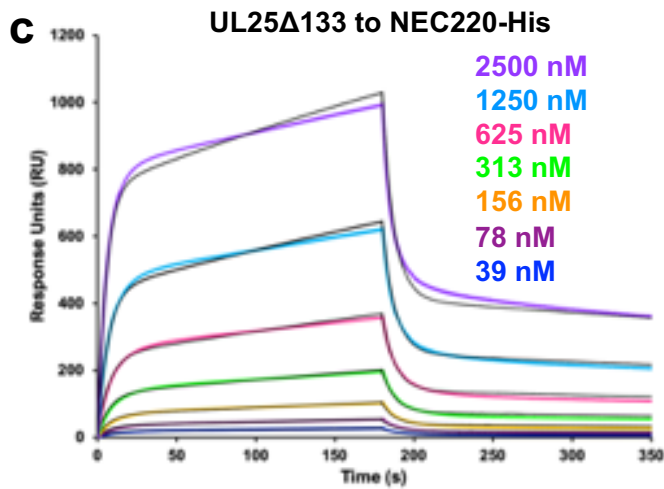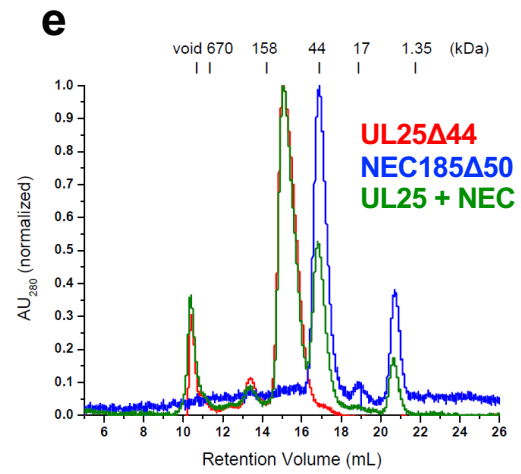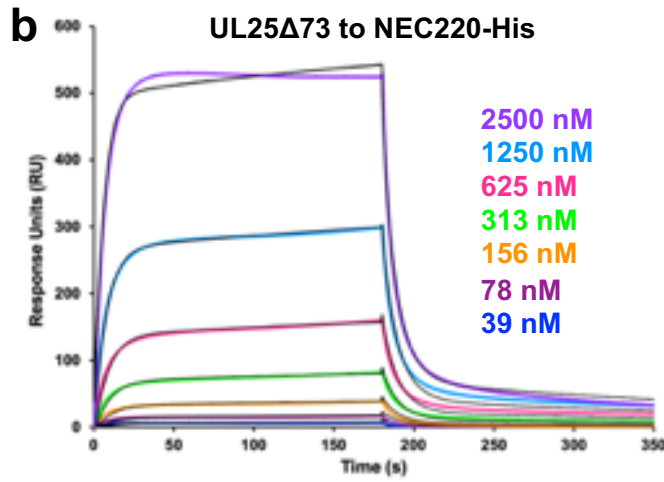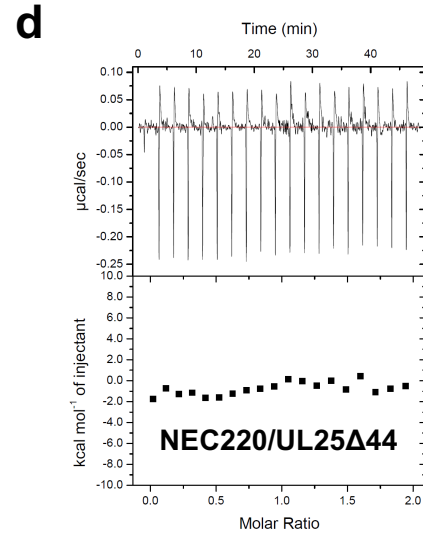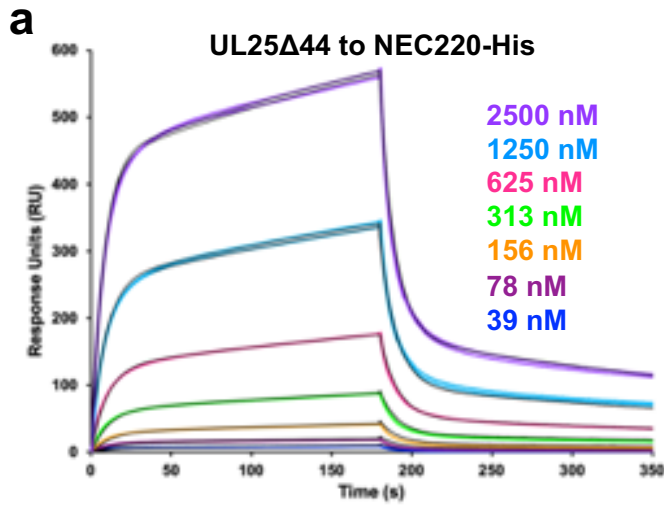
Supplemental Figure 3- 1. NEC-UL25 binding studies.

a) SPR binding of UL25Δ44 Q72A, b) UL25Δ73, and c) UL25Δ133 to NEC220-His, fit to a two-state model (black lines), indicating UL25 can bind the NEC if the NEC is able to form a type of scaffold. d) ITC of NEC220 and UL25Δ44 showing these two proteins do not bind in solution. e) Size-exclusion chromatography of NEC185Δ50 (crystallization construct) and UL25Δ44 shows these proteins also do not bind in solution.
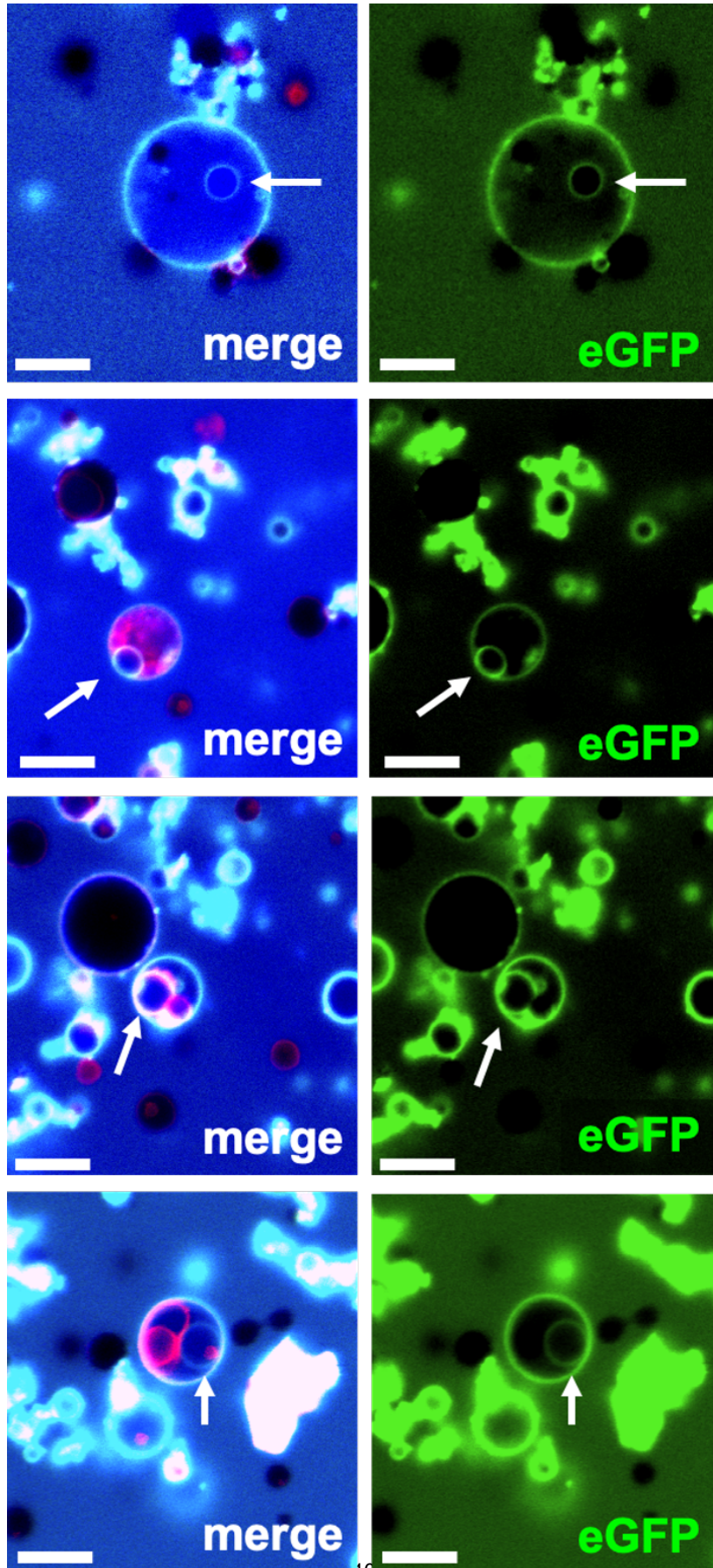
Supplemental Figure 3- 2. NEC-coated vesicle binding and internalization of eGFP-UL25Δ73 (1:5 molar ratio of NEC:UL25 to reduce aggregation).

Green intraluminal vesicles (ILVs) filled with Cascade Blue are observed (white arrows) indicating eGFP-UL25Δ73 budded into these ILVs. All three channels (ATTO-DOPE594/red, eGFP/green, and Cascade Blue/blue) are shown (merge). The eGFP channel alone is also shown (eGFP). Scale bar represents 10 μm.
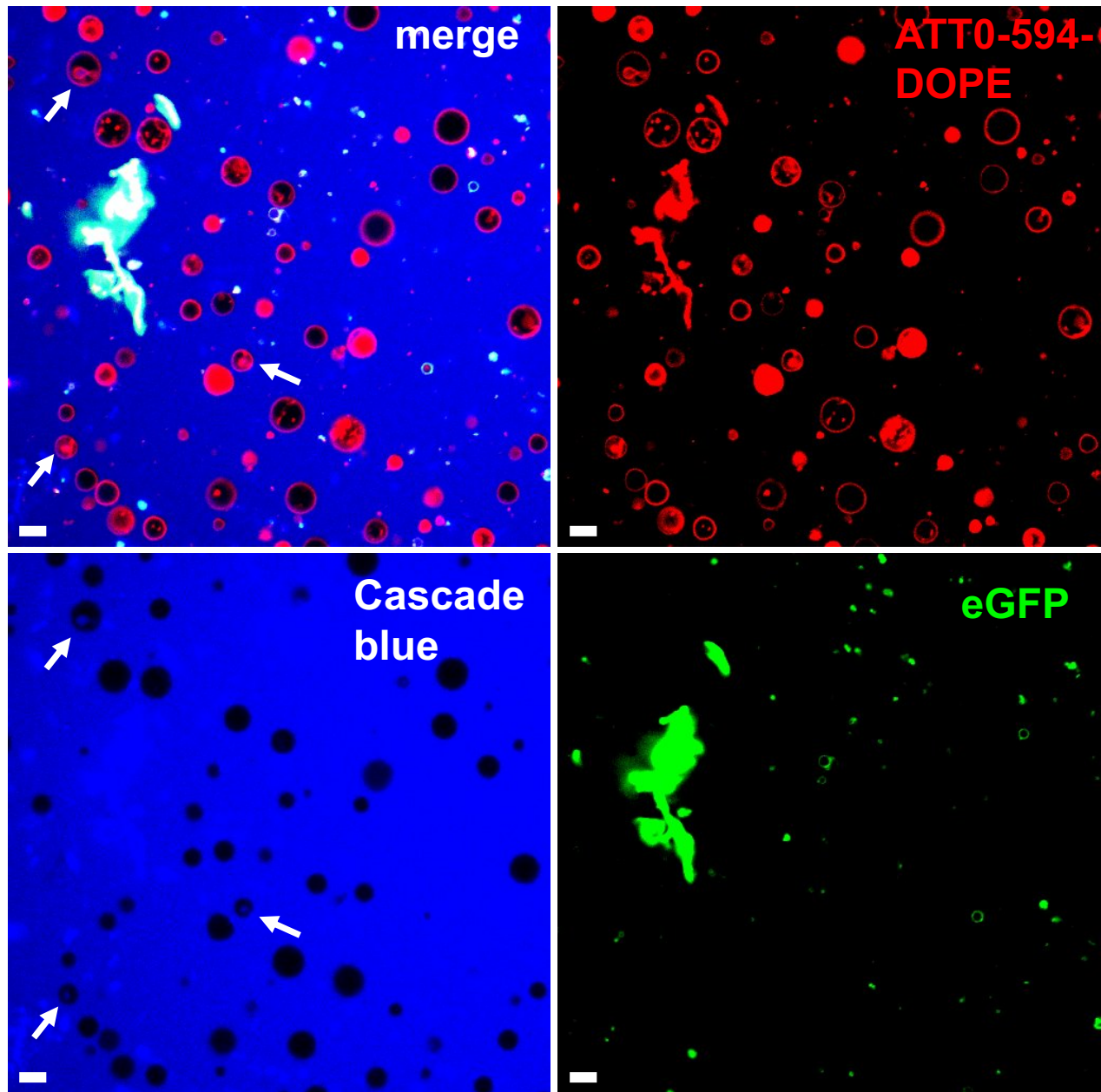
Supplemental Figure 3- 3. Lack of binding between NEC220-coated GUVs and eGFP-UL25△133 (1:10 molar ratio of NEC:UL25).

Intraluminal vesicles are observed, but none were green, indicating there was no eGFP-UL25△133 bound. Scale bar represents 10 μm.

3.13 Tables

Vesicle	Vesicle Radius (nm)	Vesicle Area[a] (nm²)	Particles forming 5-pointed stars[b,c]	Particles forming 5-pointed stars (picked for analysis)	Area of particles forming 5-pointed stars[d] (nm²)	Ratio of particles forming 5-pointed stars/vesicle (%)
1	105	34,521	233	155	7,626	22
2	263	217,317	479	319	15,695	7
3	227	161,838	458	305	15,006	9
4	317	315,940	450	300	14,760	5
5	318	317,977	450	300	14,760	5
6	202	128,650	462	308	15,154	12
7	399	500,418	459	306	15,055	3
8	187	109,794	195	130	6,396	6
9	249	195,456	453	302	14,858	8
10	267	223,775	453	302	14,858	7
Average	**253**	**220,569**	**409**	**273**	**13,417**	**8**

Table 3- 1. UL25△44 Q72A/NEC particles used for cryoET averaging.

The estimated vesicle surface coverage (%) by particles picked for cryoET data processing was calculated using certain assumptions, as indicated within the table.

[a] We assume all vesicles are spherical; Area = π * r * r; r = radius

[b] 90% of particles picked from tomograms are forming 5-pointed star shaped density (based on the cross-correlation coefficient (CCC) value)

[c] 60% of particles forming 5-pointed star shaped density were picked from tomograms

We assume the area of each 5-pointed star $= \frac{1}{4}\sqrt{5(5+2\sqrt{5})} * 9.8^2 = 164\ nm^2$ (side of UL25

pentagon = 9.8 nm; see Figure 5).

Primer Name	Primer Sequence (5'-3')	Restriction Site
UL25Δ44 fwd (JB133)	aaaaaaggatcccgtgaaaccgcagcagaacagg	BamHI
UL25Δ50 fwd (ED034A)	aaaaaaggatcccaggttgttgttctgcaggcacagcgt	BamHI
UL25Δ73 fwd (JB238)	aaaaaaggatccgccgaactgccggttgatattg	BamHI
UL25 rev (JB134)	aaaaaactcgagttattacactgcgctcagatactgagg	XhoI
UL25Δ44 Q72A (ED006)	aatgcagcaatg**gcg**gcagcc	Site-directed
UL25Δ44 Q72A (ED007)	ggctgc**cgc**cattgctgcatt	Site-directed
eGFP fwd (ED010)	aaaaaaggatccgtgagcaagggcgaggagctg	BamHI
eGFP rev (ED049)	aaaaaaggatcccttgtacagctcgtccatgccgagagtg	BamHI
UL31 CBM fwd (JB208)	ccgtgtcggccgca**gcc**atttat**gctgca**atgagg**gcc**atcagcttcgacggg	SOE
UL31 CBM rev (JB209)	cccgtcgaagctgat**ggc**cctcat**tgcagc**ataaat**ggc**tgcggccgacacgg	SOE

Table 3- 2. List of primers used for cloning procedures described in Materials and Methods.

All primers are listed in the 5'-3' direction. Restriction sites are underlined and mutations are bolded.

Chapter 4: Cryo electron tomography with Volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*

Simon Imhof[1, *], Jiayan Zhang[1, 2, 3, *], Hui Wang[1,3,4], Khanh Huy Bui[5], Hoangkim Nguyen[1, 6], Ivo Atanosov[3], Wong H. Hui[3], Shun Kai Yang[5], Z. Hong Zhou[1, 2, 3,@] and Kent L. Hill[1, 2, 3, @]

[*] co-first authors contribute equally to this project.

[@] co-corresponding authors

1 Department of Microbiology, Immunology and Molecular Genetics, UCLA, Los Angeles, California, USA

2 Molecular Biology Institute, UCLA, Los Angeles, California, USA

3 California NanoSystems Institute, UCLA

4 Department of Bioengineering, UCLA

5 Department of Anatomy and Cell Biology, McGill University, Montreal, CA

6 Present affiliation - Teva Pharmaceuticals, Redwood City, California, USA

Address for Correspondence

Kent L. Hill (kenthill@microbio.ucla.edu) or Z. Hong Zhou (Hong.Zhou@UCLA.edu)

UCLA, Dept. of MIMG

609 Charles Young Dr. East

Los Angeles, CA 90095

4.1 Abstract

The 96-nm repeating unit of an axoneme includes dynein motors and accessory structures that are the foundation for motility of eukaryotic flagella and cilia. However, high-resolution 3D structures of the axoneme are unavailable for organisms in the Excavata phylogenetic lineage, which includes pathogens of medical and economic importance. Here we report cryo electron tomography structure of the 96-nm axonemal repeat from *Trypanosoma brucei*, a protozoan parasite in the Excavata lineage that causes African trypanosomiasis. We examined axonemes from bloodstream and procyclic life cycle stages, and an RNAi knockdown lacking the DRC11/CMF22 subunit of the nexin dynein regulatory complex (NDRC). Sub-tomogram averaging yields a resolution of 21.8 Å for the entire 96-nm repeat. We discovered several lineage-specific structures, including novel inter-doublet linkages and microtubule inner proteins (MIPs). We establish that DRC11/CMF22 is required for the NDRC proximal lobe that binds the adjacent doublet microtubule. We propose lineage-specific elaboration of axoneme structure in *T. brucei* reflects adaptations to support unique motility needs in diverse host environments.

4.2 Introduction

Flagella (also called cilia) are hair-like structures that protrude from the surface of eukaryotic cells and perform motility and signaling functions (1). These activities are essential for health, development and reproduction in humans and other multicellular organisms and to power movement of protists, including microbial pathogens that afflict nearly one billion people worldwide and present an economic burden as agricultural pests (2-5).

The structural basis for the flagellum is the axoneme, and in motile flagella the axoneme typically has a "9+2" arrangement, consisting of 9 doublet microtubules (DMTs) arrayed symmetrically around a pair of singlet microtubules, with radial spokes (RS) extending inward from each DMT (6). Axoneme beating is driven by dynein motors and associated structures arranged in a repeating unit of 96-nm periodicity along each DMT. This 96-nm axonemal repeat is thus the foundational unit of motility for eukaryotic flagella. Canonical features of the 96-nm repeat are outer arm dyneins (OAD) (four per repeat, each having two or three motor domains, depending on species), inner arm dyneins (IAD) (one having two motor domains and six having a single motor domain), the IC/LC complex of inner arm dynein f, the nexin dynein regulatory complex (NDRC) inter-doublet linkage, and two or three RS (7). Within each 96-nm repeat, dynein motors are permanently affixed to the A-tubule of one DMT and use ATP-dependent binding, translocation and release of the B-tubule on the adjacent DMT to drive microtubule sliding (8). DMT attachment to the basal body at one end, together with ATP-independent connections, called nexin links, between adjacent DMTs, limits sliding and therefore causes DMTs to bend in response to dynein activity (9-11). Precise, spatiotemporal coordination of dynein activity on different DMTs enables the bend to be propagated along the length of the axoneme, giving rise to axonemal beating (9, 12). RS, together with the NDRC, are thought to provide a means for transmitting mechanochemical signals across the axoneme as part of a complex and as yet incompletely understood system for regulating dynein activity (7, 10, 13, 14).

Recent advances in cryo electron tomography (cryoET) have made high-resolution, 3D structural analyses of the 96-nm repeat possible, providing insights into mechanisms of axoneme assembly and motility (12, 15-18). However, such analyses have been limited to a restricted number of cell types and phylogenetic lineages. In particular, there has been no such analysis of the 96-nm repeat in any member of the Excavata supergroup (Fig. 1), which includes several human and agricultural pathogens of importance to global public health. Consequently, we lack understanding of the full range of structural foundations for axoneme assembly and motility, and what structural variations underlie lineage-specific beating patterns observed in different organisms. For pathogens, such variations present potential therapeutic targets.

African trypanosomes, *Trypanosoma brucei (T. brucei)* and related species, are parasitic protists in the Euglenozoa branch of the Excavata supergroup (Fig.1) (19). They are medically and economically important pathogens of humans and other mammals (2). Critical to *T. brucei* infection of a mammalian host (20) and to their transmission via a tsetse fly vector (21), is motility of these parasites within and through host tissues. Motility of trypanosomes is driven by a single flagellum that is laterally connected to the cell body along most of its length (Fig. 2A) (2, 22). The *T. brucei* flagellum consists of a 9+2 axoneme and a lineage-specific extra-axonemal structure, termed the paraflagellar rod (PFR), which runs alongside the axoneme for most of its length (2, 23-25). While the PFR exerts influence on the axoneme (24, 26), motility itself is driven by axoneme beating, which is transmitted directly to the cell, deforming the cell membrane and underlying cytoskeleton as the waveform propagates along the axoneme (27). Unlike most organisms, trypanosome axoneme beating propagates from the distal tip to proximal end in a helical wave, creating torsional strain and causing the cell to rotate on its long axis as it translocates with the flagellum tip leading (22, 28-30) (Videos 1 and 2). In essence, the entire cell rotates like an auger as it moves forward. This distinctive form of locomotion provides advantages for moving in viscous environments (31, 32) such as within human and fly tissues, and gives the genus its name, as *Trypanosoma* combines the Greek words for auger (trypanon) and body (soma) (33).

The combination of unusual locomotion mechanism, unique connections to other structures, and adaptation to diverse environmental conditions, suggests that the 96-nm repeating unit of the trypanosome axoneme might harbor lineage-specific elaborations. To investigate this possibility, we employed cryoET and sub-tomogram averaging to determine the 3D structure of the *T. brucei* 96-nm axonemal repeat. We report the 96-nm axonemal repeat structure for wild type parasites in bloodstream (BSF) and procyclic (PCF) stages, and for an RNAi knockdown targeting the CMF22/DRC11 subunit of the NDRC. Our results reveal lineage-specific adaptations, including novel inter-doublet linkages and microtubule inner proteins (MIPs). We also identify an NDRC subunit involved in inter-doublet connections between adjacent DMTs. We propose that lineage-specific adaptations to the 96-nm repeat may support the unique motility needs of these pathogens.

4.3 Results

4.3.1 3D Structure of the trypanosome 96-nm axonemal repeat

A critical element of defining any structure is to ensure the sample is pristine. Our analyses

demonstrated that flagellar skeletons purified from bloodstream form (BSF) trypanosomes are

intact, including intact PFR, basal body and distal tip (Fig. 2B-F). Next it is critical that freezing does

not distort the sample. A single zero-degree tilt image of a flagellum embedded in ice demonstrated

that the axoneme, PFR and axoneme-PFR connectors remain intact following plunge freezing (Fig.

2G). Having established high quality of vitrified samples, tilt series were collected from the center

part of full-length flagella, spanning the middle third between the basal body and tip (Fig. 2B).

Major axonemal and PFR structures were resolved in slices through a single tomogram (Fig. 2H, I,

Video 3), indicating the 3D structure is well-preserved and relatively uncompressed (Suppl. Fig.

S1).

Sub-volumes, *i.e.* particles, encompassing the 96-nm repeat of DMTs were extracted from 10

tomograms and averaged as described in Materials and Methods. In total, 763 particles were

averaged to determine the 3D structure of the axonemal repeat (Fig. 3A-D, Video 4). The average

resolution of the entire structure is 21.8 Å based on the 0.143 Fourier shell correlation criterion

(Suppl. Fig. S2A). The resolutions at different regions vary based on visual inspection, and

assessments by both local Fourier shell correlation (FSC) and *ResMap* ([34](#)) calculations (Suppl.

Fig. S2A and C-F); the resolution of DMT region with MIPs reached 19.0 Å based on local FSC

calculation (Suppl. Fig. S2A).

The 3D structure of the 96-nm repeat clearly resolves the expected major substructures,

including OAD, IAD, RS, the IC/LC complex of IAD-f and the NDRC (Fig. 3B-E). Individual

protofilaments are well-resolved and even alpha and beta tubulin monomers within protofilaments

are clearly resolved (Fig. 3F). Several MIPs are also observed (Fig. 3B). At this resolution, we

observe a filamentous structure on the outside of the DMT that spans the entire 96-nm repeat (Fig.

3E-G, red and white arrows). The location and extended conformation of this structure leads us to propose it to be the FAP59/172 molecular ruler described in *Chlamydomonas* that defines the 96-nm repeat ([17](#)). Supporting this idea, the structure makes direct contact with RS, whose position depends on the FAP59/172 ruler ([17](#)). The position of this ruler was previously determined in *Chlamydomonas* through mass-tagging, but the structure itself was not resolved ([17](#)). We also observe a novel globular structure outside the B-tubule, between protofilaments B7 and B8, having a periodicity of 8 nm (Suppl. Fig. S3A, B blue arrow). The function of this structure is unknown, but it might influence dynein binding, because the microtubule binding domain of OADα contacts the B-tubule at this position (see Fig. 4E red arrow), and its 8-nm periodicity is in the range of estimated step size for dynein and kinesin motors ([35-37](#)).

Two holes are present in the inner junction between the A- and B-tubules (red arrows in Fig. 3C). We termed these "proximal" and "distal" holes, based on their position relative to the proximal end of the axoneme. The distal hole is near the site of NDRC attachment to the DMT and corresponds to the hole reported in other organisms ([38](#), [39](#)). The distal hole in *Chlamydomonas* is dependent on the presence of the NDRC on the external face of the DMT ([40](#)). The proximal hole is specific to *T. brucei*. Unlike the distal hole, there are no structures on the external face of the DMT at the site where the proximal hole is located. This indicates the proximal hole reflects structural properties imparted by proteins of the inner junction or inside the microtubules and is not dependent on the presence of external structures.

Interconnections are observed between substructures on the A-tubule, including between individual OADs (Fig. 3D and 4), between OAD and the IAD-f complex (Fig. 3B, D, Suppl. Fig. S3A, D). Particularly noteworthy are extensive contacts between RS3, IAD-d, and the A and B-tubules (Fig. 3C, Suppl. Fig. S3C, D). At the base of RS3 is a structure that extends over four A-tubule protofilaments and attaches to the inner junction. Unlike the case for *Chlamydomonas* ([15](#)), the NDRC does not make direct contact with the OAD in *T. brucei* (Suppl. Fig. S3D), suggesting differences in mechanisms for coordinating inner and outer dynein motor activities.

4.3.2 Axonemal dynein arrangement in T. brucei

An earlier cryoET study of the *T. brucei* axoneme revealed the expected 4 OADs/repeat but did not resolve individual dynein motors ([23](#)). With sub-tomogram averaging, the beta and alpha OAD motors are now clearly resolved (Fig. 3B, D, 4A). This result provides the first direct demonstration that OADs contain two motor domains in *T. brucei*, making it the first protist shown to have two motors per OAD and correcting a misconception that all protists contain three motors ([12](#)). Together with three radial spokes per repeat, the entire arrangement of the *T. brucei* axoneme therefore resembles that of humans more so than does *Chlamydomonas* or *Tetrahymena*, which are used as models for human cilium structure and function (Fig. 5) ([39](#), [41](#), [42](#)).

Axoneme motility is driven by rotation of the dynein AAA+ ring relative to the linker and tail domains, causing translocation of adjacent DMTs as the dynein transitions from pre-powerstroke to post-powerstroke position ([12](#), [35](#), [43](#)). The AAA+ ring, linker and tail domains are resolved in the OAD-beta dynein and are in the post-power stroke position (Fig. 4B, C), consistent with the fact that samples were prepared without exogenous ATP. This result thus supports structural assignments in the averaged structure. The dynein stalk domain, which contacts the adjacent DMT is visible (Fig. 4E).

Six IADs are well-resolved (Fig. 3C, D) and annotated f, a, b, e, g, and d, according to standard nomenclature ([44](#)). Notably, IAD-c, which is important for movement of *Chlamydomonas* in high viscosity ([45](#)), is absent from the trypanosome structure. This finding is notable, given the very viscous environments experienced by trypanosomes during movement through tissues of the mammalian host ([22](#), [32](#), [46](#), [47](#)) and tsetse fly vector ([48](#)).

4.3.3 Extensive Inter-doublet connections in the T. brucei axoneme

Nexin links are connections between adjacent DMTs, that are visible in axoneme TEM thin sections. They stabilize the axoneme and are a fundamental component of the sliding filament model for axoneme motility ([9](#), [10](#), [14](#)). Prior studies indicate the NDRC is the only nexin link in

Chlamydomonas (Fig. 5A) ([49](#)). In *T. brucei*, however, two prominent inter-doublet connections are observed, the NDRC and the IC/LC complex of IAD-f (Fig. 3B-D, Fig. 7). We term this second connection the "f-connector". The NDRC and f-connector each extend from the A-tubule of one DMT to contact near protofilament B9 of the adjacent DMT. NDRC contact is through the proximal and distal lobes defined by ([49](#)) and extends approximately 31 nm. The f-connector contact region extends approximately 11 nm. A structure analogous to the f-connector is observed between neighboring DMTs of three specific DMT pairs in *Chlamydomonas* ([16](#)). However, the prominence of the f-connector in *T. brucei* suggests it is present between neighboring DMTs of most and perhaps all DMTs, a conclusion supported by analysis of individual DMTs (see below), indicating that nexin links in *T. brucei* include both the NDRC and the f-connector, as well as the OAD inter-doublet connector described below. This distinguishes the *T. brucei* axoneme from 3D axoneme structures from other organisms so far reported (Fig. 5) ([39](#), [41](#), [42](#)).

A conspicuous structure not previously reported in any organism is a large protrusion at the junction between the tail and stalk domains of OAD-alpha (Fig. 4D, F). This protrusion, which we termed the "OAD inter-doublet connector", extends to the space between protofilament B6 and B7 of the adjacent DMT. The OAD inter-doublet connector is thus distinguished from the OAD-alpha stalk, which extends from the AAA+ ring to the space between protofilament B7 and B8 of the adjacent DMT (Fig. 4E). The OAD inter-doublet connector is present on all four OAD-alpha motors in the 96-nm repeat but is not observed in OAD-beta.

4.3.4 Doublet-specific features of the 96-nm repeat

The 96-nm repeat structure described above represents an average of all nine DMTs and does not reflect heterogeneity that may distinguish individual DMTs, as reported for *Chlamydomonas* ([44](#)). To address this, we did sub-tomogram averaging on each DMT separately. The PFR restricts axoneme orientations on the EM grid and consequently, individual DMT structures suffer from the missing wedge. This was most severe for DMT 3 and 7 and we therefore cannot comment on

these DMTs. For the remaining seven DMTs, distortion due to the missing-wedge problem obscured some details, particularly MIPs and OADs. However, main features of the 96-nm repeat were resolved (Suppl. Fig. S4). Each DMT was distinct, but careful examination revealed some similarities, particularly in the region of IAD-b, between DMTs 1+5, 2+6 and 8+9 (Suppl. Fig. S4). Therefore, to reduce the impact of the missing wedge, we averaged DMTs within these pairs together (Fig. 6).

As shown in Figure 6 and Supplementary Figure S4, we identified doublet-specific structures that were not evident in the entire averaged structure. DMT 8 and 9 are distinguished from all other DMTs in that they do not have an IAD-b. In the place of IAD-b is a previously undescribed arch-like structure that extends upward from between RS1 and RS2, which we termed "arch". DMT 1 and 5 are distinguished by the presence of a novel inter-doublet connecter, which we termed "b-connector", that connects IAD-b to the adjacent DMT and includes a "tail" domain that connects with the "Modifier of Inner Arms" MIA complex (50). DMT 2 and 6 contain a b-connector that lacks the tail domain. DMT 4, 8 and 9 lack the b-connector. Structural variation of the b-connector on different DMTs explains why it was not evident in the entire averaged structure. DMTs 1, 4, 5, 6, 8 & 9 each have an f-connector structure. DMT 2 does not have a clear f-connector, but this may reflect a missing wedge artifact since the density of the NDRC connection is also reduced (Suppl. Fig. S4). The analysis of individual DMTs supports the interpretation that the f-connector is present on most DMTs. Additionally, this analysis identified a new lineage specific inter-doublet connection not present in other organisms, the b-connector.

The PFR is attached to DMT 4, 5, 6 and 7 and we therefore considered whether this attachment alters the 96-nm repeat. As detailed above, two PFR-attached DMTs, DMT 5 and 6, each show similarities to non-attached DMTs, DMT 1 and 2, that are not shared by each other (Suppl. Fig. S4). Therefore, PFR attachment does not seem to correlate with specific structural changes in the 96-nm repeat, at least at the current resolution. PFR-attachment complexes themselves, have a 56-nm periodicity (23, 24) and therefore would not be resolved in our 96-nm repeat structure.

4.3.5 CMF22/DRC11 is part of the NDRC proximal lobe involved in binding the adjacent DMT

The NDRC functions in axoneme stability and motility and these functions are thought to be mediated in part through inter-doublet connections (14, 51-53). The NDRC is composed of at least 11 subunits and some of these have been positioned within the complex (49, 50, 54-61). However, subunits that contact the B-tubule of the adjacent DMT are unknown. We identified CMF22 as a subunit of the *T. brucei* NDRC (55), and the *Chlamydomonas* CMF22 orthologue is DRC11 (57). RNAi knockdown of CMF22/DRC11 abolishes forward motility in *T. brucei*, demonstrating the importance of DRC11 in axoneme motility (Video 5 and Video 6) (55). The position of CMF22/DRC11 in the NDRC is unknown, but biochemical data indicate it may be within the proximal or distal lobe structures that contact the adjacent DMT (55, 57, 62). We therefore used cryoET and sub-tomogram averaging to determine the structural basis of the CMF22/DRC11 RNAi knockdown. We used procyclic culture form (PCF) *T. brucei*, because loss of axonemal components is lethal in bloodstream forms (53, 63, 64).

The 96-nm repeat of WT PCF (Fig. 7A) axonemes was very similar to that of BSF (Fig. 3, 4), including the presence of the novel OAD inter-doublet connector and the f-connector, as well as the missing IAD-c. In the CMF22 knockdown, the only structure clearly affected is the NDRC (Fig. 7C-E). The entire structure of the complex is mostly preserved, but the proximal lobe of the linker region is severely reduced (Fig. 7E). The affected structures encompass a large portion of the inter-doublet contact area for the *T. brucei* NDRC and include both regions reported to contact the adjacent DMT in the *Chlamydomonas* NDRC (49). The remaining NDRC domains, including dynein contacts were not grossly affected, although connection from NDRC to the MIA complex (50) might be altered. Therefore, inter-doublet connection mediated by the NDRC is critical for axoneme motility.

4.3.6 Extensive, lineage-specific MIPs in T. brucei

One major advance resulting from cryoET studies is the discovery that protein structures inside the microtubule, first observed in trypanosomes based on transmission EM studies more than fifty years ago ([65](#), [66](#)), are ubiquitous in axonemal microtubules ([38](#), [67](#)). A striking feature of *T. brucei* axonemal microtubules is the presence of extensive MIP complexes not only in the A-tubule, but also in the B-tubule (Fig. 3B, 8A and Table 1). Figure 8A shows a cross-section view of the averaged 96-nm repeat looking from the proximal end of the axoneme, with MIPs colored and external structures removed for clarity. The B-tubule is on top and the A-tubule is below, with 13 protofilaments of the A-tubule and 10 protofilaments of the B-tubule labeled according to convention (Fig. 8A). The shape, position and periodicity of the structure inside the B-tubule, next to the inner junction between the A- and B-tubules (Fig. 8A, B), indicate that this structure corresponds to MIP3 described in other organisms ([38](#), [67](#)). Notably however, the relationship of other MIPs in *T. brucei* to previously described MIPs is unclear and most TbMIPs in both the A- and B-tubules appear to be trypanosome-specific (Fig. 5).

When viewed in longitudinal section from within the B-tubule, TbMIP3 consists of two lobes, 3a and 3b (Fig. 8B), as reported for *Chlamydomonas* and *Tetrahymena* ([38](#), [67](#)). There are six such TbMIP3 structures in each 96-nm repeat. Subtle structural variations in the sizes of lobe 3b and connections to lobe 3a yield a 48-nm repeating pattern of three adjacent TbMIP3 structures, colored red, gold and orange (Fig. 8B). These TbMIP3 variations coincide with other structural variations within the microtubule, such as presence of inner junction holes (arrows in Fig. 8B), unique contacts to Snake MIP (see Snake MIP description below), and attachment to a structure identified as MIP3c in *Chlamydomonas* ([41](#)) (asterisks in Fig. 8B). Variation in lobe 3b between the two gold TbMIP3 structures could suggest a 96-nm repeat unit, but this variation probably results from interference from the DRC base plate on the outside of the DMT at the site of the distal hole.

Facing TbMIP3, on the opposite side of the B-tubule lumen, are several trypanosome-specific MIPs, MIP B5, B4, B2 and a MIP that extends across the entire lumen, thus corresponding to the ponticulus structure previously observed in classical thin section TEM (Fig. 8C) ([65](#), [66](#), [68](#)).

To our knowledge, the ponticulus was the first structure observed within the microtubule lumen in any organism and is the only structure so far described to extend across the entire microtubule. Our 3D structure shows that the ponticulus is not a single structure, but rather is comprised of 3 discrete MIPs, which we termed Pa, Pb and Pc (Fig. 8C-F). Each ponticulus MIP extends across the entire B-tubule lumen, connecting the A-tubule lattice to a different B-tubule protofilament. Pa, Pb and Pc connect protofilament A12 to protofilaments B3, 5 and 4, respectively and exhibit 48nm periodicity (Fig. 8C-F). The ponticulus is assembled after construction of the axoneme ([68]). Therefore, proteins comprising these structures must be delivered into a fully formed DMT.

The A tubule also contains a diverse cohort of MIPs each with a repeating unit of 48 nm (Table 1, Fig. 8A, Suppl. Fig. S6). Rather than constituting several isolated structures however, TbMIPs form a network of interconnected complexes, similar to, but more extensive than, that reported for *Tetrahymena* ([67]). Two A-tubule MIPs are particularly notable. One, which we termed "ring MIP", is unique among MIPs so far described because it forms a ring structure protruding into the microtubule lumen (Fig. 9B). The ring MIP is attached to the protofilaments A8 and 9 and contacts another MIP complex on the protofilaments A8-12 termed "Ring Associated MIP" (RAM) (Fig. 9B, C). Another MIP, which we termed "snake MIP", presents as a serpentine structure that appears to weave in and out of the A and B-tubules (Fig. 10 and Video 7). The continuity of this density suggests it might be a contiguous structure, extending 48 nm and spanning multiple tubulin subunits, although we cannot rule out the possibility that protofilament subunits contribute to this structure.

4.4 DISCUSSION

The ciliary axoneme is one of the most iconic features of eukaryotic cells and is considered to have been present in the last eukaryotic common ancestor (LECA) ([6]). To date, however, high-resolution structures of the 96-nm axoneme repeat have only been reported for two of the three eukaryotic supergroups. Here we report the 3D ultrastructure of the *T. brucei* 96-nm axonemal

repeat. This is the first such structure reported for any pathogenic organism and first representative from the eukaryotic supergroup Excavata, a basal group that includes many pathogens of global importance to human health and agriculture ([69], [70]). Our studies indicate the diversity of structures comprising the 96-nm repeat is under appreciated, give insight into principles of axoneme structure and function, and identify pathogen-specific features that may support unique motility needs of trypanosomes and thus present therapeutic targets.

The genus *Trypanosoma* was discovered more than 175 years ago and named for its unique cell motility ([33]), which is driven by a single flagellum. The functional unit of the eukaryotic flagellum is the 96-nm axonemal repeat, which encompasses dynein motors and regulatory proteins that direct flagellum beating ([7]). In trypanosomes, the PFR exerts influence on the axoneme ([24], [26], [71]), but motility is powered by the axoneme, which is the focus of the current work. Despite intense study for several decades, axoneme structures that underpin the parasite's unique mechanism of cell propulsion remained hitherto unclear. A main finding from our studies is the discovery of lineage-specific features of the *T. brucei* 96-nm axonemal repeat, including extensive and novel MIP structures and novel inter-doublet connections between adjacent DMTs (Fig. 3-6, 8-11). We hypothesize these parasite-specific structures support unique motility needs of trypanosomes and thereby contribute to the transmission and pathogenic capacity of these organisms. The *T. brucei* axoneme is distinguished by mechanical strain experienced due to lateral attachment to the PFR and cell body, vigorous helical beating, encounter with host tissues and frequent reversals of beat direction ([20], [24], [26], [32]). MIPs have been shown to stabilize the axoneme in other organisms ([41], [72], [73]) and the expanded and MIP network of *T. brucei* may therefore help maintain stability of individual DMTs. Likewise, novel inter-doublet connections are expected to help maintain axoneme integrity under these conditions, analogous to the role of NDRC inter-doublet links in maintaining alignment of DMTs in *Chlamydomonas* ([57]).
The diversity and placement of *T. brucei* MIPs are suggestive of functions beyond stability. It is difficult to imagine for example, how a ring structure like the RingMIP, protruding into the

microtubule lumen, would solely provide stability. MIPs in other organisms have been demonstrated to modulate axoneme beating ([41], [73]). Given the presence of numerous trypanosome-specific MIPs, together with MIP differences reported between other species (Fig. 5), we suggest that lineage-specific MIPs may provide a mechanism for fine-tuning the beating of axonemes between species that otherwise share a basic architecture. Extra connections between DMTs can also influence axoneme beating. It has been suggested that vortical beating of nodal cilia in vertebrates axoneme may involve transmission of regulatory signals from DMT to DMT, circumferentially around the axoneme ([13]). Extensive inter-doublet connections identified in our studies provide a means for direct interaction between DMTs and could thus contribute to helical beating that is a hallmark of *T. brucei* motility. Finally, given the recent demonstration that motility is critical for *T. brucei* virulence ([20]), parasite-specific features of the 96-nm repeat, which is the foundational unit of motility, may present novel therapeutic targets. Future work to identify novel *T. brucei* MIP and connector proteins will allow these ideas to be tested directly.

By defining the structural basis of the motility defect in the CMF22/DRC11 knockdown, we demonstrate a specific requirement for inter-doublet connections in axoneme motility because the defect disrupts inter-doublet connections without affecting dyneins. This contrasts to NDRC mutants analyzed previously in *Chlamydomonas*, which typically exhibit structural defects in connections to dyneins or in dyneins themselves ([49], [62], [74]). An exception is *sup-pf4* ([49]), but this mutant has only subtle effects on motility and beat frequency ([62]), which contrasts to the CMF22/DRC11 knockdown in which propulsive motility is ablated ([55]). Our CMF22/DRC11 knockdown studies therefore provide several important insights. Firstly, they demonstrate that penetrance of RNAi makes knockdown lines suitable for differential cryoET structural analysis in *T. brucei.* Secondly, they demonstrate CMF22/DRC11 is required for NDRC proximal lobe assembly and B-tubule attachment and, together with biochemical data ([55], [57], [62]), indicate that CMF22/DRC11 is part of the proximal lobe. Thirdly, because inter-doublet contacts are specifically affected, without affecting dyneins, the results demonstrate that the NDRC itself and B-tubule

contacts specifically are required for control of axoneme motility. This last point is particularly significant, as dynein-independent connection between adjacent DMTs is considered to be a founding principle of the sliding filament model for axoneme motility ([9](9), [11](11), [14](14)), yet direct tests of this idea have been limited.

The 96-nm spacing of the axoneme is controlled by a molecular ruler ([17](17)), which is visible in the averaged BSF 96-nm repeat structure. The *T. brucei* MIP repeating unit is 48 nm, suggesting existence of a separate ruler inside the DMT to guide MIP placement. Such a ruler would need to extend 48 nm, exhibit structural heterogeneity along its length, and form contacts with other MIPs. The snake MIP satisfies these criteria. Notice, for example, that structural heterogeneities along the snake MIP coincide with unique contacts to each TbMIP3a, b structure within the 48-nm repeat (Fig. 8). The snake MIP appears to extend into both the A- and B-tubules, which would make it possible to establish patterns in both tubules. Extensive interconnections between MIPs (Video 7) might allow a single ruler to guide placement of all MIPs, or there might be more than one ruler, as is suggested for the outside of DMTs in *Chlamydomonas* ([75](75)), where the 24-nm repeat of OADs is dictated by something other than the FAP59/172 ruler ([17](17)). Besides the snake MIP, another structure inside the B-tubule (spine MIP) appears to exhibit properties required of a 48-nm molecular ruler - forming a contiguous structure, spanning 48 nm and having heterogeneities that make unique contacts to adjacent MIPs (Suppl. Fig. S5).

4.5 Materials and Methods

4.5.1 Preparation of demembranated flagellum skeletons for cryoET

BSF single marker (BSSM) and PCF (29-13) ([76](76)) *T. brucei* cells were cultured as described ([77](77), [78](78)). Cells, 2×10^8 for BSF or 4×10^8 for PCF, were washed three times in sterile 1xPBS. Supernatant was aspirated to ensure all of the PBS is removed. To remove the cell membrane and other soluble proteins and release the DNA,160 µl Extraction buffer (20mM HEPES pH: 7.4, 1mM $MgCl_2$, 150mM NaCl, 0.5% NP40 IGEPAL CA-630 detergent, 2x Protease Inhibitors Cocktail-Sigma

EDTA-free) + 1/10 volume 10x DNase buffer + 1/10 volume DNase (TURBO, Life Technologies 2U/μl) was added and incubated at room temperature for 15 minutes. In order to solubilize the subpellicular microtubules, 1mM $CaCl_2$ (2 μl of 100 mM $CaCl_2$) was added and incubated on ice for 30 minutes. Then flagellum skeletons (axoneme with PFR, basal body and FAZ filament) were centrifuged (4,000 rpm, 1500g at 4°C for 10 minutes) and the supernatant was removed. Then flagellum skeletons were purified away from cell body remnants and debris by one further centrifugation step over a 30% sucrose cushion at 300rpm, 800g at 4°C for 5 minutes (Extraction buffer w/o NP-40; 30% w/v sucrose). Flagellum skeletons from 200μl of the upper fraction of the buffer-sucrose interface were collected and washed twice in 200μl Extraction buffer, centrifugation at 4,000 rpm 1500g at 4°C for 10 minutes, then resuspended in 40 μl buffer. Samples were either mixed with gold beads and plunge frozen immediately, as described below, or assessed directly for sample quality. To assess sample quality, BSF samples were negative-stained and analyzed using an FEI T12 transmission electron microscope equipped with a Gatan 2kX2k CCD camera. Samples were intact with uniform length distribution and a mean length of 25.2, +/-3.5 μm (Fig. 2B-F). PCF samples were examined by light microscopy to ensure uniform length distribution.

4.5.2 CryoET sample preparation and tilt-series acquisition

BSF or PCF samples in the amount of 40 μl was mixed with either 5-nm (for BSF) or 10-nm (for PCF) diameter fiducial gold beads in 12:1 ratio. An aliquot of 3 μl of the axoneme-gold beads solution was applied onto Quantifoil (3:1) holey carbon grids (for BSF) or continuous carbon-coated EM grids (for PCF) which were freshly glow-discharged for 30s at −40 mA. Excess of the sample on the grid was blotted away with a filter paper, at a blot force of -4 and blot time of 5s, and vitrified by immediately plunging into liquid nitrogen-cooled liquid ethane with an FEI Mark IV Vitrobot cryo-sample plunger. Axoneme architectural integrity and gold bead concentration were assessed and plunge-freezing conditions optimized by obtaining low-resolution cryoET tilt series in an FEI TF20 transmission electron microscope equipped with an Eagle 2K HS CCD camera. From these tilt

series, cryoET tomograms   were evaluated to ensure structural integrity of the axoneme and PFR. Vitrified cryoET grids were stored in liquid nitrogen until use.

For high-resolution cryoET tilt series acquisition, vitrified specimens were transferred with a cryo-holder into an FEI Titan Krios 300kV transmission electron microscope equipped with a Gatan imaging filter (GIF) and a Gatan K2 Summit direct electron detector. Samples were imaged under low-dose condition using an energy filter slit of 20 eV. CryoET tilt series were recorded with *SerialEM* ([79](#)) by tilting the specimen stage from -60° to +60° with 2° increments. The cumulative electron dosage was limited to 100~110 e$^-$/Å^2 per tilt series. All 4kx4k frames were recorded on a Gatan K2 Summit direct electron detector in counting mode with the dose rate of 8-10 e$^-$/pixel/s. For each tilt angle, a movie consisting of 7 to 8 frames was recorded. For the PCF samples, the nominal magnification was x26,000, giving rise to a calibrated pixel size of 6.102 Å. The defocus value was targeted at -4 μm. When the BSF samples were ready to be imaged, the same instrument was upgraded with a VPP, allowing us to obtain higher contrast images at closer to focus and higher magnification conditions. To obtain tilt series for the BSF samples with VPP, we follow the procedures previously described ([80](#), [81](#)) and used the same GIF and K2 parameters as indicated above. Before starting each tilt series, we moved to a new VPP slot, waited for 2 minutes for stabilization, then pre-conditioned the VPP by illumination with a total electron dose of 12 nC for 60s to achieve a phase shift of ~54°. Tilt series were recorded at a nominal magnification of 53,000X (corresponding to a calibrated pixel size of 2.553 Å) and a targeted defocus value of -0.6μm. For BSF we collected a total of 50 tomograms and selected the 10 best, based on limited axoneme compression for sub-tomogram averaging. Cross sections of these 10 tomograms are shown in Suppl. Fig. S1, and have circularity, measured as ratio of short axis/long axis, ranging from 0.92 to 0.98. This yielded 763 particles that were averaged to determine the 3D structure of the BSF axonemal repeat. For WT PCF we collected 27 tomograms, and 17 of them were used for sub-tomogram averaging, resulting in 1177 particles averaged. For DRC11/CMF22 RNAi samples a total of 24 tomograms were collected and 19 of them were used for sub-tomogram averaging,

resulting in 1726 particles averaged. For sub-tomogram averaging of individual DMT (Fig. 6 and Suppl. Fig. S4), an additional 24 tomograms of BSF axonemes were used, for a total of 34 tomograms, yielding 297 to 339 particles averaged for each DMT (DMT1 = 339, DMT2 = 332, DMT3 = 297, DMT4 = 327, DMT5 = 311, DMT6 = 337, DMT7 = 316, DMT8 = 306, DMT9 = 309).

4.5.3 Data processing

For PCF and BSF samples, frames in each movie of the raw tilt series were drift-corrected, coarsely aligned and averaged with *Motioncorr* ([82](#)), which produced a single image for each tilting angle. The tilt series images were reconstructed into 3D tomograms by weighted back projections using the *IMOD* software package ([83](#)) in six steps. Micrographs in a tilt series were coarsely aligned by cross-correlation (step 1) and then finely aligned by tracking selected gold fiducial beads (step 2). The positions of each bead in all micrographs of the tilt series were fitted into a specimen-movements mathematical model, resulting in a series of predicted positions. The mean residual error was recorded to facilitate bead tracking and poorly-modeled-bead fixing (step 3). With the boundary box reset and the tilt axis readjusted (step 4), images were realigned (step 5). Finally, tomograms were generated by weighted back projection (step 6). Contrast transfer function (CTF) was corrected with the *ctfphaseflip* program ([84](#)) of IMOD in step 5 above. The defocus value for each micrograph was determined by *CTFTILT* ([85](#)), and the estimated defocus value was used as input for *ctfphaseflip*. Note, one of the benefits of using a phase plate is that the CTF is insensitive to the sign of the defocus value being negative (underfocus) or positive (overfocus) ([86](#)).

To improve the signal-to-noise ratio and enhance the resolution, sub-tomograms containing the 96-nm axonemal repeated units along each DMT were extracted/boxed out from the raw tomograms. Sub-tomogram averaging and the missing-wedge compensation were performed using *PEET* program ([15](#), [87](#)) as detailed previously ([81](#)), except for a new script we wrote to pick sub-volumes as outlined in the subsequent paragraphs.

In our sub-tomogram averaging scheme, each particle is defined as the 96-nm repeating unit of the

DMT. We developed a *MATLAB* script, *autoPicker*, to semi-automatically pick particles and calculate their location and orientation based on axoneme geometry. Briefly, we represent the 9+2 axoneme as a cylinder. For each axoneme in a tomogram, we used *IMOD* to visually pinpoint 11 points and save their coordinates into a file. The first two points, p_a and p_b, are the center points of the two bases of the cylinder. The remaining 9 points (p_i, *i=1…9*) identify the centers of the nine DMTs (particles) within the first 96 nm length at one end of the selected axoneme. The center is defined as the intersection point of a DMT with the middle of the three radial spokes along each particle's 96-nm unit length. Our script reads the coordinates of the 11 points, calculates vector $\overrightarrow{p_a p_b}$ that defines the orientation of the cylinder, determine the center coordinates of all other particles within this axoneme based on the following formula:

$p_{ij} = p_i - L \cdot j \cdot \frac{\overrightarrow{p_a p_b}}{|\overrightarrow{p_a p_b}|}$, where *i* = 1, 9; *j* = 1 to $|\overrightarrow{p_a p_b}|/L$, *L* is the unit length (96nm)

In order to uniquely identify the orientation of each particle, *autoPicker* also calculates a second point, p^*_{ij} for each p_{ij}. p^*_{ij} corresponds to the middle radial spoke's end near the central pair. This is accomplished by solving the following linear algebraic equations that both p^*_{ij} and p_{ij} must satisfy (see illustrations in Suppl. Fig. S7):

$$\begin{cases} \overrightarrow{p_a p_b} \cdot \overrightarrow{p_{ij} p^*_{ij}} = 0 \\ (\overrightarrow{p_a p_b} \times \overrightarrow{p_a p_{ij}}) \cdot \overrightarrow{p_{ij} p^*_{ij}} = 0 \\ \left| \overrightarrow{p_{ij} p^*_{ij}} \right| = Length\ of\ the\ radial\ spoke\ (60nm) \end{cases}$$

We ran *autoPicker* for each axoneme in our tomograms to generate a *PEET* mod file that contains a list of the above described p_{ij} and p^*_{ij} pairs for all particles in that axoneme. Program *stalkInit* in *PEET* then read this mod file and generate an initial *motive* list file, a RotAxes file and three model files containing the coordinates for each particle. *PEET* then read the coordinate and orientation information from these files and automatically extracted the particles from the tomograms to perform iterative sub-tomogram averaging until no further improvement can be obtained.

Sub-tomogram averaging of the individual DMTs was performed in two steps. Step1:

particles (96-nm repeat units), picked from all 9 DMTs were classified into 9 classes, corresponding to the DMT from which each particle was picked, DMT 1-9. Step 2: for particles in each of the 9 classes, sub-tomogram averaging was performed using PEET.

The resolutions of the sub-tomogram averages were evaluated by two different approaches, one based on Fourier shell correlation (FSC) calculated by *simpleFSC* in *PEET* ([15](#), [87](#)) and the other by *ResMap* ([34](#)). To calculate FSC curves, we split all particles into two of equal-sized subsets following the PEET tutorial. Specifically, particles are separated into two subsets with the *PEET* specific *motive list* file by designating each sub-volume as either "1" or "2" so that it would be placed into one of the two sub-sets. *PEET* then performed sub-tomogram averages independently for particles in each of the two equal-sized sub-sets, yielding two sub-tomogram averages of the 96-nm axonemal structure. These two independently calculated sub-tomogram averages were then used as the input maps of the *simpleFSC* program in the PEET package to calculate the FSC curve for the entire 96-nm axonemal repeat (Suppl. Fig. S2A). We also calculate FSC curves for local regions encompassing DMT with MIPs, OAD, IAD, NDRC or RS. To do so, a cuboid mask was used in *ChimeraX* ([88](#)) to extract two corresponding local density regions that primarily containing either DMT with MIPs, or OAD, or IAD, or NDRC or RS from the 2 sub-tomogram averages. Each set of two corresponding cuboid volumes (Suppl. Fig. S2B) was then used as the input maps of the *simpleFSC* program in the PEET package to calculate an FSC curve for the local region, which is plotted as a function of spatial frequency (Suppl. Fig. S2A). Local resolution across the entire averaged 96-nm axonemal repeat was also evaluated with *ResMap* ([34](#)) using the above two independently calculated sub-tomogram averages as input maps and the result is visualized from different views in Suppl. Fig. S2C.

4.5.4 3D visualization

IMOD ([83](#)) was used to visualize the reconstructed tilt-series and the 2D tomographic slices of the sub-tomogram averages. UCSF *ChimeraX* ([88](#)) was used to visualize the resulting sub-tomogram

averages in their three dimensions. Segmentation of densities maps and surface rendering for the different components of the 96-nm repeated unit were performed by the tools *volume tracer* and *color zone* in UCSF *Chimera ([89](89)). GIMP 2.8.18 (GNU Image Manipulation Program)* was used to color regions of interest (Fig. 5, 6B-D, 8B-F, 9B-C; Suppl. Fig. S3C-D, S5B, S6B; Table 1). For rendering, no filters were applied on MIPS but we applied low pass filters on the other components to improve the clarity of individual structures described in the text. For the structures in Fig. 3C-E; Fig. 4A, B, D; Fig. 7A-E, we filtered the DMT, NDRC, RS, IC/LC, OAD and IAD to 30Å. For the structures in Fig. 5; Fig. 6; Suppl. Fig. S4, we filtered the entire map to 50Å).

4.5.5 Trypanosome motility videos

Motility videos of BSF cells were obtained exactly as described in ([90](90)). Motility videos of PCF cells were obtained exactly as described in ([55](55)). All videos were recorded and played back at 30 frames per second. The PCF tetracycline-inducible DRC11/CMF22 RNAi knockdown line has been described previously ([55](55)). WT and mutant PCF videos correspond to this knockdown line cultured in the absence (WT) or presence (mutant) of tetracycline to induce RNAi.

4.6 Author contributions:

KH and ZHZ initiated and oversaw the project; SI, HN and JZ prepared samples; JZ, SI, IA and WH prepared and imaged the samples; JZ, WH, HW, SI and SKY processed the data; SI, JZ, HW, KH, ZHZ and HB analyzed and interpreted the data; SI, JZ, KH, ZHZ wrote the paper; all authors reviewed and approved the paper.

4.7 Acknowledgements

4.8 Data availability

The cryoET sub-tomogram average maps have been deposited in the EM Data Bank under the

accession codes EMD-20012, EMD-20013 and EMD-20014, for the wild-type bloodstream form,

wild-type and DRC11-knock-down procyclic form, respectively.

4.9 References:

1.   Smith EF, Rohatgi R. Cilia 2010: the surprise organelle of the decade. Sci Signal.
2011;4(155):mr1.

2.   Langousis G, Hill KL. Motility and more: the flagellum of Trypanosoma brucei.
Nature reviews Microbiology. 2014;12(7):505-18.

3.   Gerdes JM, Davis EE, Katsanis N. The vertebrate primary cilium in development,
homeostasis, and disease. Cell. 2009;137(1):32-45.

4.   Ibanez-Tallon I, Heintz N, Omran H. To beat or not to beat: roles of cilia in
development and disease. Hum Mol Genet. 2003;12 Spec No 1:R27-35.

5.   Anvarian Z, Mykytyn K, Mukhopadhyay S, Pedersen LB, Christensen ST. Cellular
signalling by primary cilia in development, organ function and disease. Nat Rev Nephrol.
2019;15(4):199-219.

6.   Khan S, Scholey JM. Assembly, Functions and Evolution of Archaella, Flagella and
Cilia. Current Biology. 2018;28(6):R278-R92.

7.   Porter ME, Sale WS. The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. J Cell Biol. 2000;151(5):F37-42.

8.   Gibbons IR, Rowe AJ. Dynein: A Protein with Adenosine Triphosphatase Activity from Cilia. Science. 1965;149(3682):424.

9.   Satir P. STUDIES ON CILIA. The Journal of Cell Biology. 1968;39(1):77.

10.   Satir P, Heuser T, Sale WS. A Structural Basis for How Motile Cilia Beat. Bioscience. 2014;64(12):1073-83.

11.   Holwill ME, Satir P. A physical model of microtubule sliding in ciliary axonemes. Biophys J. 1990;58(4):905-17.

12.   Lin J, Nicastro D. Asymmetric distribution and spatial switching of dynein activity generates ciliary motility. Science. 2018;360(6387):eaar1968.

13.   King SM. Turning dyneins off bends cilia. Cytoskeleton. 2018;75(8):372-81.

14.   Viswanadha R, Sale WS, Porter ME. Ciliary Motility: Regulation of Axonemal Dynein Motors. LID - a018325 [pii] LID - 10.1101/cshperspect.a018325 [doi]. 2017(1943-0264 (Electronic)).

15.   Nicastro D, Schwartz C, Pierson J, Gaudette R, Porter ME, McIntosh JR. The molecular architecture of axonemes revealed by cryoelectron tomography. Science. 2006;313(5789):944-8.

16.   Bui KH, Sakakibara H, Movassagh T, Oiwa K, Ishikawa T. Asymmetry of inner dynein arms and inter-doublet links in Chlamydomonas flagella. J Cell Biol. 2009;186(3):437-46.

17.   Oda T, Yanagisawa H, Kamiya R, Kikkawa M. A molecular ruler determines the repeat length in eukaryotic cilia and flagella. Science. 2014;346(6211):857-60.

18.   Jordan MA, Diener DR, Stepanek L, Pigino G. The cryo-EM structure of intraflagellar transport trains reveals how dynein is inactivated to ensure unidirectional anterograde movement in cilia. Nat Cell Biol. 2018;20(11):1250-5.

19.   Koonin EV. The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol. 2010;11(5):209.

20.   Shimogawa MM, Ray SS, Kisalu N, Zhang Y, Geng Q, Ozcan A, et al. Parasite motility is critical for virulence of African trypanosomes. Scientific Reports. 2018;8(1):9122.

21.   Rotureau B, Ooi Cp Fau - Huet D, Huet D Fau - Perrot S, Perrot S Fau - Bastin P, Bastin P. Forward motility is essential for trypanosome infection in the tsetse fly. Cell Microbiol. 2013(1462-5822 (Electronic)).

22.   Heddergott N, Kruger T, Babu SB, Wei A, Stellamanns E, Uppaluri S, et al. Trypanosome motion represents an adaptation to the crowded environment of the vertebrate bloodstream. PLoS Pathog. 2012;8(11):e1003023.

23.   Hughes LC, Ralston KS, Hill KL, Zhou ZH. Three-Dimensional Structure of the Trypanosome Flagellum Suggests that the Paraflagellar Rod Functions as a Biomechanical Spring. Plos One. 2012;7(1).

24.   Koyfman AY, Schmid MF, Gheiratmand L, Fu CJ, Khant HA, Huang D, et al. Structure of Trypanosoma brucei flagellum accounts for its bihelical motion. P Natl Acad Sci USA. 2011;108(27):11105-8.

25.   Cachon J, Cachon M, Cosson M-P, J C. The paraflagellar rod: a structure in search of a function. Biol Cell. 1988;63:169-81.

26.   Santrich C, Moore L, Sherwin T, Bastin P, Brokaw C, Gull K, et al. A motility function for the paraflagellar rod of Leishmania parasites revealed by PFR-2 gene knockouts. Mol Biochem Parasitol. 1997;90(1):95-109.

27.   Sun SY, Kaelber JT, Chen M, Dong X, Nematbakhsh Y, Shi J, et al. Flagellum couples cell shape to motility in Trypanosoma brucei. Proc Natl Acad Sci U S A. 2018;115(26):E5916-E25.

28.   Walker PJ. Organization of function in trypanosome flagella. Nature. 1961;189:1017-8.

29.   Walker PJ, Walker JC. Movement of trypanosome flagella. J of Protozoology. 1963;10, suppl(3, abstract 109):32.

30.   Rodriguez JA, Lopez MA, Thayer MC, Zhao Y, Oberholzer M, Chang DD, et al. Propulsion of African trypanosomes is driven by bihelical waves with alternating chirality separated by kinks. Proc Natl Acad Sci U S A. 2009;106(46):19322-7.

31.   Jahn TL, Bovee EC. Locomotion of Blood Protists. In: Weinman D, Ristic M, editors. Infectious Blood Diseases of Man and Animals. 1. NY: Academic Press; 1968. p. 393-436.

32.   Bargul JL, Jung J, McOdimba FA, Omogo CO, Adung'a VO, Kruger T, et al. Species-Specific Adaptations of Trypanosome Morphology and Motility to the Mammalian Host. PLoS Pathog. 2016;12(2):e1005448.

33.   Gruby M. Recherches et observations sur une nouvelle espèce d'hématozoaire, Trypanosoma sanguinis. Comptes rendus hebdomadaire des séances de l'Académie des Sciences, Paris. 1843;17:1134-6.

34.   Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nat Methods. 2014;11(1):63-5.

35.   Kikkawa M. Big steps toward understanding dynein. The Journal of Cell Biology. 2013;202(1):15.

36.   Reck-Peterson SL, Yildiz A, Carter AP, Gennerich A, Zhang N, Vale RD. Single-Molecule Analysis of Dynein Processivity and Stepping Behavior. Cell. 2006;126(2):335-48.

37.   Coy DL, Wagenbach M, Howard J. Kinesin takes one 8-nm step for each ATP that it hydrolyzes. J Biol Chem. 1999;274(6):3667-71.

38.   Nicastro D, Fu X, Heuser T, Tso A, Porter ME, Linck RW. Cryo-electron tomography reveals conserved features of doublet microtubules in flagella. Proceedings of the National Academy of Sciences. 2011;108(42):E845.

39.   Pigino G, Maheshwari A, Bui KH, Shingyoji C, Kamimura S, Ishikawa T. Comparative structural analysis of eukaryotic flagella and cilia from Chlamydomonas, Tetrahymena, and sea urchins. Journal of Structural Biology. 2012;178(2):199-206.

40.   Heuser T, Dymek EE, Lin J, Smith EF, Nicastro D. The CSC connects three major axonemal complexes involved in dynein regulation. Molecular Biology of the Cell. 2012;23(16):3143-55.

41.   Owa M, Uchihashi T, Yanagisawa HA, Yamano T, Iguchi H, Fukuzawa H, et al. Inner lumen proteins stabilize doublet microtubules in cilia and flagella. Nat Commun. 2019;10(1):1143.

42.   Lin J, Yin W, Smith MC, Song K, Leigh MW, Zariwala MA, et al. Cryo-electron tomography reveals ciliary defects underlying human RSPH1 primary ciliary dyskinesia. Nature Communications. 2014;5:5727.

43.   Burgess SA, Walker ML, Sakakibara H, Knight PJ, Oiwa K. Dynein structure and power stroke. Nature. 2003;421:715.

44.   Bui KH, Yagi T, Yamamoto R, Kamiya R, Ishikawa T. Polarity and asymmetry in the arrangement of dynein and related structures in the Chlamydomonas axoneme. J Cell Biol. 2012;198(5):913-25.

45.   Yagi T, Minoura I, Fujiwara A, Saito R, Yasunaga T, Hirono M, et al. An Axonemal Dynein Particularly Important for Flagellar Movement at High Viscosity: IMPLICATIONS FROM A NEW CHLAMYDOMONAS MUTANT DEFICIENT IN THE DYNEIN HEAVY CHAIN GENE DHC9. Journal of Biological Chemistry. 2005;280(50):41412-20.

46.   Capewell P, Cren-Travaillé C, Marchesi F, Johnston P, Clucas C, Benson RA, et al. The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. eLife. 2016;5:e17716.

47.   Trindade S, Rijo-Ferreira F, Carvalho T, Pinto-Neves D, Guegan F, Aresta-Branco F, et al. Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. Cell Host Microbe. 2016;19(6):837-48.

48.   Schuster S, Krüger T, Subota I, Thusek S, Rotureau B, Beilhack A, et al. Developmental adaptations of trypanosome motility to the tsetse fly host environments unravel a multifaceted in vivo microswimmer system. eLife. 2017;6:e27656.

49.   Heuser T, Raytchev M, Krell J, Porter ME, Nicastro D. The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. The Journal of Cell Biology. 2009;187(6):921.

50.   Yamamoto R, Song K, Yanagisawa H-a, Fox L, Yagi T, Wirschell M, et al. The MIA complex is a conserved and novel dynein regulator essential for normal ciliary motility. The Journal of Cell Biology. 2013;201(2):263.

51.   Olbrich H, Cremers C, Loges Niki T, Werner C, Nielsen Kim G, Marthin June K, et al. Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the

Nexin-Dynein Regulatory Complex. The American Journal of Human Genetics. 2015;97(4):546-54.

52.   Wirschell M, Olbrich H, Werner C, Tritschler D, Bower R, Sale WS, et al. The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. Nature Genetics. 2013;45:262.

53.   Ralston KS, Hill KL. Trypanin, a component of the flagellar dynein regulatory complex, is essential in bloodstream form African trypanosomes. Plos Pathogens. 2006;2(9):873-82.

54.   Ralston KS, Lerner AG, Diener DR, Hill KL. Flagellar Motility Contributes to Cytokinesis in Trypanosoma brucei and Is Modulated by an Evolutionarily Conserved Dynein Regulatory System. Eukaryotic Cell. 2006;5(4):696-711.

55.   Nguyen HT, Sandhu J, Langousis G, Hill KL. CMF22 Is a Broadly Conserved Axonemal Protein and Is Required for Propulsive Motility in Trypanosoma brucei. Eukaryotic Cell. 2013;12(9):1202-13.

56.   Kabututu ZP, Thayer M, Melehani JH, Hill KL. CMF70 is a subunit of the dynein regulatory complex. Journal of Cell Science. 2010;123(Pt 20):3587-95.

57.   Bower R, Tritschler D, VanderWaal K, Perrone CA, Mueller J, Fox L, et al. The N-DRC forms a conserved biochemical complex that maintains outer doublet alignment and limits microtubule sliding in motile axonemes. Molecular Biology of the Cell. 2013;24(8):1134-52.

58.   Lin J, Tritschler D, Song K, Barber CF, Cobb JS, Porter ME, et al. Building Blocks of the Nexin-Dynein Regulatory Complex in Chlamydomonas Flagella. Journal of Biological Chemistry. 2011;286(33):29175-91.

59.   Huang B, Ramanis Z, Luck DJL. Suppressor mutations in chlamydomonas reveal a regulatory mechanism for flagellar function. Cell. 1982;28(1):115-24.

60.   Song K, Awata J, Tritschler D, Bower R, Witman GB, Porter ME, et al. In Situ Localization of N and C Termini of Subunits of the Flagellar Nexin-Dynein Regulatory Complex (N-DRC) Using SNAP Tag and Cryo-electron Tomography. Journal of Biological Chemistry. 2015;290(9):5341-53.

61.   Oda T, Yanagisawa H, Kikkawa M. Detailed structural and biochemical characterization of the nexin-dynein regulatory complex. Molecular Biology of the Cell. 2014;26(2):294-304.

62.   Awata J, Song K, Lin J, King SM, Sanderson MJ, Nicastro D, et al. DRC3 connects the N-DRC to dynein g to regulate flagellar waveform. Mol Biol Cell. 2015;26(15):2788-800.

63.   Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, et al. Flagellar motility is required for the viability of the bloodstream trypanosome. Nature. 2006;440(7081):224-7.

64.   Ralston KS, Hill KL. The flagellum of Trypanosoma brucei: new tricks from an old dog. Int J Parasitol. 2008;38(8-9):869-84.

65.   Vickerman K. On The Surface Coat and Flagellar Adhesion in Trypanosomes. Journal of Cell Science. 1969;5(1):163.

66.   ANDERSON WA, ELLIS RA. Ultrastructure of Trypanosoma lewisi: Flagellum, Microtubules, and the Kinetoplast. The Journal of Protozoology. 1965;12(4):483-99.

67.   Ichikawa M, Liu D, Kastritis PL, Basu K, Hsu TC, Yang S, et al. Subnanometre-resolution structure of the doublet microtubule reveals new classes of microtubule-associated proteins. Nature Communications. 2017;8:15035.

68.   Vaughan S, Shaw M, Gull K. A post-assembly structural modification to the lumen of flagellar microtubule doublets. Current Biology. 2006;16(12):R449-R50.

69.   Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A. 2009;106(10):3859-64.

70.   Dawson SC, Paredez AR. Alternative cytoskeletal landscapes: cytoskeletal novelty and evolution in basal excavate protists. Curr Opin Cell Biol. 2013;25(1):134-41.

71.   Bastin P, Sherwin T, Gull K. Paraflagellar rod is vital for trypanosome motility. Nature. 1998;391(6667):548.

72.   Ichikawa M, Bui KH. Microtubule Inner Proteins: A Meshwork of Luminal Proteins Stabilizing the Doublet Microtubule. BioEssays. 2018;40(3):1700209.

73.   Stoddard D, Zhao Y, Bayless BA, Gui L, Louka P, Dave D, et al. Tetrahymena RIB72A and RIB72B are microtubule inner proteins in the ciliary doublet microtubules. Mol Biol Cell. 2018;29(21):2566-77.

74.   Bower R, Tritschler D, Mills KV, Heuser T, Nicastro D, Porter ME. DRC2/CCDC65 is a central hub for assembly of the nexin-dynein regulatory complex and other regulators of ciliary and flagellar motility. Mol Biol Cell. 2018;29(2):137-53.

75.   Song K, Shang Z, Fu X, Lou X, Girgorieff N, Nicastro D. Structure of the ciliary axoneme at nanometer resolution reconstructed by TYGRESS. BioRxiv. Unpublished;preprint.

76.   Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol Biochem Parasitol. 1999;99(1):89-101.

77.   Shimogawa MM, Saada EA, Vashisht AA, Barshop WD, Wohlschlegel JA, Hill KL. Cell Surface Proteomics Provides Insight into Stage-Specific Remodeling of the Host-Parasite Interface in Trypanosoma brucei. Mol Cell Proteomics. 2015;14(7):1977-88.

78.   Saada EA, Kabututu ZP, Lopez M, Shimogawa MM, Langousis G, Oberholzer M, et al. Insect stage-specific receptor adenylate cyclases are localized to distinct subdomains of the Trypanosoma brucei Flagellar membrane. Eukaryot Cell. 2014;13(8):1064-76.

79.   Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology. 2005;152(1):36-51.

80.   Fukuda Y, Laugks U, Lucic V, Baumeister W, Danev R. Electron cryotomography of vitrified cells with a Volta phase plate. J Struct Biol. 2015;190(2):143-54.

81.   Si Z, Zhang J, Shivakoti S, Atanasov I, Tao C-L, Hui WH, et al. Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. PLOS Pathogens. 2018;14(12):e1007452.

82.   Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods. 2013;10(6):584-90.

83.   Kremer JR, Mastronarde DN, McIntosh JR. Computer visualization of three-dimensional image data using IMOD. J Struct Biol. 1996;116(1):71-6.

84.   Xiong Q, Morphew MK, Schwartz CL, Hoenger AH, Mastronarde DN. CTF Determination and Correction for Low Dose Tomographic Tilt Series. Journal of structural biology. 2009;168(3):378-87.

85.   Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol. 2003;142(3):334-47.

86.   Fan X, Zhao L, Liu C, Zhang JC, Fan K, Yan X, et al. Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM. Structure. 2017;25(10):1623-30 e3.

87.   Heumann JM, Hoenger A, Mastronarde DN. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. Journal of Structural Biology. 2011;175(3):288-99.

88.   Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Science. 2018;27(1):14-25.

89.   Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004;25(13):1605-12.

90.   Kisalu NK, Langousis G, Bentolila LA, Ralston KS, Hill KL. Mouse infection and pathogenesis by Trypanosoma brucei motility mutants. Cellular Microbiology. 2014;16(6):912-24.

91.   Oberholzer M, Lopez MA, Ralston KS, Hill KL. Chapter 2 - Approaches for Functional Analysis of Flagellar Proteins in African Trypanosomes. In: King SM, Pazour GJ, editors. Methods in Cell Biology. 93: Academic Press; 2009. p. 21-57.

4.10 Figures
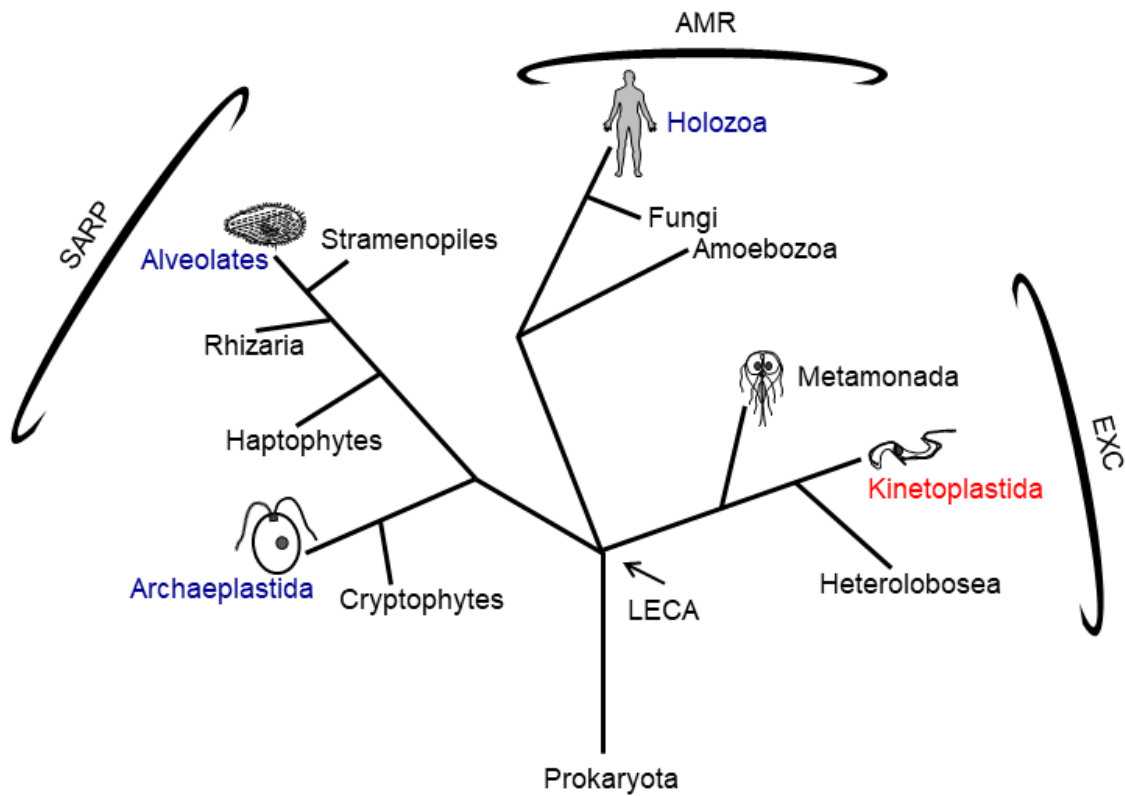
Figure 4- 1. Phylogenetic tree of eukaryotes

The tree is adapted from (84). High-resolution structures of the axoneme are published for the

clades indicated in blue, with the corresponding organism depicted in cartoon. *T. brucei* is in the

clade Kinetoplastida, indicated in red, and represents the Excavata (EXC) supergroup that includes

other pathogens, such as *Giardia* within Metamonada, also depicted in cartoon. The position of the

last eukaryotic common ancestor (LECA) is indicated. AMR: Amorphea; SARP: Stramenopila+

Alveolata + Rhizaria + Plantae; and EXC: Excavata are indicated.
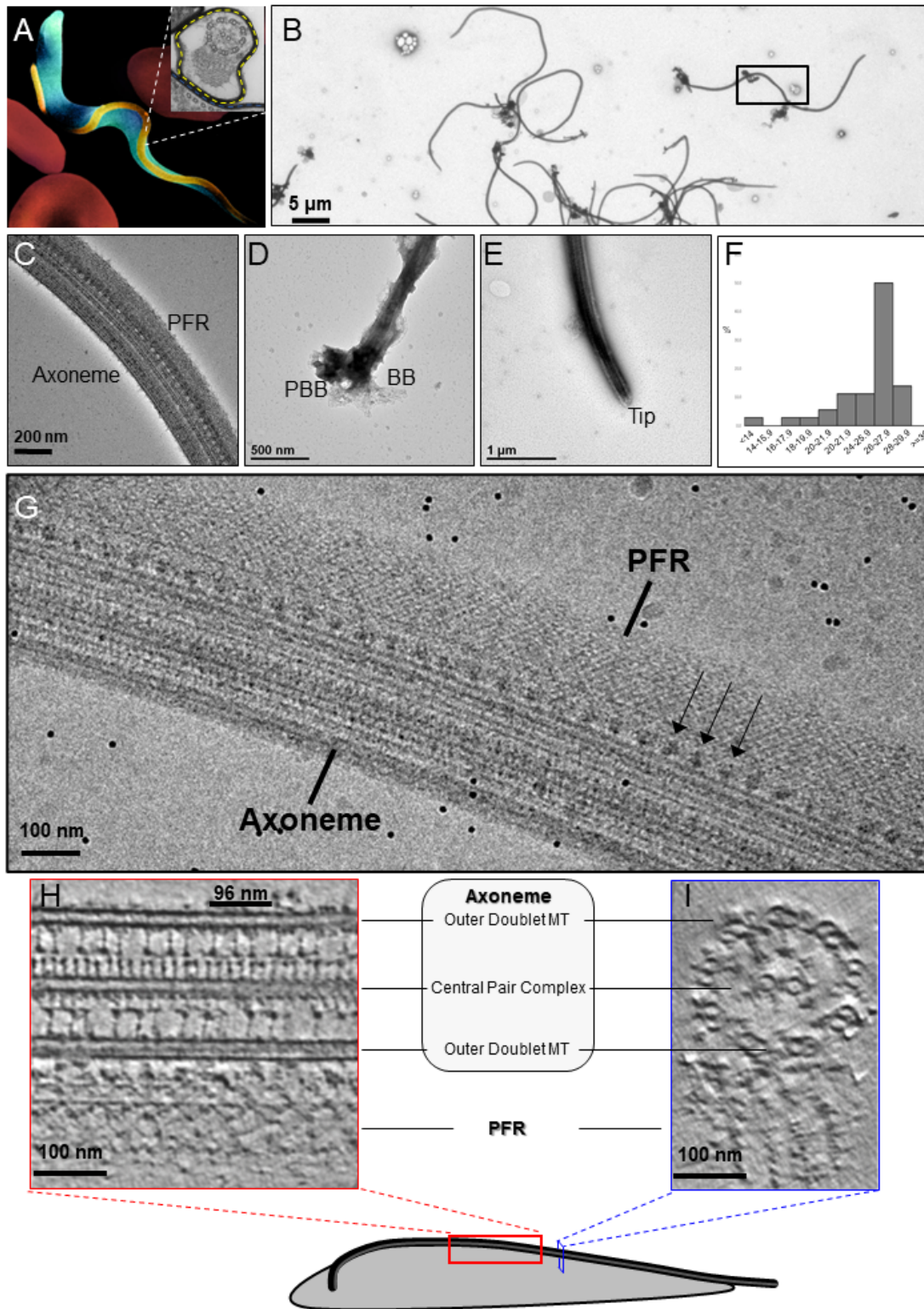
Figure 4- 2. Intact demembranated flagella from BSF *T. brucei*

(A) A representative scanning electron microscope image of a *T. brucei* parasite in blood from an

143

infected mouse, with the cell body colored blue, the flagellum yellow and host erythrocytes red. The inset is a transmission electron microscope image of the flagellum in representative transverse section, viewed from the proximal end, showing the 9+2 axoneme and PFR, enclosed within the flagellar membrane which is outlined by the yellow dotted line. (Adapted from (91)).

(B-E) Negative stain TEM images of purified flagellum samples, distributed on the grid with minimal clustering (B), showing that the axoneme and PFR are intact (C), with the basal body and pro-basal body on the proximal end (D), and a tapered tip at the distal end (E). The black box in (B) shows the approximate region chosen to image for cryoET

(F) Histogram of the length distribution of purified flagellum samples showing that the majority are full-length with a mean length of 25.2 microns (standard deviation = 3.5 microns).

(G) A zero-degree tilted cryoEM image shows intact Axoneme, PFR and Ax-PFR connectors (arrows).

(H-I) 6-Å thick digital slice from a representative tomogram showing the sample in longitudinal (H) and the transverse (I) sections, with main structures labelled. Black line indicates one 96-nm axonemal repeat.
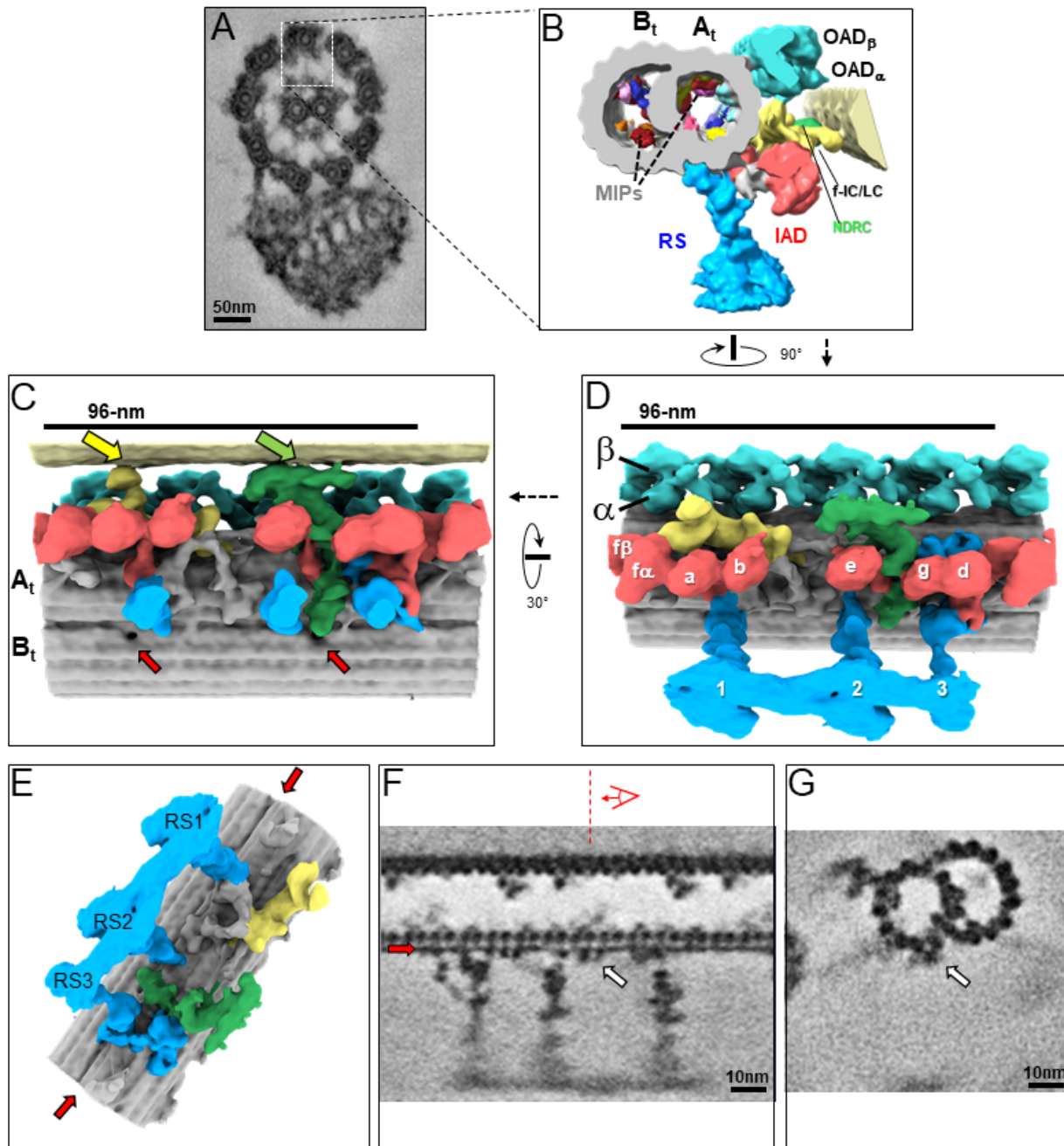
Figure 4- 3. The 3D ultrastructure of the 96-nm repeat from intact axonemes of BSF

T. brucei

(A) A representative cross-section of a demembranated and negative-stained *T. brucei* flagellum,

viewed from the proximal end (Adapted from (23). Boxed region orients the view of the averaged

96-nm repeat along a DMT shown in B.

(B) Cross-section view of the 96-nm repeat obtained by sub-tomogram averaging. Labeled are: the A- and B-tubule (At, Bt), Microtubule Inner Proteins (MIPs), Radial Spokes (RS), Inner Arm Dyneins (IAD), Nexin Dynein Regulatory Complex (NDRC), IAD-f-Intermediate Chain / Light Chain Complex (f-IC/LC), Outer Arm Dynein (OAD). The surface of the B-tubule from the adjacent DMT is visible on the right. The coloring scheme is as follows: cyan, OAD; red, IAD; blue, RS; green, NDRC; yellow, dynein f IC/LC. This scheme is consistently used throughout all main figures, supplemental figures and videos unless stated otherwise.

(C, D) Shaded surface rendering longitudinal views of the 96-nm repeat. Panel C shows the view from the center of the axoneme looking outward with the proximal end of the axoneme on the left and spoke heads removed for clarity (rotation relative to Panel D is shown). The surface of the B-tubule of the adjacent DMT is visible on top. Yellow and green arrows point to the inter-doublet connections formed by the f-connector and NDRC, respectively. White arrows point to the proximal and distal holes in the inner junction between the A- and B-tubules. Panel D shows the view from the adjacent DMT, with proximal end of the axoneme on the left (rotation relative to panel B is shown). For reference, alpha ( ) and beta ( ) OAD are indicated, individual IADs and RS are labeled.

(E) Shaded surface rendering of the averaged 96-nm repeat with the IAD, OAD and MIA complex removed, showing a massive structure at the base of the RS3 (see also Suppl. Fig. S1). Red arrows point to the density corresponding to the FAP59/172, 96-nm ruler (17) between protofilaments A2 and A3.

(F, G) Longitudinal (F) and transverse (G) density slices of the averaged 96-nm repeat. Red arrows in panels E and F point at the density of the FAP59/172 ruler between protofilaments A2 and A3. The red dashed line and perspective cartoon in panel F show the position and perspective of the cross-section shown in G, with the white arrow in panels F and G indicating the FAP59/172 ruler.
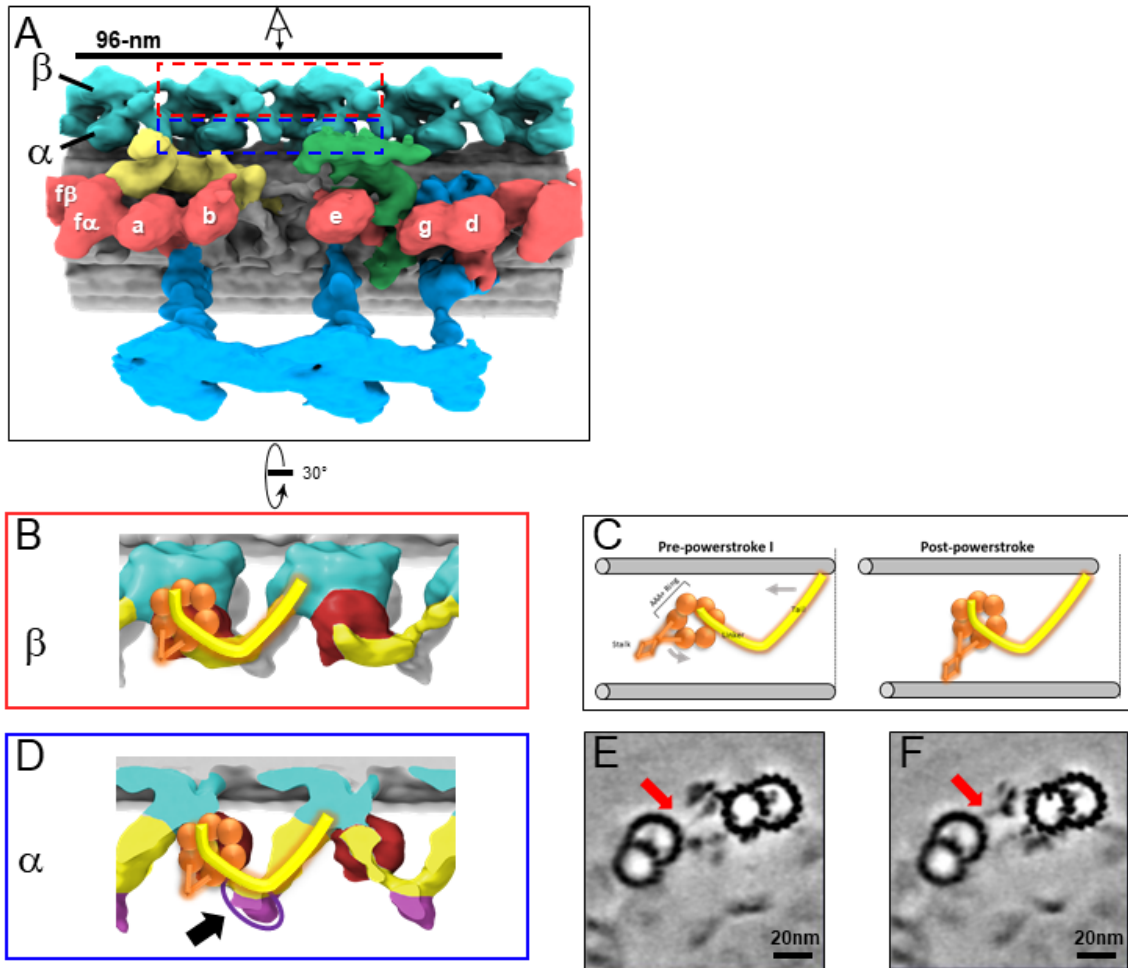
Figure 4- 4. *In situ* **structure of outer arm dyneins and novel OAD-alpha inter-doublet connector in BSF** *T. brucei*.

(A) Shaded surface rendering, longitudinal view of the averaged 96-nm repeat. Coloring as described for Figure 3A. The box around the OAD indicates the region and perspective shown in B (red box) and D (blue box).

(B, D) Shaded surface renderings of outer arm dyneins from the averaged 96-nm repeat. (B) Two adjacent OAD   dyneins. The linker and tail domains are colored yellow and the AAA+ ring is red. Cartoon overlay shows the post-powerstroke position of dynein. (D) Top view of two adjacent OADα dyneins. The linker and tail domains are colored in yellow and the AAA+ ring is colored in red. The arrow points to the OADα   connector (purple), at the junction between the tail and linker

domains. Cartoon overlay shows the post-powerstroke position of dynein.

(C) A schematic illustrating relative DMT movement as dynein moves from pre-powerstroke 1 state (left) to post-powerstroke state (right).

(E-F) Density slices of the averaged 96-nm repeat, viewed in cross-section, viewed from the distal tip of the axoneme. Red arrows indicate the dynein stalk domain in (E), and the OADα connector in (F), contacting the neighboring DMT.
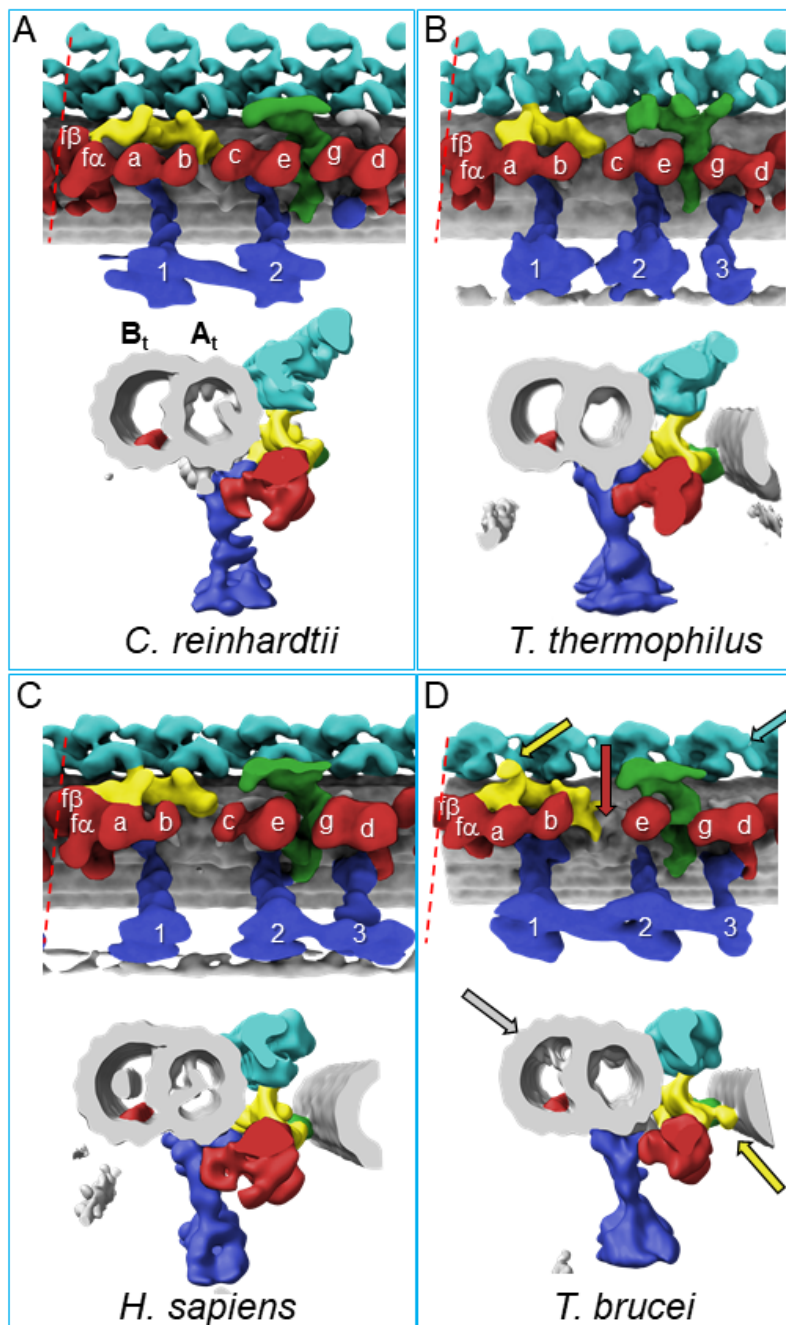
Figure 4- 5. Comparison of 96-nm axonemal repeat structures across species.

(A-D) Structure of the 96-nm axonemal repeat is shown for *Chlamydomonas reinhardtii* (A) (41), *Tetrahymena thermophilus* (B)(39) , *Homo sapiens* (C) (42) and BSF *Trypanosoma brucei* (D) (this work). Longitudinal (top) and cross-sectional (bottom) views are shown for each. Canonical features of the 96-nm repeat are colored, including outer arm dyneins (cyan), inner arm dyneins (red and numbered according to convention) the IC/LC complex of inner arm dynein f (yellow), the NDRC inter-doublet linkage (green) and radial spokes (blue). The microtubule lattice is gray and the A- and B-tubules are indicated. MIP3 (red) is present in all organisms shown and is colored in the B-tubule for reference. For all structures except that from *C. reinhardtii*, the surface of the B-tubule from the adjacent DMT is shown. Inner dyneins and radial spokes are labeled for reference. The red dashed line indicates the position of viewing for the cross-section shown. All structures are filtered to resolution of 50 Å. Features that distinguish the *T. brucei* repeat include the f-connector (yellow arrow), missing dynein-c (red arrow), lineage specific MIPs within the A- and B-tubules (gray arrow), and two OAD motors in a protist (cyan arrow). Other *T. brucei*-specific structures, such as the OAD-alpha inter-doublet connector and b-connector are not visible in this view.
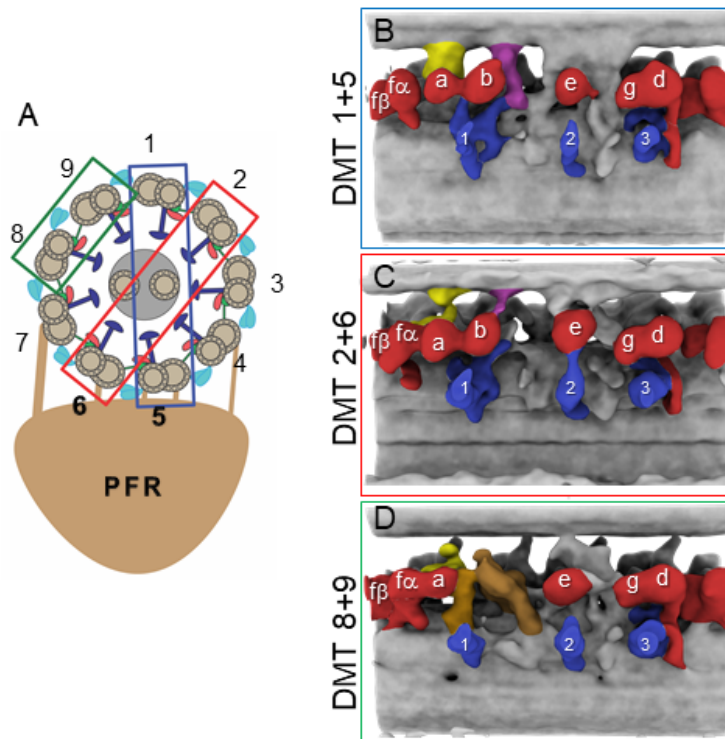
Figure 4- 6. Doublet-specific structures of the BSF *T. brucei* 96-nm repeat.

(A) Schematic showing the numbering of individual DMTs.

(B-D) Panels show averaged structures for DMT pairs 1+5 (B), 2+6 (C), and 8+9 (D). Inner arm dyneins (red) and radial spokes (blue) are labeled for reference. The f-connector, b-connector and the arch that distinguish DMTs 8 and 9 are colored yellow, purple and brown, respectively.
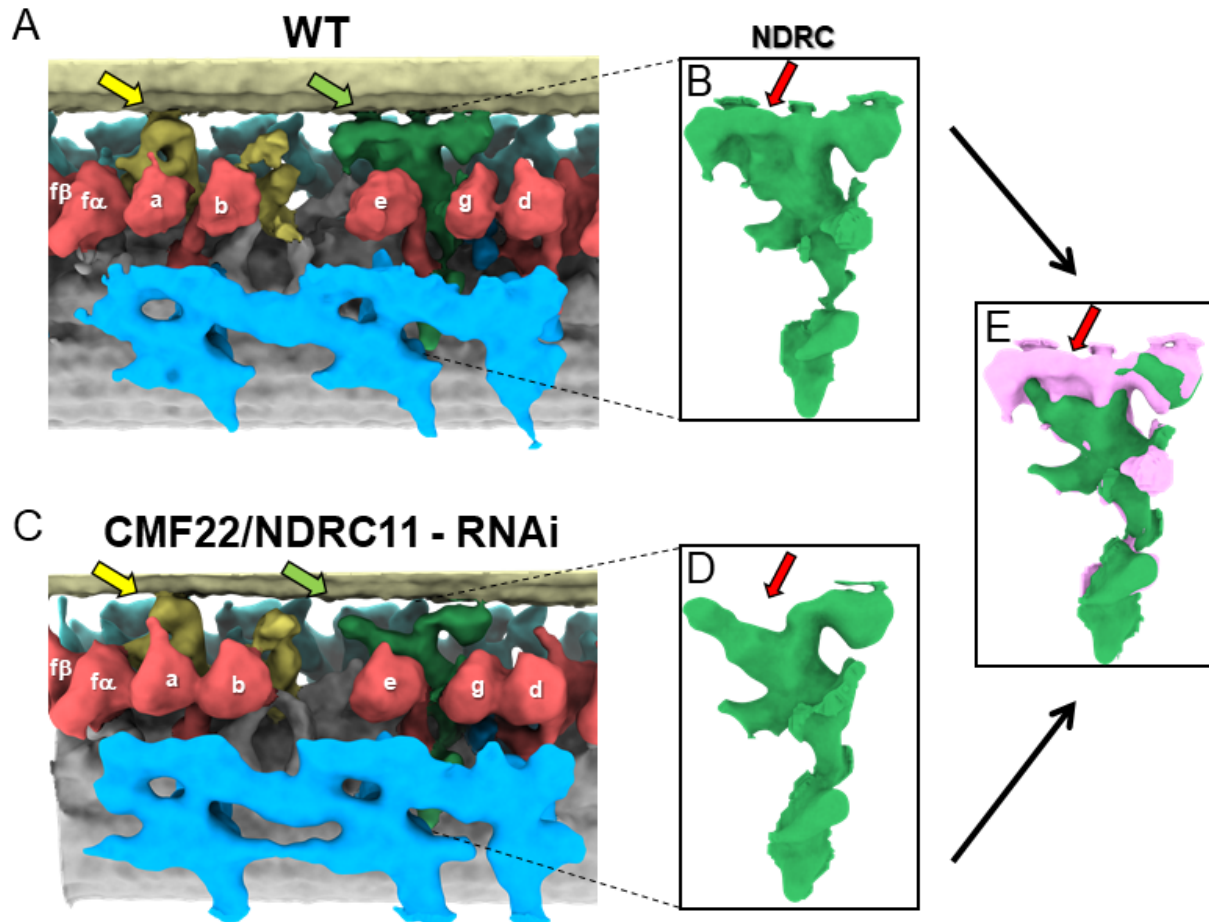
Figure 4- 7. Comparison between averaged 96-nm repeats of wild-type and CMF22/DRC11 knockdown PCF *T. brucei*

(A, C) Sub-tomogram averages of the 96-nm repeats of wild-type (A), and CMF22/DRC11 knockdown mutant (C). Yellow and green arrows point to the region of the B-tubule contacted by the f-connector and NDRC, respectively.

(B, D) Zoomed-in view of the NDRC from WT (B) and CMF22/DRC11 knockdown (D). The red arrow in each panel denotes the structure most substantially affected in the knockdown.

(E) Superposition of the NDRC structures shown in B and D, with WT in pink and the mutant in green. The red arrow indicates the most striking difference, corresponding to inter-doublet contacts made by the NDRC.
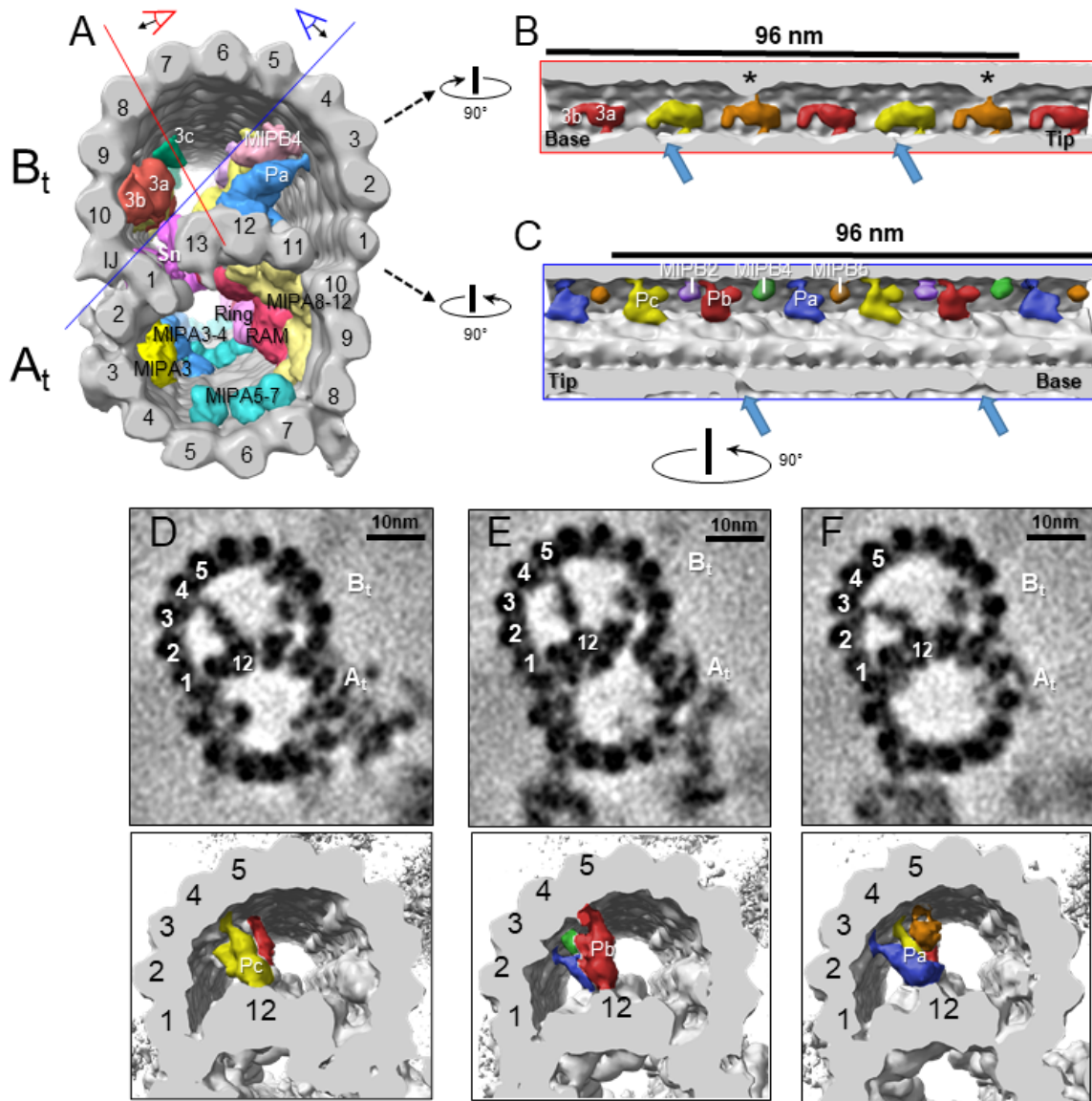
Figure 4- 8. TbMIP3 and ponticulus in the B-tubule of BSF *T. brucei*.

(A) Guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Red and blue lines indicate sections and viewing perspectives shown in panels (B) and (C), respectively.

(B) Longitudinal view into the inside of the B-tubule showing structural variations of TbMIP3 (red, yellow and orange) described in the text, with a periodicity of 48 nm. Arrows indicate the proximal and distal holes in the inner junction. Asterisk indicates MIP3a attachment to a structure identified

as MIP3c in *Chlamydomonas* (41). Proximal (base) and distal (tip) ends of the repeat are indicated and rotation relative to panel A is shown.

(C) Longitudinal view into the inside of the B-tubule showing ponticulus complexes Pa, Pb and Pc with a periodicity of 48nm. Arrows indicate the distal and proximal holes in the inner junction and rotation relative to panel A is shown.

(D-F) Top panels show cross-sections of average density maps viewed from the axoneme's distal tip to proximal end into the DMT. A subset of protofilaments are labeled for reference and rotation relative to panel C is shown. The trypanosome-specific Ponticulus (Pa, Pb and Pc) is seen bridging the entire lumen of the B-tubule from protofilament A12 to protofilaments B3, B5, and B4, respectively. The corresponding 3D isosurface renderings, looking from the same position are shown below, with Ponticulus-Pa, Pb and Pc, colored in blue, red and yellow respectively.

Figure 4- 9. The RingMIP and Ring Associated MIP (RAM) in the A-tubule of BSF *T. brucei*.

(A) Guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Red and blue lines indicate sections and viewing perspectives shown in panels (B) and (C), respectively.

(B-C) Longitudinal view of the A-tubule, showing the RingMIP and RAM. Left panels are sections through averaged density maps and right panels are corresponding isosurface renderings showing the same structures. The RingMIP (fuchsia), as well as its neighboring Ring Associated MIP (RAM) (red) and MIPA5-7 (cyan) are shown. The proximal (base) and distal (tip) ends of the axoneme are indicated and rotation of panel C relative to panel B is shown.
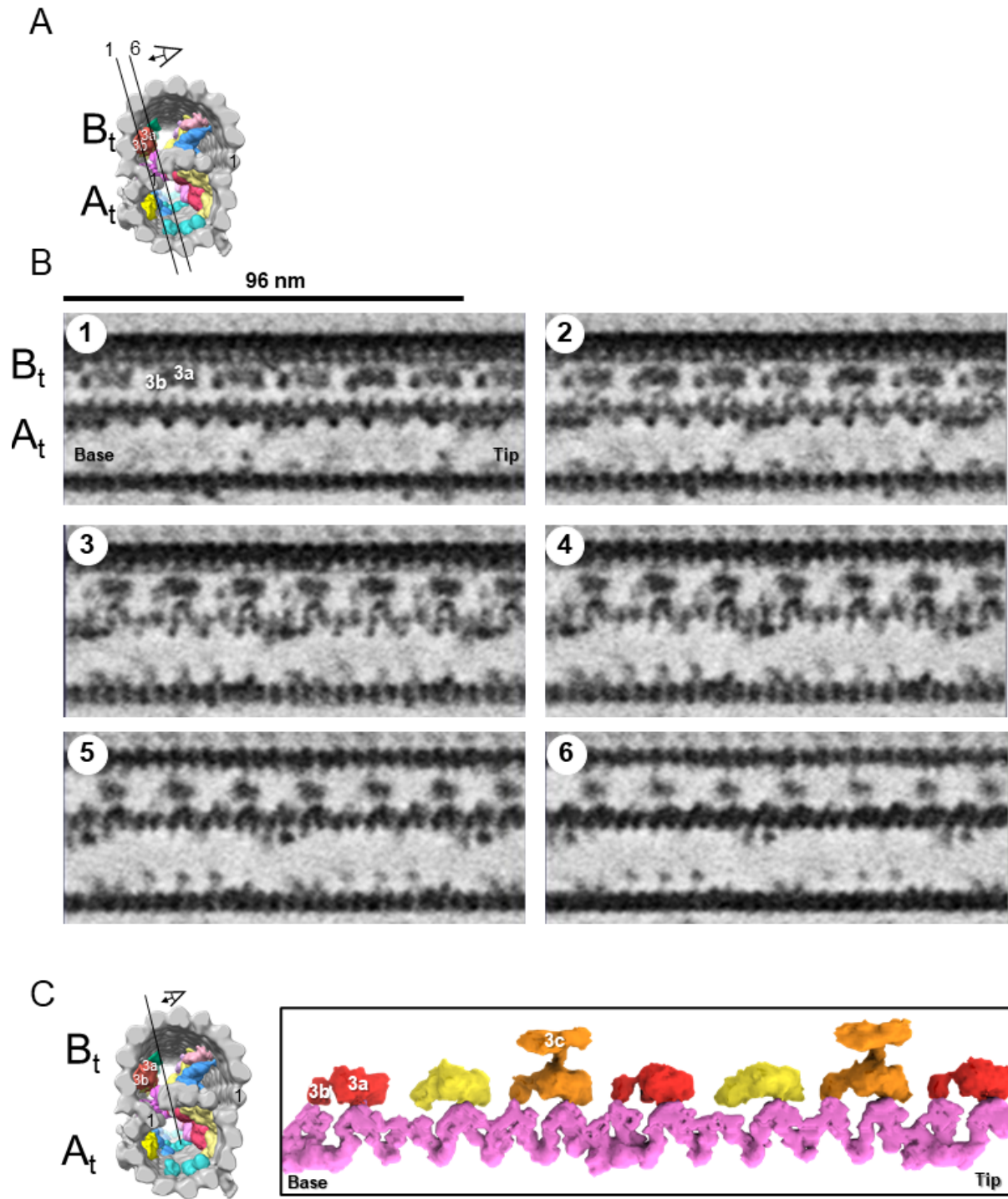
Figure 4- 10. The snake MIP connects the A-tubule and the B-tubule of BSF *T. brucei*.

(A) Guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the

proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Black lines 1 and 6 show the position and perspective of sections shown in B.

(B) Longitudinal view of the averaged density map. A and B-tubules are labeled. Panels 1 through 6 show six 6-Å thick, consecutive digital sections (the distance between 2 sections is 6.2 Å) through the snake MIP.
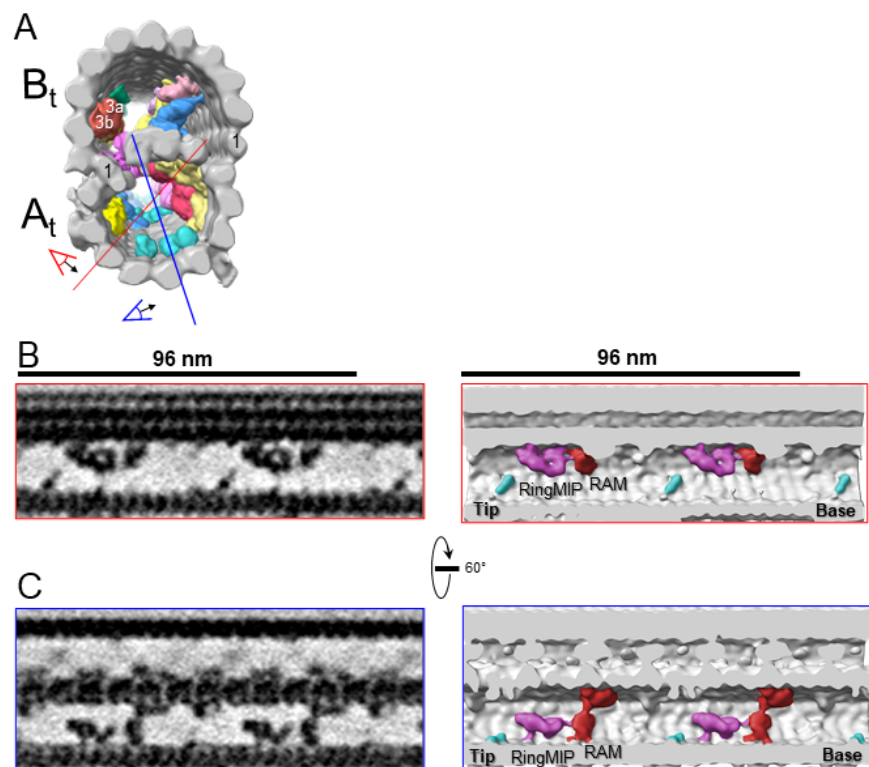
(C) Left panel is a guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Black line shows position and perspective for view of snake MIP shown in the right panel. Right panel shows segmented TbMIP3 (red, yellow and orange, as described for Figure 8B) and Snake MIP (mauve). (See also Video 7.)
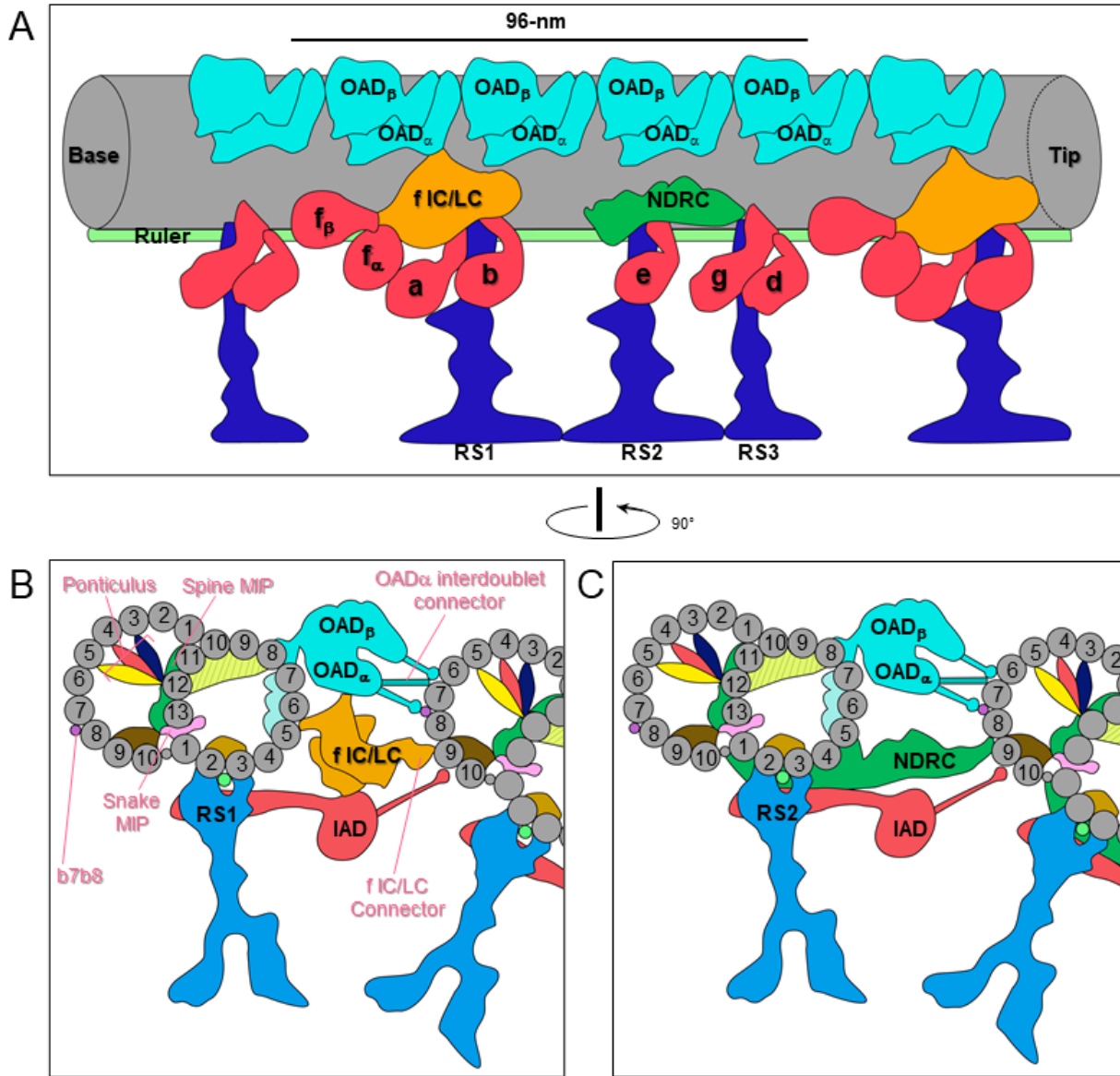
Figure 4- 11. Schematic overview of the trypanosome axoneme.

(A) Cartoon longitudinal view of the entire averaged 96-nm axonemal repeat. Major labeled structures are Outer Arm Dyneins (OAD), Inner Arm Dyneins (IAD), dynein-f IC/LC, Nexin Dynein Regulatory Complex (NDRC), Radial Spokes (RS) and Ruler. Image is oriented with proximal end (base) at the left.

(B and C) Cartoon cross-section view of the axoneme (viewed from the proximal end) at roughly the position of RS1 (B) and RS2 (C). Protofilaments are numbered and structures are labelled as

for panel A. Major trypanosome-specific structures described in text are labelled in pink. Note that

additional *T. brucei*-specific structures, RingMIP, RAM MIP and b-connector are not visible in this

simplified depiction. Summary of MIP structures is provided in Table 1.

4.11 Supplemental Figures
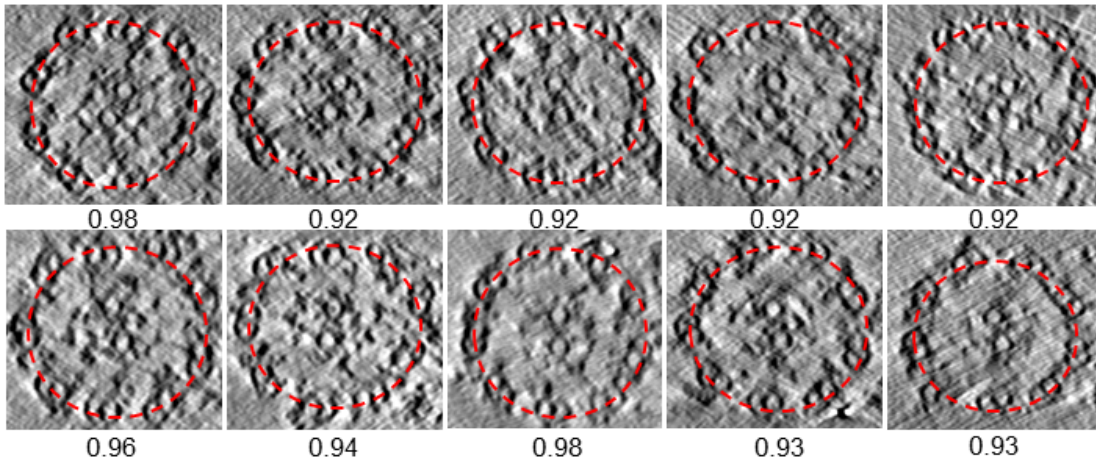
Supplemental Figure 4- 1. Cross sections of the ten tomograms used to obtain the entire averaged BSF 96-nm axonemal repeat structure. [related to Fig. 3]

Cross sections fitted with red dashed ellipses. The circularity, ratio of short axis/long axis of the

ellipse, is given below each axoneme and ranges from 0.92 to 0.98.
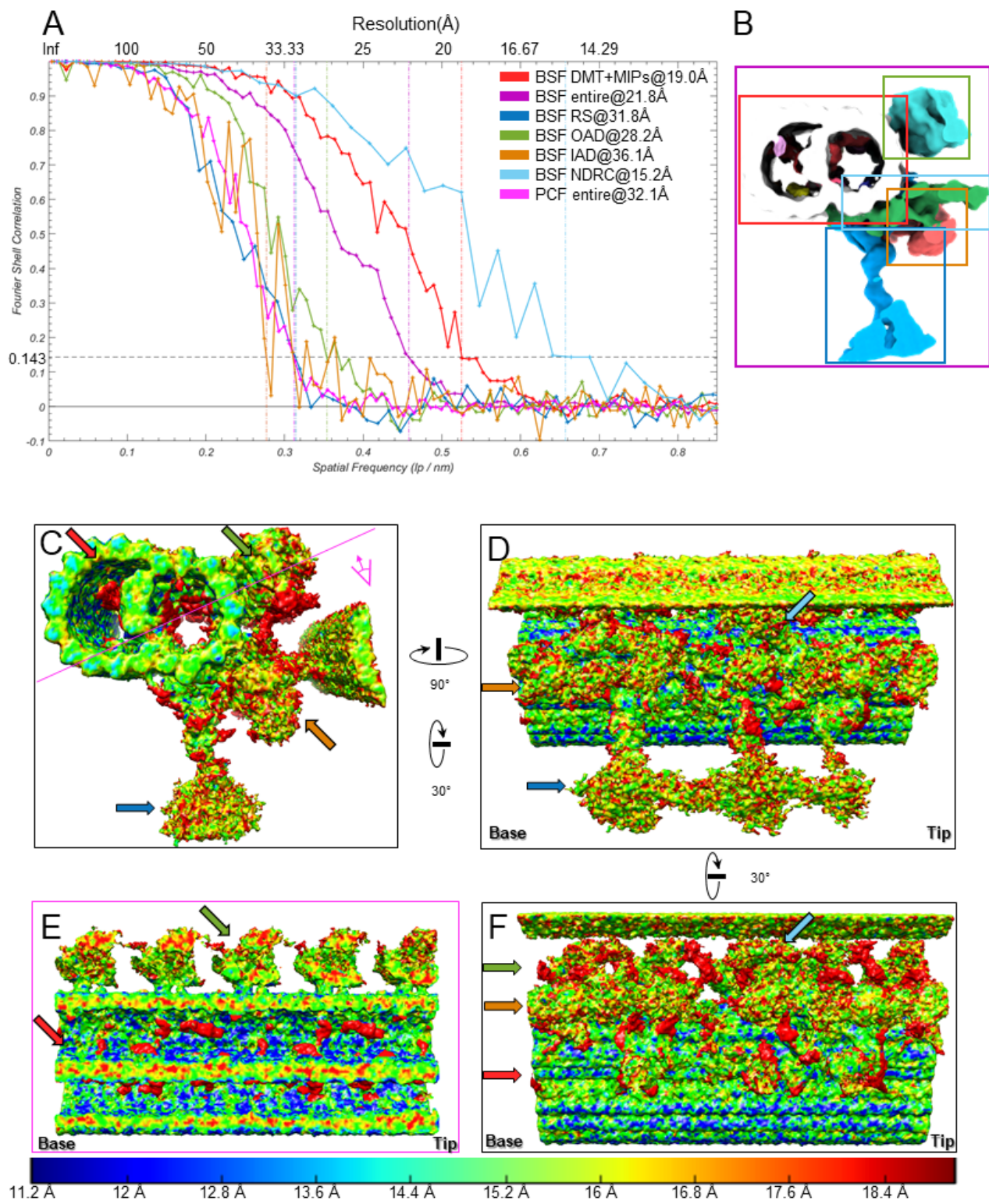
Supplemental Figure 4- 2. Gold Standard Fourier shell correlation (FSC) and ResMAP analyses. [related to Fig. 3]

(A) Fourier shell correlation (FSC) coefficients are plotted as a function of spatial frequency for sub-tomogram averages of the 96-nm axonemal repeats for either the entire maps of the BSF and PCF 96-nm repeat or the indicated local regions of the BSF 96-nm repeat. The resolutions based on the 0.143 criterion are indicated by vertical dashed lines. The FSC curve for some of the regions are rather ragged likely due to noisier densities and/or relative tight/small box used for selecting local regions.

(B) Local regions used for FSC analysis in panel A. The local regions were cut using a cuboid containing primarily DMT with MIPs, or OAD , or IAD, or NDRC, or RS as illustrated by the colored rectangles. Note that each cuboid also contains residual densities from nearby structures.

(C-F) Local resolution evaluation using the *ResMap* (34) program for the BSF structure shown in different views with local resolutions indicated in the bottom color bar. Colored arrows indicate different local regions used for local FSC calculation in panel A. Red = BSF DMT with MIPs, Purple = BSF entire 96-nm repeat, Blue = BSF RS, Green = BSF OAD, Brown = BSF IAD, Cyan = BSF NDRC and Pink = PCF entire 96-nm repeat. The views in C, D, F correspond roughly to Fig. 3 B, D, C respectively. Pink line in panel C indicates section and viewing perspective shown in panel E. The proximal (base) and distal (tip) ends of the axoneme are indicated and rotation of panel D relative to C and of panel F relative to D is shown.
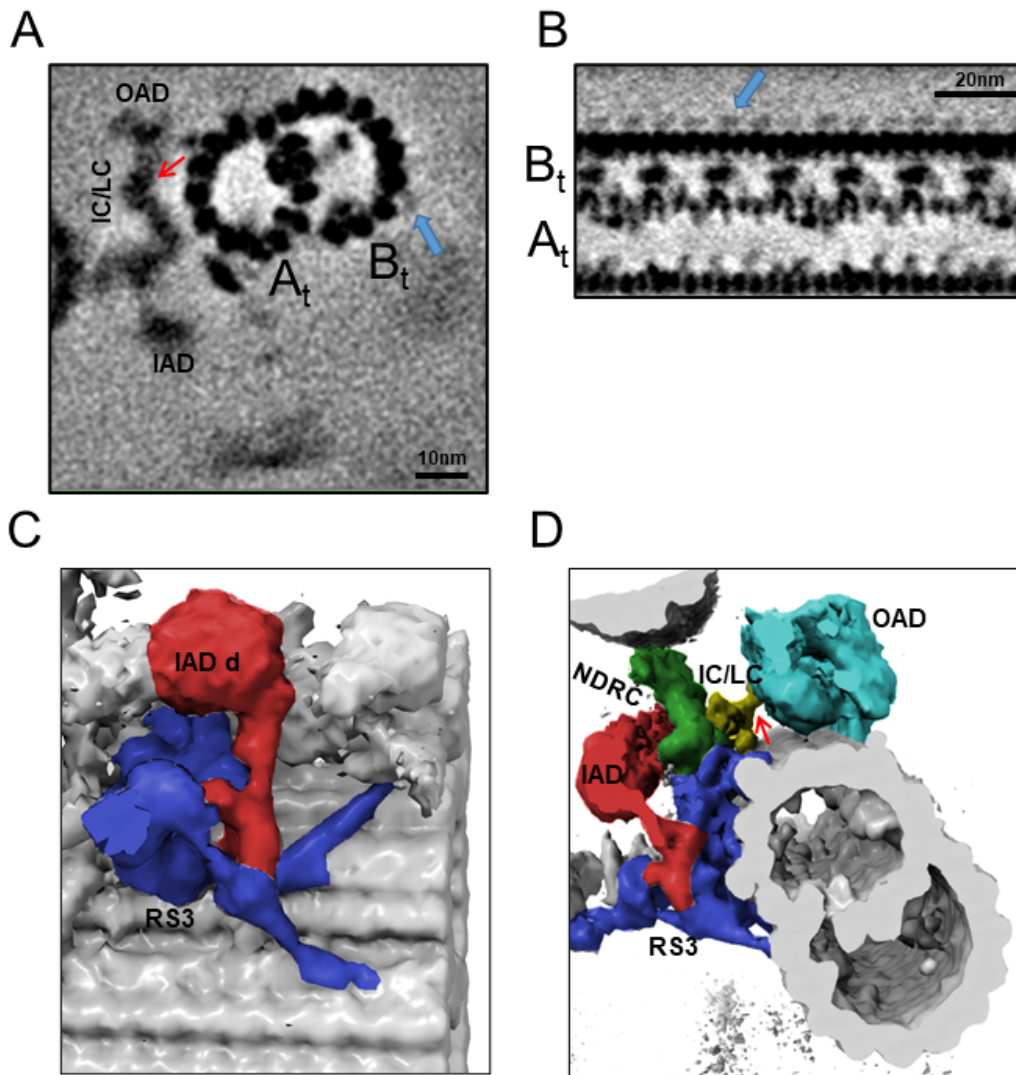
Supplemental Figure 4- 3. Extra densities outside protofilaments b7b8 and massive density at the base of RS3 in BSF *T. brucei*. [related to Fig. 3]

(A) Cross-section through the BSF averaged density map, viewed from the distal end of the axoneme and showing connections of the f-IC/LC complex to the IAD and OAD, red arrow. A and B-tubules are labeled and blue arrow points to the density outside protofilaments b7 and b8.

(B) Longitudinal section through the BSF averaged density map, with the proximal end of the axoneme at the left. A and B-tubules are labeled and blue arrow points to the novel density outside protofilaments b7 and b8.

(C) 3D isosurface rendering of the BSF averaged 96-nm repeat, showing the connection between IAD-d (red) and RS3 (blue).

(D) 3D isosurface rendering of the BSF averaged 96-nm repeat, viewed from the distal end. Extensive connection of the RS3 base to the A-tubule and inner junction is shown. No connection between the NDRC and the OAD is observed. Red arrow points to connection between f-IC/LC and OAD.
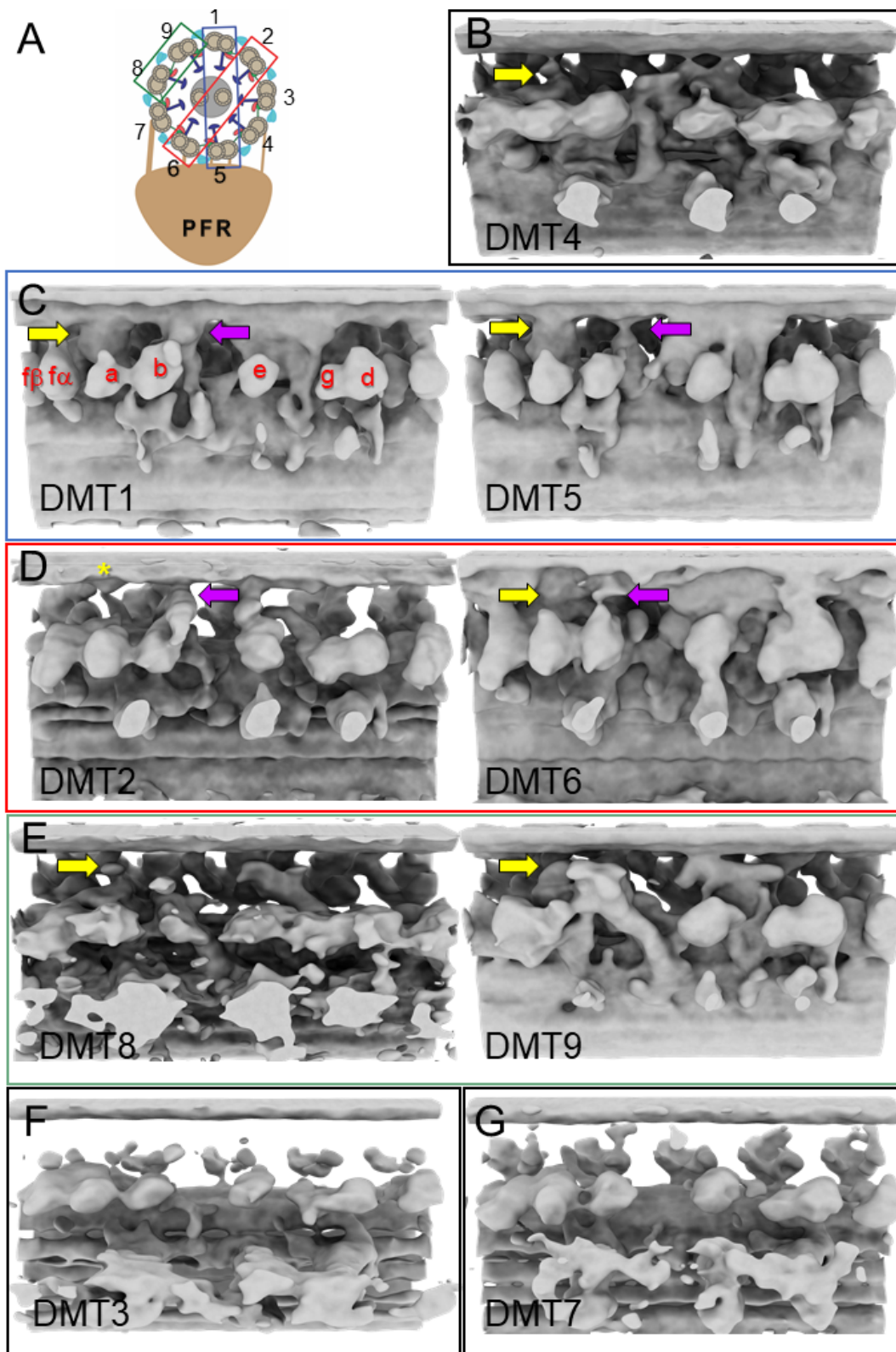
Supplemental Figure 4- 4. Sub-tomogram averages of the 96-nm repeat of individual DMTs of BSF *T. brucei*. [related to Fig. 6]

(A) Schematic showing the numbering of individual DMTs. Blue, red and green boxes indicate pairs of DMTs used for averaging in Figure 6.

(B-G) Images show the 96-nm repeat structure obtained by sub-tomogram averaging of each DMT individually. The f-connector (yellow arrow) and b-connector (purple arrow) are indicated. The asterisk on DMT2 indicates a density at the site where the f-connector typically would contact the B-tubule of the adjacent DMT. Inner arm dyneins are labelled a-f in DMT1 for reference.
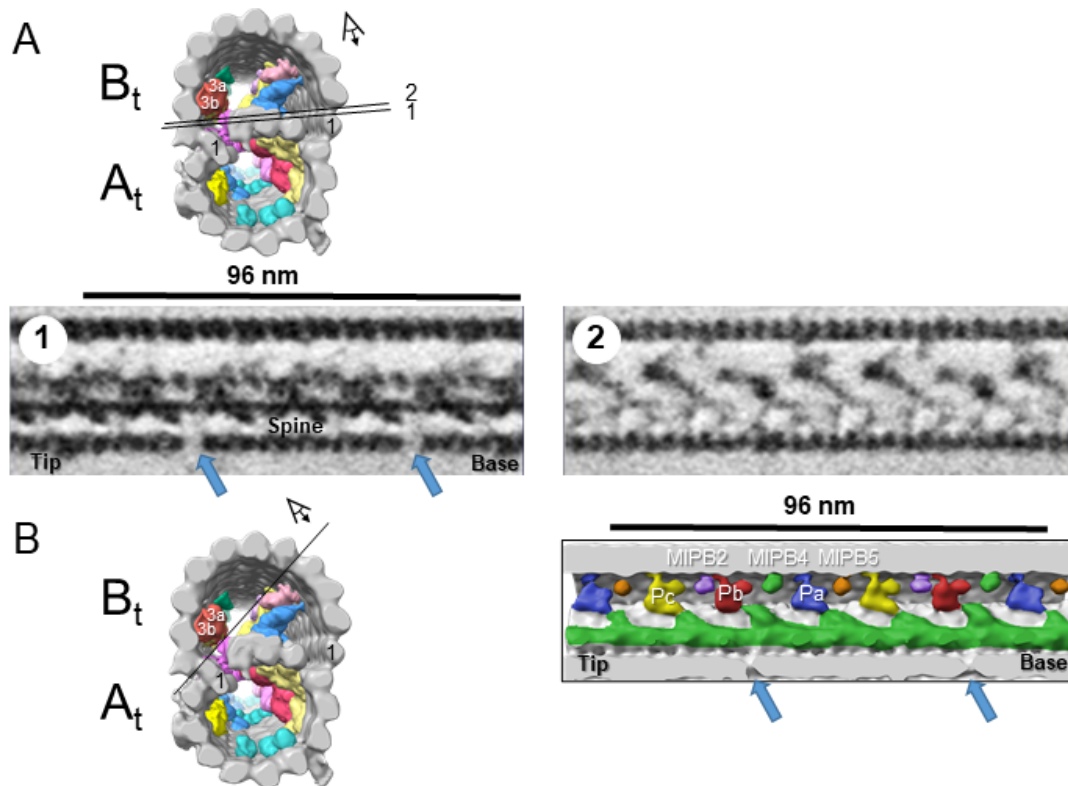
Supplemental Figure 4- 5. The Spine MIP is a contiguous structure, spanning 48 nm and contacting adjacent MIPs in BSF *T. brucei*. [related to Fig. 8]

(A) Top image is a guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Protofilament number 1 of the A and B-tubules are labeled for reference. Black lines 1 and 2 indicate the section and perspective of the two longitudinal views of the averaged density map shown below. Bottom images show longitudinal views 1 and 2, oriented with the proximal

(base) and distal (tip) ends of the axoneme indicated. The spine MIP runs along protofilament A13 and is connected to the ponticulus (Pa, Pb, Pc). The proximal and distal holes of the inner junction are visible in longitudinal view number 1 (blue arrows).

(B) Left image is a guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Protofilament number 1 of the A and B-tubules are labeled for reference. Black line indicates section and viewing perspective for the longitudinal surface rendering shown on the right. Right image shows connections of the Spine MIP to the ponticulus (Pa, Pb, Pc) in the B-tubule. The proximal (base) and distal (tip) ends of the axoneme indicated. The proximal and distal holes of the inner junction are visible in longitudinal view number 1 (blue arrows).
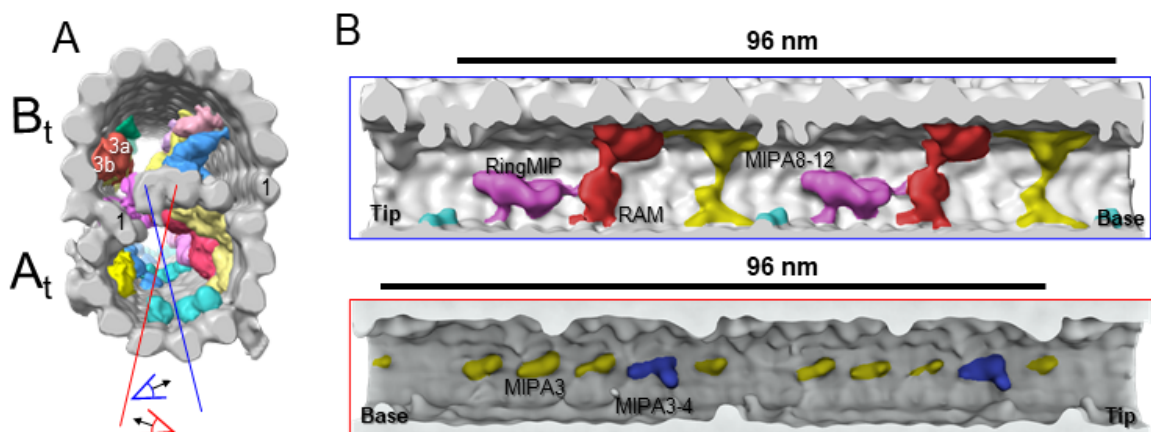
Supplemental Figure 4- 6. MIPs in the A-tubule of BSF *T. brucei*. [related to Fig. 9]

(A) Guide figure showing cross-section view of the averaged 96-nm repeat, viewed from the proximal end of the axoneme with MIPs colored and densities external to the DMT removed. Protofilament number 1 of the A and B-tubules are labeled for reference. Red and blue lines indicate sections and viewing perspectives shown in B.

(B) Shaded surface renderings show two different longitudinal views inside the A-tubule, oriented with the proximal (base) and distal (tip) ends of the axoneme as indicated. A-tubule MIPS: RingMIP, RAM, MIPA8-12, MIPA3 and MIPA3-4, are labeled and exhibit a 48-nm periodicity.
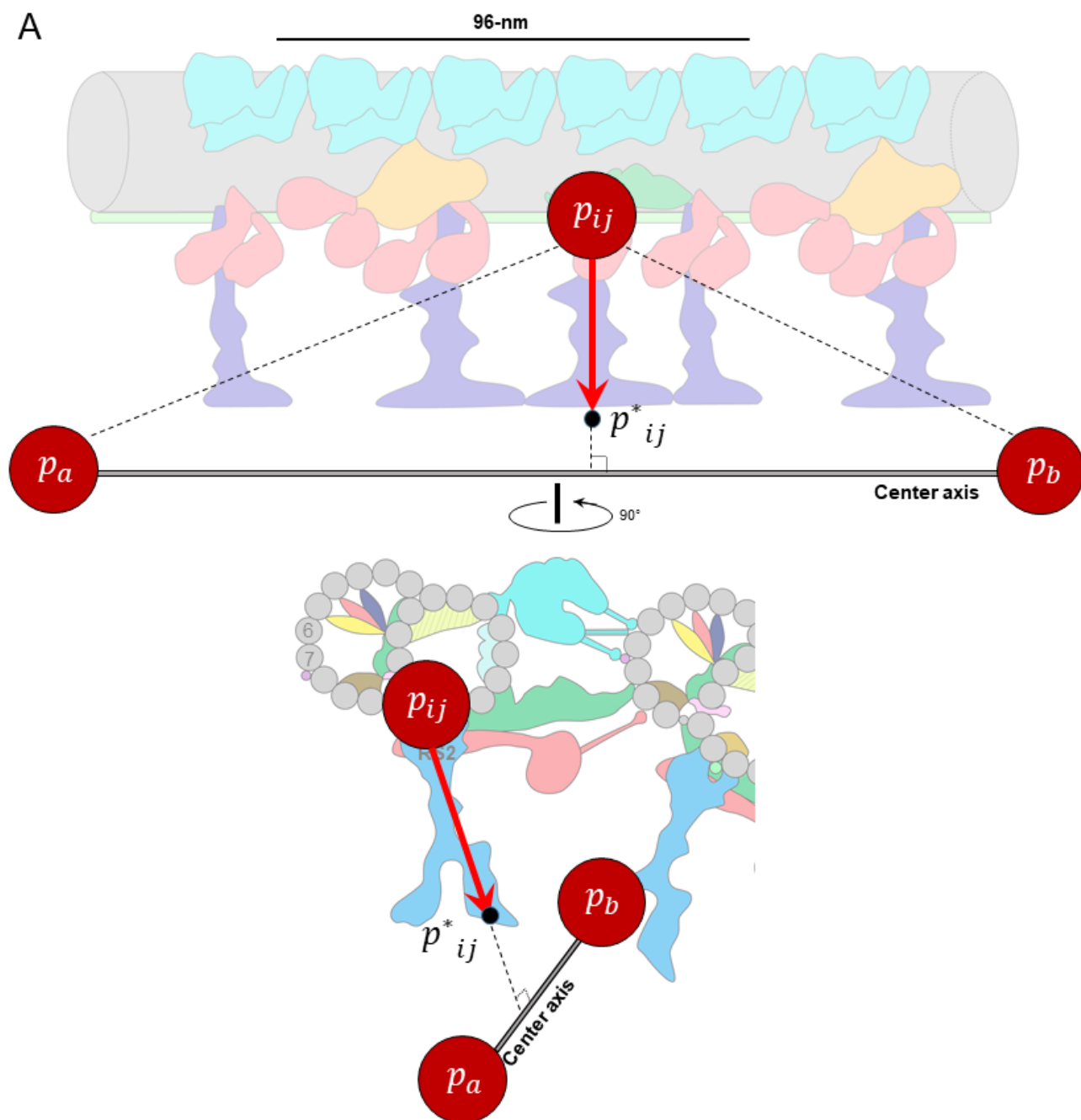
Supplemental Figure 4- 7. Illustration of the principles of *autoPicker*. [related to Fig. 3]

(A)p_{ij} is the intersection point of a DMT with the middle of the three radial spokes, RS2, along each particle's 96-nm unit length; The first two points, p_a and p_b , are the center points of the two bases of the cylinder. $\overrightarrow{p_{ij}p^*_{ij}}$ is perpenticular to $\overrightarrow{p_a p_b}$, and parallel to the plane defined by the three points of

p_a, p_b and p_{ij}.

4.12 Table
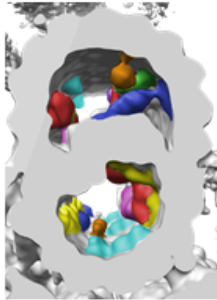
MIP Number	Tubule	PF	Times/4 8nm	Color	Figure
TbMIP3 a,b	B	B9-10	3	■ ■ ■	Figure 8
TbMIP3 c	B	B8	1	■	Figure 8,10
Spine MIP	B	A11-13	1		Suppl Fig. S5
Ponticulus a	B	A12, B3	1	■	Figure 8
Ponticulus b	B	A12, B5	1	■	Figure 8
Ponticulus c	B	A12, B4	1	■	Figure 8
MIPB5	B	B5	1	■	Figure 8
MIPB4	B	B4	1	■	Figure 8
MIPB2	B	B3	1	■	Figure 8
Snake MIP	A+B	A1+A13	1	■	Figure 10
MIPA3	A	A3	4	■	Suppl Fig. S6
MIPA3-4	A	A3-4	1	■	Suppl Fig. S6
MIPA5	A	A5	3	■	Figure 8
MIPA5-7	A	A5-7	1	■	Figure 9
MIPA8-12	A	A8-12	1	■	Suppl Fig. S6
RAM	A	A8-12	1	■	Figure 9
Ring MIP	A	A8, 9	1	■	Figure 9

Table 4- 1. MIPs of BSF *T. brucei*

Illustration of the tubule locations, protofilament numberings, periodicities, and their color coded in

each of the figures, of all different kinds of MIPS.

Chapter 5: Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells

Jiayan Zhang[1, 2, 3, *], Hui Wang[1, 3, 4, *], Simon Imhof[1], Xueting Zhou[1], Shiqing Liao[1], Ivo Atanasov[3], Wong H Hui[3], Kent L. Hill[1, 2, 3, @] & Z. Hong Zhou[1, 2, 3, @]

[*] co-first authors contribute equally to this project.

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

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

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

[4]Department of Bioengineering, UCLA, Los Angeles, California 90095, USA

Address for Correspondence:

Z. Hong Zhou (Hong.Zhou@UCLA.edu, 1-310-694-7527) or Kent L. Hill (kenthill@microbio.ucla.edu, 1-310-267-0546)

5.1 Abstract

Eukaryotic flagella (synonymous with cilia) rely on a microtubule-based axoneme, together with accessory filaments to carryout motility and signaling functions. While axoneme structures are well characterized, 3D ultrastructure of accessory filaments and their axoneme interface are mostly unknown, presenting a critical gap in understanding structural foundations of eukaryotic flagella. In the flagellum of the protozoan parasite *Trypanosoma brucei*, the axoneme is accompanied by a paraflagellar rod (PFR) that supports non-planar motility and signaling necessary for disease transmission and pathogenesis. Here, we employed cryogenic electron tomography (cryoET) with sub-tomographic averaging, to obtain structures of the PFR, PFR-axoneme connectors (PACs), and the axonemal central pair complex (CPC). The structures resolve how the 8nm repeat of the axonemal tubulin dimer interfaces with the 54nm repeat of the PFR, which consist of proximal, intermediate, and distal zones. In the distal zone, stacked "density scissors" connect with one another to form a "scissors stack network (SSN)" plane oriented 45° to the axoneme axis; and ~370 parallel SSN planes are connected by helix-rich wires into a paracrystalline array with ~90% empty space. Connections from these wires to the intermediate zone, then to overlapping layers of the proximal zone and to the PACs, and ultimately to the CPC point out a contiguous pathway for signal transmission. Together, our findings provide insights into flagellum-driven, non-planar helical motility of *T. brucei* and have broad implications ranging from cell motility and tensegrity in biology to engineering principles in bionics.

Key words: structural biology | *T. brucei* | PFR | cryoET | cell motility | cilium | flagellum

5.2 Introduction

Eukaryotic cells depend on flagella (synonymous with cilia [1]) to move through and respond to their external environment. In humans, flagellum motility and signaling are essential for normal development, physiology, and reproduction [2-4]. In protists and fungi, flagella enable navigation through diverse environments [1, 5-8], direct movement and interaction of gametes for reproduction [9, 10], and contribute to transmission and pathogenesis of microbial pathogens [11-14]. The structural foundation of the flagellum is the axoneme, a microtubule-based molecular machine that drives motility and provides a platform for assembly of signaling machinery [15]. In addition to the axoneme, flagella of many organisms contain accessory structures, such as outer dense fibers and fibrous sheath of human sperm [16], mastigonemes of algae [17], and the paraflagellar rod (PFR) of euglenoids and kinetoplastids [18, 19]. Biochemical and genetic analyses have demonstrated that these extra-axonemal structures contribute to flagellum motility and signaling functions, but the structural foundation for how they achieve this is unclear [17, 20-22]. Recent studies have resolved structures of the axoneme and axoneme subcomplexes in great detail, providing important insights into mechanisms and structural foundations of flagellum function [23-28]. However, much less is known about 3D ultrastructures of extra-axonemal filaments, and this presents a gap in understanding structural foundations of flagellum function in eukaryotes.

Among the most enigmatic of extra-axonemal structures is the PFR of euglenoids and kinetoplastids, a taxonomic group that includes several human and animal pathogens, such as *Trypanosoma brucei* (Fig. 1a) and related kinetoplastid parasites, as well as free-living *Euglena* and related species [18, 19]. *T. brucei* causes fatal sleeping sickness in humans and related diseases in livestock throughout sub-Saharan Africa, while other kinetoplastid parasites cause Chagas disease in the Americas and Leishmaniasis in tropical and subtropical regions globally [29]. The *T. brucei* PFR is a massive paracrystalline filament that runs parallel to the axoneme along most of its length and is connected to axoneme doublet microtubules (DMTs) 4-7 [19, 30, 31]. The exact function of the PFR is not known, but it is required for cell motility, and studies of mutants lacking major

171

portions of the PFR suggest it provides elastic resistance to axoneme bending [20, 21, 32]. Such internal resistance would be required for efficient movement in viscous environments where an organism must push against high external resistance, *e.g.*, blood and other tissues encountered by *T. brucei* [5, 33]. *T. brucei* motility is characterized by a vigorous, non-planar helical motion that must accommodate frequent flagellum beat reversals and collisions with external structures [5, 34-36]. Therefore, the PFR must have flexibility while maintaining structural integrity. The PFR also provides a platform for cAMP and Ca^{++} signaling systems that control motility and host-pathogen interactions [13, 37-42], and for metabolic activities that may participate in energy transfer within the flagellum [42, 43]. Trypanosome motility and PFR-dependent cAMP signaling are required for transmission and pathogenesis of these deadly pathogens [11-13, 39]. Therefore, the PFR presents both a model for understanding functions of extra-axonemal structures of eukaryotic flagella, and an attractive drug target in a group of organisms that pose a substantial global public health burden.

Proteomic and biochemical analyses have provided information on PFR composition [40, 44-46] and conventional electron microscopy and early tomography studies have provided a low resolution model for PFR structure [30, 31, 47-50]. However, high resolution 3D structures of the PFR and the PFR-axoneme interface are not available. Consequently, how the PFR and axoneme combine to direct and regulate the hallmark non-planar helical movement of *T. brucei* [5, 35, 36] remains a mystery.

Here, we have employed a combination of cryoET with Volta phase plate (VPP), energy filtering and direct electron-counting imaging, together with sub-tomographic averaging [51], to determine the structure of the entire *T. brucei* axoneme with PFR. The sub-tomographic averaged structure of the PFR distal zone reveals a largely hollow architecture comprised of planar networks of stacked "scissors densities" placed each 54nm along and oriented at 45° to the axoneme axis and connected by thin "wire densities". Such an architecture suggests tensegrity [52], rather than space-filling observed in other cellular structures [53], as the means to achieve competing needs of

172

integrity and flexibility. Structural features of wire densities are reminiscent of bundled helices, consistent with secondary structure predictions of major PFR proteins. Based on the sub-tomographic averaged structure of the PFR-axoneme interface and interconnections within the PFR, our work also provides details on interconnections within and between the PFR and axoneme that may provide a means for signaling within these complexes suggested previously [24, 27, 45, 54]. Together, our results provide insights into flagellar motility of *T. brucei* and have broad implications regarding functions of extra-axonemal filaments that are a common, yet enigmatic feature of eukaryotic flagella.

5.3 Results

5.3.1 Resolving T. brucei flagellum components with different periodicities

Presence of the PFR in *T. brucei* flagellum poses two major challenges for structural studies: possible structure distortion due to increased sample thickness and restricted flagellum orientation on the cryoEM grid due to deviation from circularity. To cope with these challenges, we developed a procedure to evaluate the width of the axoneme prior to recording [26] and relied on sub-tomographic averaging with wedge-mask differences [55]. We also used a machine learning-assisted method to compensate for the missing wedge problem (Methods). Tilt-series of *T. brucei* detergent-extracted flagellum samples from bloodstream form parasites were recorded in a Titan Krios electron microscope equipped with a VPP, an energy filter and a direct electron detector in electron-counting mode. Tomograms (Figs. 1b, c) were assembled as described in methods. The structure was well-preserved as indicated by presence of all major flagellum components, including the "9+2" axoneme, extra-axonemal PFR, and PFR-axoneme connectors (PACs) (Figs. 1b, c; Supplementary Fig. S1 and Supplementary Movie S1). Asymmetry of the PFR-axoneme interface, local arrangement of A and B tubules of DMTs, and orientation of axonemal dyneins, allowed unambiguous identification of the nine DMTs, numbered according to established convention [56] (Fig. 1c).

Based on cross-sectional views in traditional transmission electron microscopy (TEM) studies, the PFR consists of three structurally distinct zones: proximal, intermediate, and distal [30], and these are evident in tomograms (Figs. 1b, c; Supplementary Movie S1). One can readily identify repeating densities in the PFR distal zone that have the appearance of "comb teeth" in longitudinal sectional views (Figs. 1b, d; cyan arrows). Periodicities of major axoneme substructures, *e.g.,* radial spokes (RS), outer arm dyneins (OADs) or inner arm dyneins (IADs), CPC and microtubule inner proteins (MIPs), vary but all are integer multiples of the underlying tubulin dimer repeat of 8nm [26, 57]. By contrast, we determined the periodicity of PFR distal zone repeating units to be 54nm along the axoneme axis (Methods). A 54nm repeat interval is consistent with earlier measurements of 54 - 57nm [44, 48-50]. Since 54 is not a multiple of 8, resolving the details of axoneme and PFR structures simultaneously by sub-tomographic averaging is not possible. We therefore had to design a stepwise workflow incorporating both interactive and automatic particle-picking strategies (see Methods) to obtain sub-tomographic averaged structures for individual flagellum components, and then fitted them into a montage (Figs. 1d, e; Supplementary Movie S2) (Methods). Visualizing the RS, tubulin dimer, OAD, CPC, PACs, and different zones of PFR in this way allows us to decipher interactions among these components, as detailed below.

5.3.2 The PFR distal zone consists of a series of parallel SSN planes, aligned at 45° to the axoneme axis and interconnected by coiled-coil wires

Previous electron tomography studies have provided important insights into building blocks of the PFR [48-50], but were unable to fully resolve the PFR organization. Using our newly developed script called *Propagate* (Methods), together with the *PEET* program [55, 58], we were able to iteratively identify and refine the paracrystalline lattice parameters of the distal zone, first in one dimension along the axoneme axis and then in all three dimensions (see details in Methods) (Fig. 2). Sub-tomographic averaging of the distal zone using 1362 sub-tomograms from 12 tomograms yielded an averaged 3D structure of the distal zone at ~28.5Å resolution based on Fourier shell correlation (FSC) analysis at the 0.143 coefficient criterion (Supplementary Fig. S4a).

The refined paracrystalline lattice is triclinic (*i.e.*, none of the three lattice angels is orthogonal) with the unit lengths of vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ of 18.0nm, 47.0nm, and 41.5nm and angles of α, β, γ at 76°, 78°, and 79°, respectively (Fig. 2h). The dimension $\vec{a}$ is along the vertical direction in Figs. 2a-e and 2h$_{1, 3, 4}$; that is, $\vec{a}$ is perpendicular to the axoneme axis. Surprisingly, the other two axes of the unit cell are neither along the axoneme axis (dashed cyan line, Fig. 2h$_2$) nor orthogonal to dimension $\vec{a}$. Thus, the unit cell is neither orthorhombic [48], nor helical [44]. Rather, when examining from the top of the axoneme down to the PFR (Figs. 2f, h$_2$), dimension $\vec{b}$ is oriented approximately 45° from the axoneme axis, which may account for the ~45° angle previously observed in negative-stained samples [44, 49]. Dimension $\vec{c}$ is oblique (76°) to dimension $\vec{b}$. The main building block of this plane is what we call density scissors (delineated by the tan surface rendering in Fig. 2c), which is visible when viewing the face of the plane from the perpendicular perspective. This scissors-like building block differs from a previously proposed "jackscrew" model, which was based on viewing the distal zone from the axoneme side of the PFR [50]. Density scissors stack vertically upon one another along dimension $\vec{a}$ (Figs. 2c, h$_3$), with 4-6 scissors per stack. Each stack in turn connects horizontally with adjacent scissors stacks, forming a planar network of stacks, which we term a "scissors stack network" plane (SSN plane), that is oriented 45° to the axoneme axis. Each SSN plane encompasses 5 stacks, 1 with 5 pairs of scissors, 3 with 6 pairs of scissors, and one near the boundary of the PFR on DMT4 side with 4 pairs of scissors but with densities exhibiting differences from the other 4 stacks. Thus, a 20μm long PFR would include ~370 SSN planes with ~27 pairs of scissors per plane.

Extending between SSN planes are thin densities, which we call wires. When viewed parallel to SSN planes (Figs. 2d, e; Supplementary Movie S3), wires from adjacent planes appear to extend contiguously to form a smile-like arc (0.8 radians) that spans four SSN planes, with two end segments (wires 1 and 5) and three middle segments (wires 2 – 4). When viewed from above (Fig. 2f), the end segments of each arc deviate slightly from a straight line formed by the three middle segments. Wire 1 from one arc abuts wire 5 of an adjacent arc and their structures appear

175

distinct (Fig. 2f). When the density threshold is lowered, wires 1 and 5 appear to interact (Figs. 2f, g; Supplementary Movie S3). In addition, near the interface of the distal and intermediate zones, wires 1 and 5 extend to make direct contact with densities in the intermediate zone (Fig. 2g; Supplementary Movie S4; see also Figs. 4b, c). Therefore, in the structure examined, the distal zone is a remarkable and intricate 3D nanoscale crystal consisting of many SSN planes aligned 45° to the axoneme axis and joined together by density wires, with ~90% empty space when calculated using a threshold as in Fig. 2f.

To interpret structures constituting the unit cell of the distal zone, we examined elements that make up each SSN plane (same color in Fig. 3a) and connect with neighboring SSN planes (different colors in Fig. 3a). We segmented these elements in such a way that each scissors density and its connecting wires remain together (Figs. 3b, c). These structural elements can be brought together within a single unit cell by translation of one unit length along the vectors of the triclinic unit cell (Figs. 3d-f). Fig. 3b is the view looking perpendicularly at the SSN plane and highlights one scissors stack in blue, with a single pair of scissors density colored tan. The handles of the scissors are at the bottom with curved blades projecting toward the top (Figs. 3b, e). Rotating this view 90 degrees (Fig. 3c) reveals the wire densities that extend between adjacent SSN planes described above. Figs. 3d-f show three orthogonal views of the structural elements that comprise a single unit cell, with dimensions of the unit cell (as defined in Fig. 2h) indicated with vectors and wires 1-5 labeled.

Current resolution of the PFR structure precludes identification of molecules that make up the scissors densities and wires, but the extended rod-shape of wires is consistent with structures exhibited by coiled-coil helices. Biochemical analyses indicate that major structural components of the PFR are two proteins, PFR1 and PFR2 [59, 60]. Deletion mutants of either PFR1 or PFR2 failed to assemble a complete PFR structure [21, 32, 61, 62], demonstrating that, despite sharing high-percentage amino acid identity, PFR1 and PFR2 are essential and non-redundant components of the PFR. In some cases, a rudimentary proximal domain is retained [21, 32, 62]. Secondary structure predictions of

PFR1 and PFR2 from *T. brucei* and homologs from other *kinetoplastids* showed that more than 80% of the protein is predicted to form helices (Fig. 3g) [60]. The N-terminal half contains helices of short lengths, followed by unusually long helices. The longest predicted helices of PFR1 and PFR2 contain 258 amino acids and 222 amino acids, respectively, which correspond to a length of 41.3nm and 35.5nm, respectively (~1.6 Å/amino acid in an alpha helix). In our sub-tomographic averaged structure, the lengths of individual rod-shaped density wires range from 40nm to 45nm. The diameter of the wire, ~4nm, would accommodate a coiled coil of 3-4 helices (Figs. 3h-j), suggesting that multiple subunits of PFR1 and PFR2 molecules could contribute to each arm of the wire. The shorter helices and coiled-coil sequences predicted could contribute to the formation of the globular region of the scissors densities. Additional PFR components have been identified [42], including PFC3 and PAR1 that are predicted to assemble into extended coiled-coil structures, and we expect these may also contribute to wires or other PFR structures. However, unlike PFR1 and 2, RNAi knockdown of PFC3 or PAR1 does not noticeably affect PFR ultrastructure [63], indicating they are not required for assembly of the main PFR structural elements.

5.3.3 Contiguous overlapping layers in the proximal zone and flexible linkages in the intermediate zone

The organizations of the PFR proximal and intermediate zones remain a mystery, likely due to difficulties in sub-tomographic averaging caused by lack of knowledge about the periodicity, flexible nature of the structure, and the large volume to resolve. In 3D reconstructed tomograms after missing-wedge compensation, one can readily discern multiple densities in the intermediate zone, consistent with TEM from thin cross-section of embedded trypanosome flagella [49] (Figs. 4a, b). These densities connect to wires 1 and 5 of the distal zone (Fig. 4b, orange arrows), providing a direct link between structurally distinct PFR zones. Sub-tomographic averaging did not improve resolution of these densities (Fig. 4c), suggesting that they might be flexible or present polymorphic features.

In the sub-tomographic average of the proximal zone, the overall cross section contour approximates a right-angle trapezoid with the right-angle side corresponding to the DMT7 side of the axoneme (Fig. 4d). The approximate dimensions of this trapezoid are base lengths of 210nm and 70nm, and height of 87nm. Unlike the clearly resolved density elements in the distal zone described above, proximal zone density elements are convoluted and hard to distinguish from one another. Nonetheless, we were able to segment these densities into four layers when viewed in cross section, by following the gaps visible as shown in Fig. 4d. Three of these layers are readily visible in longitudinal views from outside the PFR (pink, blue and light purple in Fig. 4e; Supplementary Movie S5). When viewed from the axoneme looking toward the PFR, the light purple layer is dominated by an elongated density extending perpendicular to the PFR axis and spanning from DMT7 to DMT4 of the axoneme. This density measures ~210nm long and ~40nm wide, with thickness up to 30nm on the DMT4 side. This density repeats along the axoneme axis every 54nm (Fig. 4f, g), which is the same interval observed for the distal zone repeat. The other proximal zone layers (pink, blue, green) are best visualized when viewed from the PFR distal zone looking toward the axoneme, which allows the visualization of all four layers of the PFR proximal zone (Fig. 4g). Diagonal densities join adjacent repeats within each layer and form interconnections between layers (Figs. 4h, i; Supplementary Movie S5). These densities form lattice-like networks, when examined from the axoneme interface looking toward the PFR (Figs. 4h, i; Supplementary Movie S5). Though different from one another, these networks and that in the distal zone (Fig. 4j) appear to be congruent with the rectangular, 36nm by 40nm repeating units, observed in the green layer of proximal zone (Fig. 4j), This congruence suggests presence of organizational linkages extending from the distal zone into the intermediate and proximal zones, and a mechanism for mediating continuity between the linear repeat of the axoneme with the diagonal repeat of the PFR distal zone.

5.3.4 PAC structures bridge different repeats of the PFR and axoneme

A central question about *T. brucei* flagellum biology concerns the mechanism of PFR attachment to the axoneme at DMT4-7. Our finding that the PFR and DMT repeating unit dimensions are mismatched, 54 versus 8nm, makes this a particularly challenging problem. Filaments connecting the PFR to DMT7 have been described [47, 49, 64], but the 3D arrangement and structural details were limited and connections to DMT4, 5, and 6 are almost completely uncharacterized. We therefore performed sub-tomographic averaging of local volumes at the PFR-axoneme interface to resolve individual structures of connections at DMT4-7, which we term PFR-axoneme connectors (PACs) (Fig. 5a).

Prior studies describe PAC7 as filamentous connections between the PFR and DMT7 [47, 65]. Our structure reveals PAC7 is actually comprised of four components (Figs. 5b$_2$-b$_5$): a PFR-proximal baseplate (~30nm in diameter), a peri-axonemal plate (PAP, approximately 42 by 20 by 12nm in size), filaments (~35nm in length) connecting the baseplate and PAP, and small globular densities that connect the PAP to the microtubule lattice of the DMT7 B-tubule at protofilaments B2-4 (Supplementary Fig. S3). There are typically three connecting filaments per baseplate, with two parallel filaments connecting to one PAP and the third extending from the baseplate to contact the PAP of an adjacent PAC7. Positioning of connecting filaments along the PFR-axoneme interface is not entirely uniform (Figs. 5b$_2$, b$_6$) and structural details of these filaments become smeared in the sub-tomographic average of PAC7 (Figs. 5b$_3$, b$_4$). The PAC7 baseplate exhibits a periodicity of 54nm, consistent with the PFR repeating unit. However, the PAP does not repeat in an entirely regular fashion (Figs. 5b$_2$, b$_6$), likely reflecting the need to interface the different repeats of the PFR (54nm) and axoneme (96nm). Thus, our cryoET structure resolves the structural basis for bridging the distinct repeating unit dimensions of two megastructures, the PFR and axoneme, which must act together to support unique motility of the trypanosome cell.

PAC5 includes three separate connections: PAC5-1, PAC5-2, and PAC5-3 (Fig. 5c; Supplementary Figs. S3d, h, I), which connect to OAD, and protofilaments B1, 2 and B5 of DMT5, respectively (Fig. 5c$_1$; Supplementary Fig. S3). These results support an earlier report that PFR

may influence motility through direct interaction with dynein [49]. In longitudinal views, PAC5-1, 5-2 and 5-3 each include two densities spaced 54nm apart.

PAC4, when viewed in cross-section, is a large globular density that forms a U-like structure at the axoneme, making separate contacts with the A and B-tubules of DMT4 at protofilaments A9, 10 and B1, 3 (Fig. 5d$_1$; Supplementary Fig. S3). PAC4 has a periodicity of 54nm along the flagellum (Fig. 5d$_2$), consistent with the repeating unit of the PFR. Because 216 is the least common multiple of 8, 24 and 54, a structural unit encompassing the PFR and its contact points on the axoneme, α/β-tubulin dimer, and OAD, would repeat each 216nm (Fig. 5d$_3$), while a unit that also includes the 96nm repeat of RS-IAD-NDRC would repeat each 864nm.

We did not observe obvious densities for PAC6 in the PFR sub-tomographic average. However, in the axoneme sub-tomographic average we did resolve a novel "microtubule outer protein" (MOP) on DMT6 (orange-colored in Fig. 5e), which we term DMT6-MOPB-1,2, because it is attached to the DMT6 B-tubule protofilaments 1 and 2 (Fig. 5e; Supplementary Fig. S3). DMT6-MOPB-1,2 has a periodicity of 8nm, consistent with the periodicity of the α/β-tubulin dimer.

5.3.5 Structure of the T. brucei central pair complex

The CPC (Fig. 6a) is an essential regulator of axoneme motility, functioning with the RS to transmit mechanochemical signals across the axoneme [27, 54, 66]. Importance of CPC function is evidenced by several human diseases associated with CPC abnormalities [67] and a requirement of CPC proteins for *T. brucei* motility [34, 68]. We therefore determined the 3D structure of the *T. brucei* CPC *in situ*. The CPC is visible in individual tomograms (Figs. 1b, 6b; Supplementary Fig. S2). Sub-tomographic averaging at 96nm periodicity shows densities protruding outward from the central pair microtubules (Fig. 6c) and diagonally arranged densities of the tripartite bridge [69] between C1 and C2 microtubules (Fig. 6d). Major densities repeated at an interval of 16nm (Figs. 6c, d). We therefore did sub-tomographic averaging at 16nm, yielding a 25Å resolution sub-tomographic average structure (Fig. 6e; Supplementary Fig. S4b and Supplementary Movie S6), in which we

resolved the C1 and C2 microtubules, as well as densities corresponding to 11 projections and tripartite bridge between C1 and C2 described for the *Chlamydomonas* and *Strongylocentrotus* CPC [69] (Fig. 6e). Interfacing with RS is critical for CPC function [24, 25, 27] and extensive contacts between RS and CPC projections are observed in the *T. brucei* structure (Figs. 6b, 6f-h; Supplementary Fig. S2). These extensive contacts may contribute to restricted orientations of the CPC relative to the DMTs in *T. brucei* [34, 68, 70]. Our results provide the first 3D structure for the CPC in trypanosomes, revealing overall conserved features and extensive direct contacts with the RS.

5.4 Discussion

In this study, we have used a combination of cryoET, machine learning-based missing wedge compensation, and sub-tomographic averaging to resolve the previously unknown molecular structures of the PFR, CPC, and PACs in *T. brucei*. The structures reported here provide insights into flagellar motility and mechanical bionics.

To date, studies of structural mechanisms underpinning flagellum beating have focused almost exclusively on organisms with planar axoneme beating, and have mostly ignored extra-axonemal filaments [23]. However, helical waves are common among microbes, including important pathogens [5, 8, 36, 71], and even occur in human sperm [72]. Non-planar helical motion may contribute to microbial pathogenesis [12], as it is recognized to facilitate propulsion through viscous environments [73], such as host tissues. Meanwhile, extra-axonemal filaments are common features of eukaryotic flagella [16, 17, 19]. Therefore, a full understanding of biomechanics of cell propulsion requires structural analysis of flagella from organisms that support helical motion, such as *T. brucei*, and requires analysis of extra-axonemal structures, such as the PFR.

For the sake of illustration, helical waves can be decomposed into x and y oscillations each described by a sinusoidal function $H(t) = A \cdot sin(\omega t) \cdot \vec{\imath} + A \cdot cos(\omega t) \cdot \vec{\jmath} + v \cdot t \cdot \vec{k}$ where A, ω, v are amplitude, rotational speed and forward speed respectively, and $\vec{\imath}, \vec{\jmath}, \vec{k}$ are orthogonal unit vectors in the x, y and z axis, respectively (Fig. 7a). In addition, since points along the axoneme filament

are connected, we must also consider twist introduced by helical waves *(i.e.*, the situation for A > 0 in Fig. 7a). Therefore, the PFR of *T. brucei* must provide elastic bending resistance [21] while being flexible enough to support the axoneme as it executes complex helical motion [5, 36, 74, 75]. SSN planes placed along the axoneme at about 45° to the axoneme axis offers an excellent solution to these competing needs (Fig. 7b). We propose that the two-dimensional SSN planes provide a rigid component for support and resistance, while connection of these planes in the third dimension by coiled-coil helix bundles (*i.e.*, wires in Fig. 3) provides elasticity. Such organization imparts integrity to the axoneme, yet still allows sinusoidal oscillation in both x and y directions and the 45° orientation of SSN planes also provides elastic resistance and support both across and along axoneme DMTs. By contrast, placing SSN planes at 0° or 90° with respect to the axis of axoneme would prohibit oscillation in the y direction (Fig. 7c), or lend less support along axoneme DMTs (Fig. 7d), respectively. A 45° orientation may also contribute to the helical bending of the axoneme.

A distal zone comprised of SSN planes interconnected by wire-like densities differs from a previously proposed "jackscrew" model for the PFR distal zone [50]. The jackscrew model was based on viewing the distal zone from the axoneme toward the PFR, which revealed two sets of linear densities intersecting diagonally to form a lattice of diamond shapes (Figs. 3b, d, f in reference [48]). Viewing from this perspective, we also observe a lattice-like arrangement of diagonally intersecting densities (Fig. 4j). However, as noted previously [48], there is no density connecting opposing vertices of each diamond shape, i.e. no screw for a jackscrew (Fig. 4j). Moreover, the earlier work did not resolve the structural units that comprise each diagonal density in the lattice, *i.e.* stacked scissors of SSN planes and wire densities that connect them as described here.

For the PFR to fulfill motility functions in unison with the axoneme, these two massive structures must be interconnected in a way that enables mechanochemical signals to be transmitted to all participating components. Such signals may go both ways, either from axoneme to PFR or vice versa, as the PFR is a platform for Ca++ and cAMP signaling systems [40, 42]. The observed structural organization of connections within and between the axoneme and PFR is

consistent with such needs. Within the axoneme, signaling between DMTs is supported by nexin links that connect adjacent DMTs [76]. Our prior studies [26] demonstrate capacity for DMT-DMT signaling in *T. brucei*, by confirming the presence of nexin links, and identifying novel, lineage-specific connections between adjacent DMTs [26]. In the present study, our observation of direct interaction between RS and CPC (Figs. 6b, f, g; Supplementary Fig. S2), is likewise consistent with the role of the RS in transmitting signals through the CPC to DMTs across the axoneme reported in other organisms [27, 54, 66]. Extensive contact between RS and CPC may also contribute to limited rotation observed for the *T. brucei* CPC relative to DMTs [34, 68, 70]. Signals from the axoneme may in turn be transmitted directly to the PFR proximal zone through the PACs. Contiguous structural connections extending through the proximal, intermediate, and distal zones (Figs. 4b, c, h-j; Supplementary Movie S2, S4, S5) allow these signals to be propagated all the way to the distal zone, which provides a highly organized spring-like structure that can store and release mechanical energy.

Structural details of PFR connections to the axoneme provide insight into how the PFR may influence axoneme beating. The mismatch in periodicity of the axoneme (8nm) and the PFR (54nm) is reminiscent of symmetry mismatch often observed in structures involved in dynamic biological processes, such as those of the DNA translocation portals in viruses [77, 78] Dynamic interaction is also supported by the observed heterogeneity in spacing of the PAC7 - DMT7 interface (Fig. 5b), as this heterogeneity in spacing suggests capacity for sliding of PAC7 on this side of the axoneme. In addition, PAC4 and 5 appear to have more regular longitudinal spacing, and such an arrangement could support or even amplify helical bending, as it would present more resistance to bending on the DMT4 versus DMT7 side of the axoneme. A mismatch in periodicity also suggests that, although PFR assembly is coordinated with and dependent on axoneme assembly [79, 80], PFR assembly is not templated by the axoneme. This interpretation is consistent with RNAi knockdown studies showing proteins that are not part of the PFR are nonetheless required for proper PFR assembly [81-83], indicating that PFR assembly is a multistep, well-controlled

process with assembly steps in the cytoplasm as well as flagellum. Finally, direct connection of PAC5-1 to dynein motors on DMT5 (Figs. 5c$_1$, c$_2$) [49] will exert substantial influence on dynein orientation. Because reorientation of axonemal dynein during the beat cycle is a major mechanism of axonemal beat regulation [23, 24, 84], PAC5-1 provides a mechanism for the PFR to directly control axonemal beating.

In summary, our studies provide a high resolution 3D description of an extra-axonemal structure, giving insight into how these common yet enigmatic components of eukaryotic flagella contribute to axonemal beating. From the standpoint of bionics, with less than 10% space filled, the PFR of trypanosomes may provide an example of cellular tensegrity—biological entities that embody a fine balance between strength and flexibility, owing to opposing forces of compression or tension—and should inform future bioengineering and mechanical design of nanomachines and microswimmers [85, 86].

5.5 Materials and Methods

5.5.1 Sample preparation and cryoET

T. brucei bloodstream form single marker (BSSM) cells [87] were used, and details for culturing, flagella isolation, and cryoET were described previously [26]. Briefly, demembraned flagella were isolated and vitrified on quantifoil grids with 5nm gold particles. With *SerialEM* [88], tilt series were collected from straight segments near the center part of full-length flagella, spanning the middle third between the basal body and tip, in a Titan Krios instrument equipped with a VPP, a Gatan imaging filter (GIF) and a post-GIF K2 direct electron detector in electron-counting mode. Frames in each movie of the raw tilt series were aligned, drift-corrected, and averaged with *Motioncorr* [89]. The tilt series micrographs were aligned and reconstructed into 3D tomograms by either weighted back projections (WBP, for sub-tomographic averaging) or simultaneous iterative reconstruction technique (SIRT, for visualization and particle picking) using the *IMOD* software package [90]. The contrast transfer function (CTF) was determined by *ctffind4* [91] and corrected with the *ctfphaseflip*

program [92] of *IMOD*. With phase plate, the CTF is insensitive to the sign of the defocus value being negative (under-focus) or positive (over-focus) [93] so CTF of micrographs obtained with phase plate were approximated when CTF rings were not readily detected.

5.5.2 Missing-wedge compensation

Demembranated *T. brucei* flagella samples typically lie on the cryoEM grids with a preferred orientation due to the presence of the PFR. To alleviate the missing-wedge problem associated with preferred orientation, we used a novel deep learning-based method developed to compensate for missing-wedge problem (Liu et al., unpublished program). Using tilt geometry and resulting tomograms as inputs, the program iteratively learns how to fill in missing information.

5.5.3 Sub-tomographic averaging

In our sub-tomographic averaging scheme performed using *PEET*, each particle is a 3D sub-volume of the tomogram corresponding to the repeating unit of the component of interest, *i.e.*, PFR distal zone, PFR proximal zones, PACs, or CPC.

PFR distal zone.   For the PFR distal zone, sub-tomographic averaging required first defining the repeating unit through multiple rounds of *PEET* trials, and then using this defined repeating unit to re-pick particles automatically for the final round of *PEET* refinement, leading to a final sub-tomographic average at the best possible resolution.

To define the repeating unit *a priori*, we first performed crude sub-tomographic averaging with 4x binned (resulting a pixel size of 10.2Å) SIRT tomograms. Initially, we manually picked particles (190x190x190 pixels) by taking advantage of the repeating comb teeth visible in the PFR distal zone in the raw tomograms (*e.g.*, Figs. 1b, 2b, cyan arrows). For each particle, this process records the x, y, z coordinates of two points in the tomogram, one at each end of a tooth. The coordinates for all picked particles in each tomogram were saved into a coordinates model .mod file. With this model file as the input, we then ran *stalkInit* of the *PEET* package to generate three output files—motive list .csv file (translation and rotation parameters), coordinates model .mod file

(central coordinates of particles) and rotation axes .csv file (vectors representing rotation axes for all particles)—for each tomogram. All these picked particles were summed together to generate a featureless volume, which was used as the initial reference of the first cycle of *PEET*. We then ran *PEET* iteratively by gradually decreasing the search range for both *angular* and *distance* parameters. For the *angular* search range, the parameters decreased from "180° max with 60° step in *Phi* (y axis), and 9° max with 3° step in both *Theta* (z axis) and *Psi* (x axis)", to "3° max with 1° step in all *Phi* (y axis), *Theta* (z axis) and *Psi* (x axis)". For the *distance* search range, the parameters decreased from 10 pixels to 2 pixels along all three axes. In addition to decreasing the search ranges, the reference was also updated by using the result of the previous round of *PEET* sub-tomographic averaging. This process was iterated until the averaged structure converged, and no improvement in the averaged structure could be observed. In this converged average, repeating densities, characteristic of those ultimately resolved in the lattice in Figs. 4j, were observed.

The repeating unit parameters defined above were utilized to re-pick particles following the 3D lattice. To do this for many tomograms automatically, we developed a bash script, *propagate.sh,* and used it to pick particles encompassing all repeating units within each tomogram, taking advantage of the unit cell dimensions of the above-observed lattice in all three orthogonal directions. At this stage, the sub-tomographic average volume resulting from the above initial *PEET* process contains more than one repeating unit. Though our script allows multiple repeating units in three dimensions, in the current case, we only needed to pick up all repeating units in the two dimensions, *i.e.*, within an SSN plane, because the above manually picked particles cover only one unit-cell length along axoneme axis. Using the 3D visualization tools of *IMOD*, we identified 22 repeating units within each SSN plane. We carefully measured the x, y, and z distances between the center position of the current average volume and the center of each of the 22 repeating units and list each set of x, y, z distances in separate lines in a propagation list file. The script *propagate.sh* takes as inputs the follwoing: the propagation list file, and a set of three alignment files—including a motive list file (translation and rotation parameters), a coordinates model file

(central coordinates of particles) and a rotation axes file (vectors representing rotation axes for all particles) for each tomogram—generated from the above *PEET* round. The output includes a single *prm* file (file pathway information) for the entire project and a new set of three alignment files for each tomogram to be used as the input files for the next round of *PEET*. This re-picking process ended up with 22 times as many particles as in the initial manual picking process.

With these re-picked particles, one round of iterative *PEET* refinement was performed by loading the new alignment and *prm* files to generate the final sub-tomographic average for the PFR distal zone. As before, the search ranges were gradually decreased and the average result from each prior refinement cycle was used as the updated reference for the subsequent cycle of refinement. The refinement was terminated when the averaged structure converged, and no further improvement could be seen. From the *unMasked* averaged density map file automatically generated by PEET during the refinement, the repeating length along the axoneme axis was measured to be 54nm. Unit cell dimensions were also measured as reported as the final parameters in Fig. 2h.

PFR proximal and intermediate zones, and PACs. For sub-tomographic averaging of PFR proximal and intermediate zones, and PACs, the method is the same as that of the PFR distal zone, except that there was no need to run *propagate* because the repeating unit is only one dimensional. Instead, we identified prominent repeated densities from raw tomograms to be used for manual particle picking. For example, we used the baseplate density attached with 3 filaments (Fig. 5b) for manual picking PAC7. After iterative sub-tomographic averaging for the component of interest, the *unMasked* averaged density map file automatically generated by *PEET* was used to measure the dimension of a repeating unit.

CPC. To identify the periodicity of CPC, we first evaluated the tomograms and did a trial of sub-tomographic averaging using 96nm periodicity (Figs. 6c, d). One could readily recognize 16nm repeated features including the different projection densities. Therefore, we next utilized the 16nm periodicity to box particles of the CPC along the axoneme axis. These sub-

volumes centered on the CPC were then aligned to each other, and averaged together with *PEET*, with the procedures described above.

We used the following numbers of particles to generate sub-tomographic averages of various flagellar components: 763 particles from 10 tomograms for DMT, 3001 particles from 52 tomograms for the 16nm CPC, 558 particles from 52 tomograms for 96nm CPC, 246 particles from 17 tomograms for each of the PAC4, PAC5, PAC6 and PAC7, 246 particles from 17 tomograms for the PFR proximal zone, 1362 particles from 12 tomograms for the PFR distal zone.

The resolution of each sub-tomographic average was calculated by *calcFSC* in *PEET* based on the 0.143 FSC criterion (Supplementary Fig. S4).

5.5.4 Sequence alignment and secondary structure prediction of major PFR proteins

For PFR protein sequence alignments, we used PFR1 (NCBI# XP_844025.1) and PFR2 (NCBI# XP_847331.1) from *T. brucei* [59]*,* PFR-like protein of *Leishmania braziliensis* (NCBI# XP_001565953) [94], PFR1 (NCBI #AAV53924) of *Angomonas deanei* [95] from the *Strigomonadinae* family, and PFR 1D (NCBI XP_003872382.1) [96] of *Leishmania mexicana* from the *Leishmaniinae* subfamily. Sequences were aligned with multiple sequence alignment function in *Clustal Omega* (1.2.4) [97]. For each of the four proteins, secondary structures were predicted using *PSIPRED* [98]. Predicted secondary structures of proteins were combined with protein alignment results to compare the PFR protein structure between different species.

5.5.5 3D visualization

IMOD [90] and UCSF *ChimeraX* [99] were used to visualize reconstructed tomograms and sub-tomographic averages. Segmentation of densities maps was performed by the *volume tracer* and *segger* tools of UCSF *Chimera* [100]*.* For surface rendering with UCSF *ChimeraX*, maps were first low-pass filtered to either 30Å or 50Å. Montage was done in UCSF *ChimeraX* by fitting averaged structures (*i.e.,* 96nm averaged axoneme, 16nm averaged CPC, 54nm averaged PACs, 54nm

averaged PFR proximal zone, etc.) into an unaveraged tomogram after missing-wedge

compensation. Fitting of existing model of coiled-coil helix bundle was done with the *molmap* and *fit*

functions in UCSF *ChimeraX* [99]. Schematics were drawn by *Adobe Illustrator.* The built-in

denoising program in *Warp* [101] was used to enhance visualization. Tilt series were separated into

two sets and reconstructed independently, and the noise will be filtered out to improve the SNR

given that the signal is consistent in both maps, but the noise is random.

5.6 Acknowledgments

5.7 Author contributions

Z.H.Z. and K.L.H. designed the project; S.I. and J.Z. cultured and isolated samples; J.Z., I.A.,

H.H.W.. S.I. and H.W. performed cryoET imaging; J.Z. and H.W. led the data processing and

visualization with assistance from X.Z., S.L. and S.I.; X.Z., Z.H.Z., J.Z., H.W., and K.L.H. performed

sequence analysis and structure prediction; J.Z., Z.H.Z., H.W., and K.L.H. interpreted the results

and wrote the paper; all authors reviewed and approved the paper.

5.8 Conflict of interest

The authors declare that they have no conflict of interest.

5.9 Data Availability

The cryoET sub-tomographic averaged maps have been deposited in the EM Data Bank under the

accession codes EMD-23619 and EMD-23620 for the PFR distal zone and the central pair complex

in *T. brucei,* respectively. A tomogram of the *T. brucei* flagellum in its bloodstream form was deposited in the EM Data Bank under the accession code EMD-23621.

5.10 References

1.   Bloodgood, R.A. Sensory reception is an attribute of both primary cilia and motile cilia. *J Cell Sci* **123**, 505-509 (2010).

2.   Engle, S.E., Bansal, R., Antonellis, P.J. & Berbari, N.F. Cilia signaling and obesity. *Semin Cell Dev Biol* (2020).

3.   Gabriel, G.C., Young, C.B. & Lo, C.W. Role of cilia in the pathogenesis of congenital heart disease. *Semin Cell Dev Biol* (2020).

4.   Wallmeier, J. *et al.* Motile ciliopathies. *Nat Rev Dis Primers* **6**, 77 (2020).

5.   Heddergott, N. *et al.* Trypanosome motion represents an adaptation to the crowded environment of the vertebrate bloodstream. *PLoS Pathog* **8**, e1003023 (2012).

6.   Jennings, H.S. Studies on reactions to stimuli in unicellular organisms. II. - The mechanism of the motor reeactions of paramecium. *American Journal of Physiology* **II**, 311-341 (1899).

7.   Schuster, S. *et al.* Developmental adaptations of trypanosome motility to the tsetse fly host environments unravel a multifaceted in vivo microswimmer system. *Elife* **6** (2017).

8.   Rossi, M., Cicconofri, G., Beran, A., Noselli, G. & DeSimone, A. Kinematics of flagellar swimming in Euglena gracilis: Helical trajectories and flagellar shapes. *Proc Natl Acad Sci U S A* **114**, 13085-13090 (2017).

9.   Snell, W. & Goodenough, U. Flagellar Adhesion, Flagellar-Generated Signaling, and Gamete Fusion during Mating, in *THe Chlamydomonas Sourcebook*, Edn. 2nd Edition (2009).

10.   Medina, E.M. *et al.* Genetic transformation of Spizellomyces punctatus, a resource for studying chytrid biology and evolutionary cell biology. *Elife* **9** (2020).

11.   Rotureau, B., Ooi, C.P., Huet, D., Perrot, S. & Bastin, P. Forward motility is essential for trypanosome infection in the tsetse fly. *Cell Microbiol* **16**, 425-433 (2014).

12.   Shimogawa, M.M. *et al.* Parasite motility is critical for virulence of African trypanosomes. *Sci Rep* **8**, 9122 (2018).

13.   Salmon, D. *et al.* Adenylate cyclases of Trypanosoma brucei inhibit the innate immune response of the host. *Science* **337**, 463-466 (2012).

14.   House, S.A., Richter, D.J., Pham, J.K. & Dawson, S.C. Giardia flagellar motility is not directly required to maintain attachment to surfaces. *PLoS Pathog* **7**, e1002167 (2011).

15.   Smith, E.F. & Rohatgi, R. Cilia 2010: the surprise organelle of the decade. *Sci Signal* **4**, mr1 (2011).

16.   Linck, R.W., Chemes, H. & Albertini, D.F. The axoneme: the propulsive engine of spermatozoa and cilia and associated ciliopathies leading to infertility. *J Assist Reprod Genet* **33**, 141-156 (2016).

17.   Liu, P. *et al.* Chlamydomonas PKD2 organizes mastigonemes, hair-like glycoprotein polymers on cilia. *J Cell Biol* **219** (2020).

18.   Cachon, J., Cachon, M., Cosson, M.-P. & J, C. The paraflagellar rod: a structure in search of a function. *Biol. Cell* **63**, 169-181 (1988).

19.   Vickerman, K. The mechanism of cyclical development in trypanosomes of the Trypanosoma brucei sub-group: an hypothesis based on ultrastructural observations. *Trans R Soc Trop Med Hyg* **56**, 487-495 (1962).

20.   Bastin, P., Sherwin, T. & Gull, K. Paraflagellar rod is vital for trypanosome motility. *Nature* **391**, 548 (1998).

21.   Santrich, C. *et al.* A motility function for the paraflagellar rod of Leishmania parasites revealed by PFR-2 gene knockouts. *Mol Biochem Parasitol* **90**, 95-109 (1997).

22.   Zhao, W. *et al.* Outer dense fibers stabilize the axoneme to maintain sperm motility. *J Cell Mol Med* **22**, 1755-1768 (2018).

23.   Lin, J. & Nicastro, D. Asymmetric distribution and spatial switching of dynein activity generates ciliary motility. *Science* **360** (2018).

24.   Gui, M. *et al.* Structures of radial spokes and associated complexes important for ciliary motility. *Nat Struct Mol Biol* (2020).

25.   Grossman-Haham, I. *et al.* Structure of the radial spoke head and insights into its role in mechanoregulation of ciliary beating. *Nat Struct Mol Biol* **28**, 20-28 (2021).

26.   Imhof, S. *et al.* Cryo electron tomography with Volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen. *Elife* **8** (2019).

27.   Oda, T., Yanagisawa, H., Yagi, T. & Kikkawa, M. Mechanosignaling between central apparatus and radial spokes controls axonemal dynein activity. *J Cell Biol* **204**, 807-819 (2014).

28.   Ichikawa, M. *et al.* Subnanometre-resolution structure of the doublet microtubule reveals new classes of microtubule-associated proteins. *Nat Commun* **8**, 15035 (2017).

29.   Stuart, K. *et al.* Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of clinical investigation* **118**, 1301-1310 (2008).

30.   Farina, M., Attias, M., Soutopadron, T. & Desouza, W. FURTHER-STUDIES ON THE ORGANIZATION OF THE PARAXIAL ROD OF TRYPANOSOMATIDS. *Journal of Protozoology* **33**, 552-557 (1986).

31.   Fuge, H. Electron microscopic studies on the intra-flagellar structures of trypanosomes. *J. of Protozoology* **16**, 460-466 (1969).

32.   Maga, J.A., Sherwin, T., Francis, S., Gull, K. & LeBowitz, J.H. Genetic dissection of the Leishmania paraflagellar rod, a unique flagellar cytoskeleton structure. *J Cell Sci* **112 (Pt 16)**, 2753-2763 (1999).

33.   Smith, D.J., Gaffney, E.A., Gadelha, H., Kapur, N. & Kirkman-Brown, J.C. Bend propagation in the flagella of migrating human sperm, and its modulation by viscosity. *Cell Motil Cytoskeleton* **66**, 220-236 (2009).

34.   Branche, C. *et al.* Conserved and specific functions of axoneme components in trypanosome motility. *J Cell Sci* **119**, 3443-3455 (2006).

35.   Gruby, M. Recherches et observations sur une nouvelle espèce d'hématozoaire, Trypanosoma sanguinis. *Comptes rendus hebdomadaire des séances de l'Académie des Sciences, Paris* **17**, 1134-1136 (1843).

36.   Rodriguez, J.A. *et al.* Propulsion of African trypanosomes is driven by bihelical waves with alternating chirality separated by kinks. *Proc Natl Acad Sci U S A* **106**, 19322-19327 (2009).

37.   Ginger, M.L. *et al.* Calmodulin is required for paraflagellar rod assembly and flagellum-cell body attachment in trypanosomes. *Protist* **164**, 528-540 (2013).

38. Ridgley, E., Webster, P., Patton, C. & Ruben, L. Calmodulin-binding properties of the paraflagellar rod complex from *Trypanosoma brucei*. *Mol Biochem Parasitol* **109**, 195-201. (2000).

39. Shaw, S. *et al.* Flagellar cAMP signaling controls trypanosome progression through host tissues. *Nat Commun* **10**, 803 (2019).

40. Oberholzer, M. *et al.* The Trypanosoma brucei cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. *FASEB J* **21**, 720-731 (2007).

41. Sugrue, P., Hirons, M.R., Adam, J.U. & Holwill, M.E. Flagellar wave reversal in the kinetoplastid flagellate *Crithidia oncopelti*. *Biol Cell* **63**, 127-131 (1988).

42. Portman, N. & Gull, K. The paraflagellar rod of kinetoplastid parasites: from structure to components and function. *Int J Parasitol* **40**, 135-148 (2010).

43. Ginger, M.L. *et al.* Intracellular positioning of isoforms explains an unusually large adenylate kinase gene family in the parasite Trypanosoma brucei. *J Biol Chem* **280**, 11781-11789 (2005).

44. Hyams, J.S. The Euglena paraflagellar rod: structure, relationship to other flagellar components and preliminary biochemical characterization. *J Cell Sci* **55**, 199-210 (1982).

45. Portman, N., Lacomble, S., Thomas, B., McKean, P.G. & Gull, K. Combining RNA interference mutants and comparative proteomics to identify protein components and dependences in a eukaryotic flagellum. *Journal of Biological Chemistry* **284**, 5610-5619 (2009).

46.   Pullen, T.J., Ginger, M.L., Gaskell, S.J. & Gull, K. Protein targeting of an unusual, evolutionarily conserved adenylate kinase to a eukaryotic flagellum. *Mol Biol Cell* **15**, 3257-3265 (2004).

47.   Hemphill, A., Seebeck, T. & Lawson, D. The Trypanosoma brucei cytoskeleton: ultrastructure and localization of microtubule-associated and spectrin-like proteins using quick-freeze, deep-etch, immunogold electron microscopy. *J Struct Biol* **107**, 211-220 (1991).

48.   Höög, J.L., Bouchet-Marquis, C., McIntosh, J.R., Hoenger, A. & Gull, K. Cryo-electron tomography and 3-D analysis of the intact flagellum in Trypanosoma brucei. *J Struct Biol* **178**, 189-198 (2012).

49.   Hughes, L.C., Ralston, K.S., Hill, K.L. & Zhou, Z.H. Three-dimensional structure of the Trypanosome flagellum suggests that the paraflagellar rod functions as a biomechanical spring. *PLoS One* **7**, e25700 (2012).

50.   Koyfman, A.Y. *et al.* Structure of Trypanosoma brucei flagellum accounts for its bihelical motion. *Proc Natl Acad Sci U S A* **108**, 11105-11108 (2011).

51.   Langousis, G. & Hill, K.L. Motility and more: the flagellum of Trypanosoma brucei. *Nat Rev Microbiol* **12**, 505-518 (2014).

52.   Ingber, D.E. The architecture of life. *Sci Am* **278**, 48-57 (1998).

53.   Matias, V.R. & Beveridge, T.J. Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. *J Bacteriol* **188**, 1011-1021 (2006).

54.   Smith, E.F. & Lefebvre, P.A. The role of central apparatus components in flagellar motility and microtubule assembly. *Cell Motil Cytoskeleton* **38**, 1-8 (1997).

55.   Heumann, J.M., Hoenger, A. & Mastronarde, D.N. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J Struct Biol* **175**, 288-299 (2011).

56.   AFZELIUS, B. Electron microscopy of the sperm tail; results obtained with a new fixative. *J Biophys Biochem Cytol* **5**, 269-278 (1959).

57.   Nicastro, D. *et al.* Cryo-electron tomography reveals conserved features of doublet microtubules in flagella. *Proc Natl Acad Sci U S A* **108**, E845-853 (2011).

58.   Nicastro, D. *et al.* The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944-948 (2006).

59.   Deflorin, J., Rudolf, M. & Seebeck, T. The major components of the paraflagellar rod of Trypanosoma brucei are two similar, but distinct proteins which are encoded by two different gene loci. *J Biol Chem* **269**, 28745-28751 (1994).

60.   Schlaeppi, K., Deflorin, J. & Seebeck, T. The major component of the paraflagellar rod of Trypanosoma brucei is a helical protein that is encoded by two identical, tandemly linked genes. *J. Cell Biol.* **109**, 1695-1709 (1989).

61.   Lander, N., Li, Z.H., Niyogi, S. & Docampo, R. CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment. *mBio* **6**, e01012 (2015).

62.   Bastin, P., Pullen, T.J., Sherwin, T. & Gull, K. Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant snl-1. *J Cell Sci* **112 (Pt 21)**, 3769-3777 (1999).

63.   Lacomble, S., Portman, N. & Gull, K. A protein-protein interaction map of the Trypanosoma brucei paraflagellar rod. *PLoS One* **4**, e7685 (2009).

64. de Souza, W. & Souto-Padron, T. The paraxial structure of the flagellum of trypanosomatidae. *J Parasitol* **66**, 229-236 (1980).

65. Hemphill, A., Lawson, D. & Seebeck, T. The cytoskeletal architecture of Trypanosoma brucei. *J. Parasitol.* **77**, 603-612 (1991).

66. Smith, E.F. Regulation of flagellar dynein by the axonemal central apparatus. *Cell Motil Cytoskeleton* **52**, 33-42 (2002).

67. Shoemark, A. *et al.* PCD Detect: enhancing ciliary features through image averaging and classification. *Am J Physiol Lung Cell Mol Physiol* **319**, L1048-L1060 (2020).

68. Ralston, K.S., Lerner, A.G., Diener, D.R. & Hill, K.L. Flagellar motility contributes to cytokinesis in Trypanosoma brucei and is modulated by an evolutionarily conserved dynein regulatory system. *Eukaryot Cell* **5**, 696-711 (2006).

69. Carbajal-Gonzalez, B.I. *et al.* Conserved structural motifs in the central pair complex of eukaryotic flagella. *Cytoskeleton (Hoboken)* **70**, 101-120 (2013).

70. Gadelha, C., Wickstead, B., McKean, P.G. & Gull, K. Basal body and flagellum mutants reveal a rotational constraint of the central pair microtubules in the axonemes of trypanosomes. *J. Cell Sci.* **119**, 2405-2413 (2006).

71. Wilson, L.G., Carter, L.M. & Reece, S.E. High-speed holographic microscopy of malaria parasites reveals ambidextrous flagellar waveforms. *Proc Natl Acad Sci U S A* **110**, 18769-18774 (2013).

72. Su, T.W., Xue, L. & Ozcan, A. High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories. *Proc Natl Acad Sci U S A* **109**, 16018-16022 (2012).

73.   Jahn, T.L. & Bovee, E.C. Locomotion of Blood Protists, in *Infectious Blood Diseases of Man and Animals*, Vol. 1. (eds. D. Weinman & M. Ristic) 393-436 (Academic Press, NY; 1968).

74.   Alizadehrad, D., Kruger, T., Engstler, M. & Stark, H. Simulating the complex cell design of Trypanosoma brucei and its motility. *PLoS Comput Biol* **11**, e1003967 (2015).

75.   Wheeler, R.J. Use of chiral cell shape to ensure highly directional swimming in trypanosomes. *PLoS Comput Biol* **13**, e1005353 (2017).

76.   Heuser, T., Raytchev, M., Krell, J., Porter, M.E. & Nicastro, D. The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. *J Cell Biol* **187**, 921-933 (2009).

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

78.   Fang, Q. *et al.* Structural morphing in a symmetry-mismatched viral vertex. *Nat Commun* **11**, 1713 (2020).

79.   Absalon, S. *et al.* Intraflagellar transport and functional analysis of genes required for flagellum formation in trypanosomes. *Mol Biol Cell* **19**, 929-944 (2008).

80.   Kohl, L., Sherwin, T. & Gull, K. Assembly of the paraflagellar rod and the flagellum attachment zone complex during the Trypanosoma brucei cell cycle. *J Eukaryot Microbiol* **46**, 105-109 (1999).

81.   Demonchy, R. *et al.* Kinesin 9 family members perform separate functions in the trypanosome flagellum. *The Journal of cell biology* **187**, 615-622 (2009).

82. Alves, A.A. *et al.* Control of assembly of extra-axonemal structures: the paraflagellar rod of trypanosomes. *J Cell Sci* **133** (2020).

83. Harmer, J. *et al.* A centriolar FGR1 oncogene partner-like protein required for paraflagellar rod assembly, but not axoneme assembly in African trypanosomes. *Open Biol* **8** (2018).

84. Satir, P. & Matsuoka, T. Splitting the ciliary axoneme: implications for a "switch-point" model of dynein arm activity in ciliary motion. *Cell Motil Cytoskeleton* **14**, 345-358 (1989).

85. Dreyfus, R. *et al.* Microscopic artificial swimmers. *Nature* **437**, 862-865 (2005).

86. Huang, H.W. *et al.* Adaptive locomotion of artificial microswimmers. *Sci Adv* **5**, eaau1532 (2019).

87. Wirtz, E., Leal, S., Ochatt, C. & Cross, G.A. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **99**, 89-101 (1999).

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

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

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

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

92. Xiong, Q., Morphew, M.K., Schwartz, C.L., Hoenger, A.H. & Mastronarde, D.N. CTF determination and correction for low dose tomographic tilt series. *J Struct Biol* **168**, 378-387 (2009).

93. Fan, X. *et al.* Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM. *Structure* **25**, 1623-1630 e1623 (2017).

94. Peacock, C.S. *et al.* Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nat Genet* **39**, 839-847 (2007).

95. Gadelha, C., Wickstead, B., de Souza, W., Gull, K. & Cunha-e-Silva, N. Cryptic paraflagellar rod in endosymbiont-containing kinetoplastid protozoa. *Eukaryot Cell* **4**, 516-525 (2005).

96. Rogers, M.B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Res* **21**, 2129-2142 (2011).

97. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

98. Buchan, D.W.A. & Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* **47**, W402-W407 (2019).

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

100. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

101. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019).

102. Donelson, J.E. & Turner, M.J. How the trypanosome changes its coat. *Sci Am* **252**, 44-51 (1985).

103. Dietrich, M.H. *et al.* Structural and Functional Features of the Reovirus sigma1 Tail. *J Virol* **92** (2018).

5.11 Figures
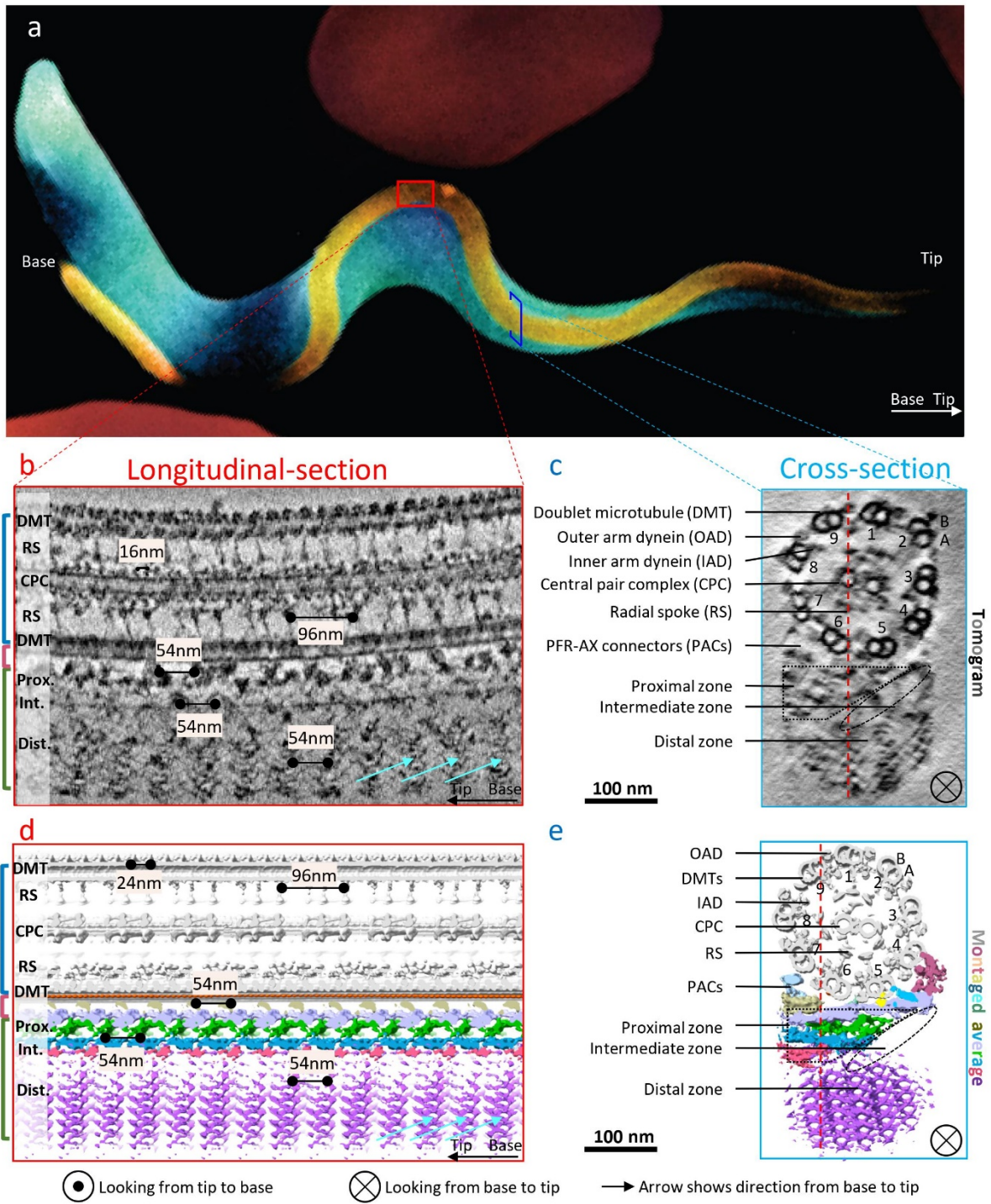
Figure 5-1. CryoET of *T. brucei* flagellum in its bloodstream form.

(a) A scanning EM image of a trypanosome cell (blue) with flagellum (gold) adjacent to red blood cells (red). Artistic rendering based on [102].

(b, c) Longitudinal (b) and cross (c) sections density slices (10nm thickness) of a tomogram of the

T. brucei axoneme and PFR. Arrows in (b) point to apparent "comb teeth" features in the PFR distal zone.

(d, e) Shaded surface views of longitudinal (d) and cross (e) sections of a 1248nm portion of the *T. brucei* axoneme and PFR, obtained by montaging sub-tomographic averages of the axoneme (96nm repeat), PACs, PFR proximal and distal zones (54nm repeat). Symbols indicating orientation are defined at the bottom and are used throughout the figures.
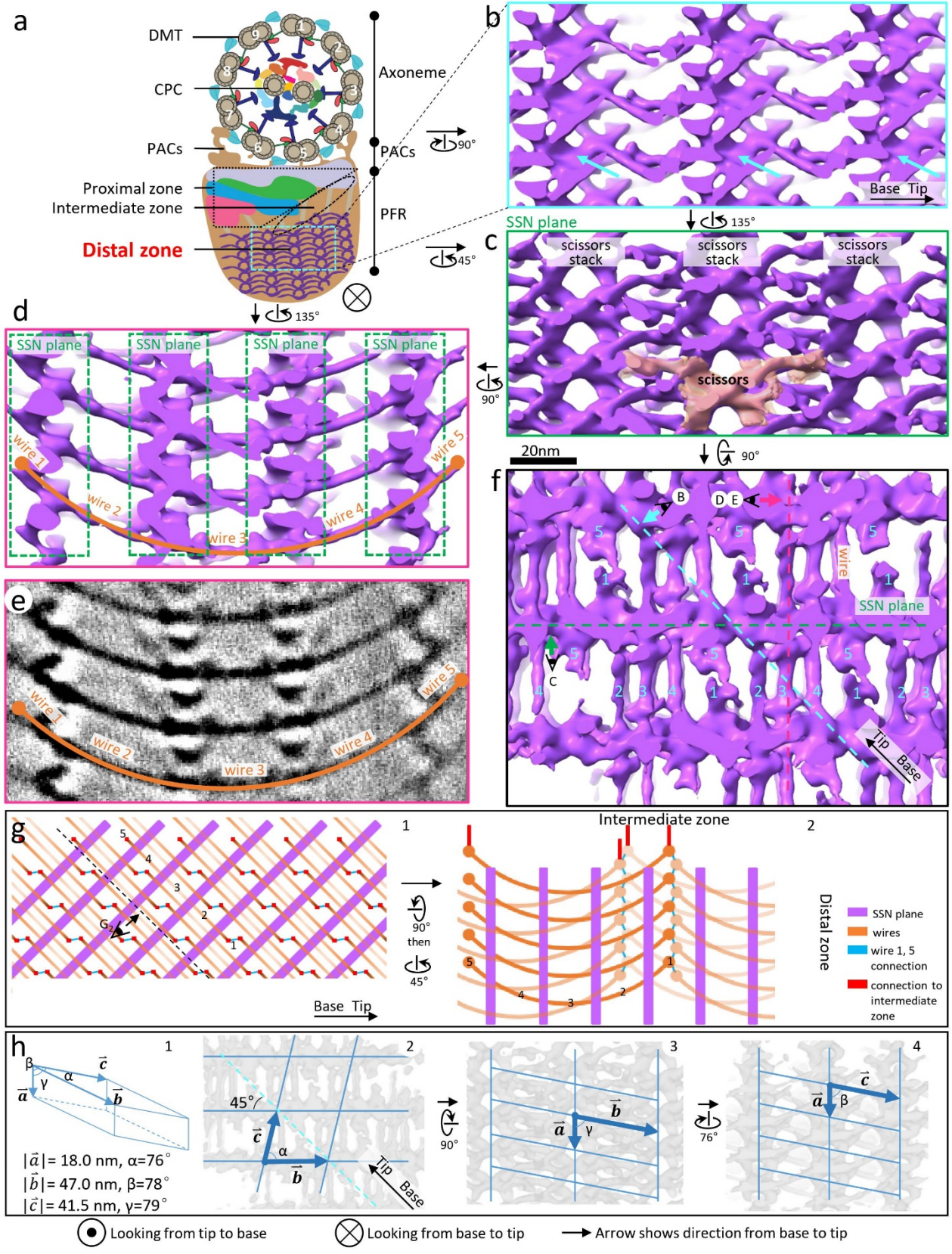
Figure 5- 2. Structure of the paracrystalline PFR distal zone.

(a) A schematic depicting a cross-section of *T. brucei* flagellum viewed from base to tip.

(b-d, f) Shaded surface views of the boxed region in (a) after rotating as indicated. The "comb teeth" densities in Fig. 1b are resolved in greater detail in b and pointed out by cyan arrows in both. b is the side view; c is looking perpendicular to an SSN plane; d is looking parallel to the SSN planes; f is looking from the axoneme to the PFR distal zone.

(e) Same view as (d), showing density slice (5nm thick). In d and e, wires 1-5 joining four consecutive SSN planes form a shape of a smiling face. The viewing directions of (b-e) are indicated within panel f, with letters corresponding to each panel and viewing directions indicated by an arrow next to the eye symbol. Scale bar for panels b-f is shown between panels c and f.

(g) Schematics depicting arrangement of SSN planes (purple) and wires (orange) at the interface of the PFR distal and intermediate zones. The viewing direction of g_1 is the same as that of f except that the horizontal direction is now along the axoneme axis. Decreased intensity of orange wires indicates further away from the viewer.

(h) Geometry schematic illustrating the three vectors of the triclinic crystalline lattice discovered in the distal zone. The semi-transparent gray densities in the background of h_2 is the same as that in (f), and rotations for h_3 and h_4 are indicated.
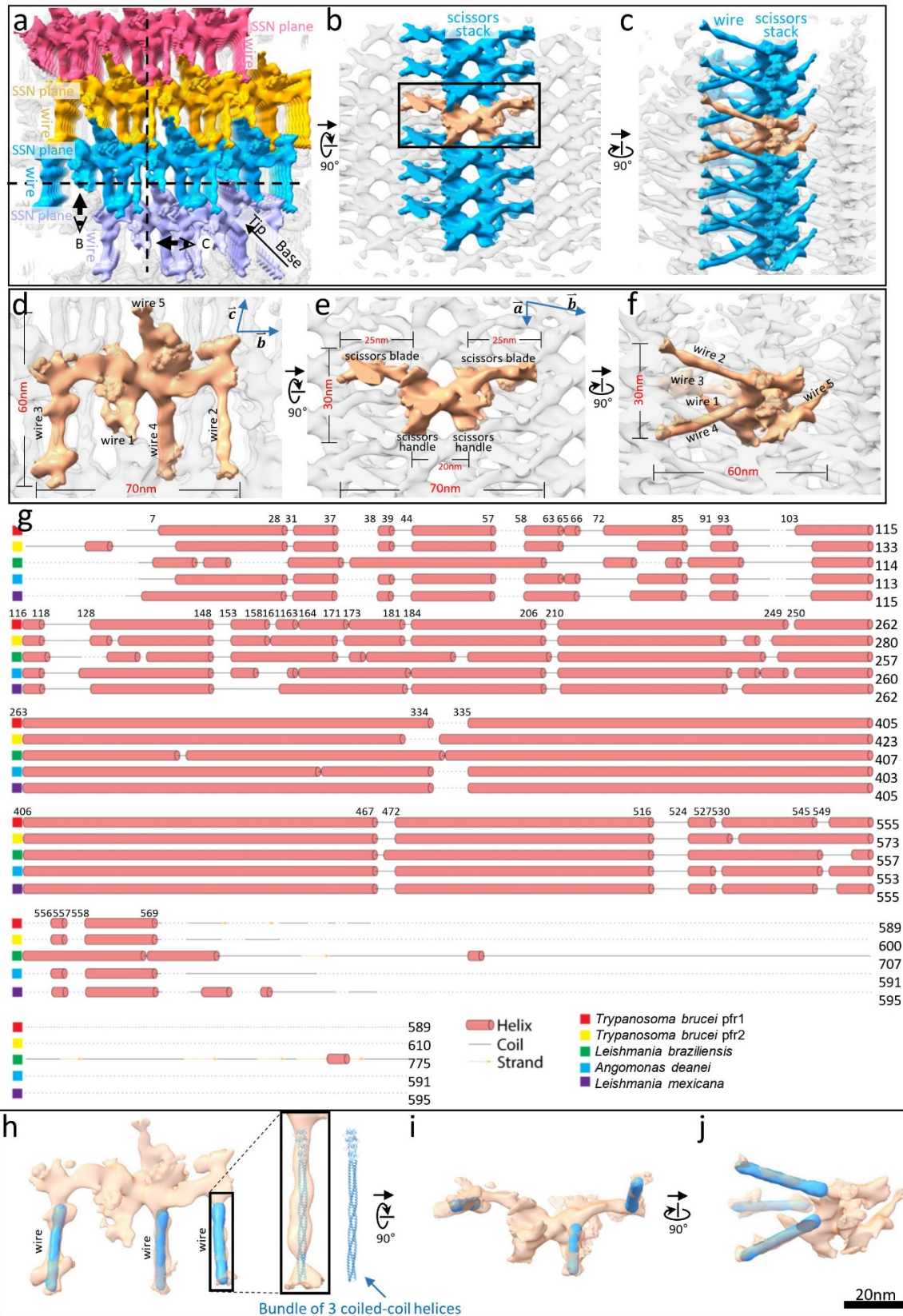
Figure 5- 3. Paracrystalline network of the PFR distal zone and structural

interpretation of subunits

(a) An overall surface view of a PFR distal zone encompassing four SSN planes (different colors) and their connecting wires, showing the paracrystalline arrangement. The view is looking at the PFR distal zone from the axoneme.

(b, c) Two orthogonal surface views of (a) at the sectional planes indicated by the dashed lines in (a), showing a single scissors density with wires (tan) in a stack (blue) and joined to adjacent scissors stacks (grey), forming an SSN plane of the paracrystalline zone.

(d-f) Three orthogonal zoom-in views of a single scissors density with wires (tan) within the paracrystalline distal zone (grey). Viewing directions in d, e and f are the same as in a, b and c, respectively.

(g) Predicted secondary structures of major PFR proteins from the indicated organisms, shown according to their sequence alignment. Amino acid residue numbers of *T. brucei* PFR1 are indicated on the top.

(h-j) Three orthogonal views of a single scissors density with wires (transparent tan) showing a 3 coiled-coil helix bundle (PDB: 6GAO) [103] (blue) fitted into the wires.
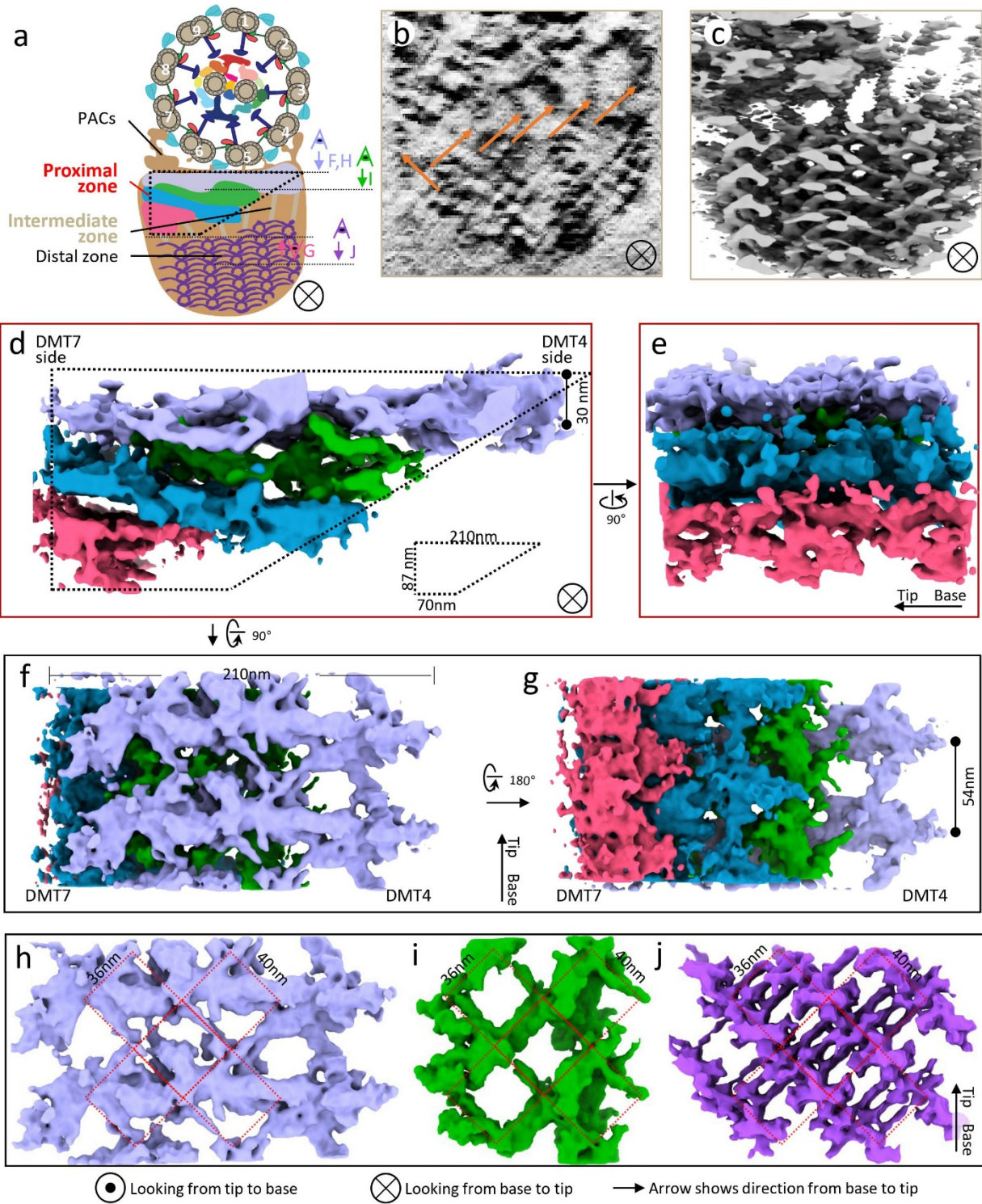
Figure 5- 4. Structure of the proximal zone and its structural continuity with the intermediate and distal zones

(a) A schematic illustrating the location of the trapezoid-shaped proximal zone with respect to the

rest of the flagellum.

(b, c) Cross-section density slice (10nm thickness) from a single tomogram (b) and the surface view of the same region after sub-tomographic averaging displayed at a low-density threshold (c). The connecting densities of the intermediate zone (arrows in b) contact wires of the distal zone.

(d-g) Cross-section (d), side (e), top (f), and bottom (g) surface views of the trapezoid-shaped proximal zone after sub-tomographic averaging. The four density layers distinguishable by gaps are colored differently. Viewing angles are shown in (a).

(h-j) Top views of the light purple (h) and green (i) proximal zone layers, alongside the same view of the distal zone (j), highlighting the contiguous 36 by 40nm lattice pattern (red) in all three.
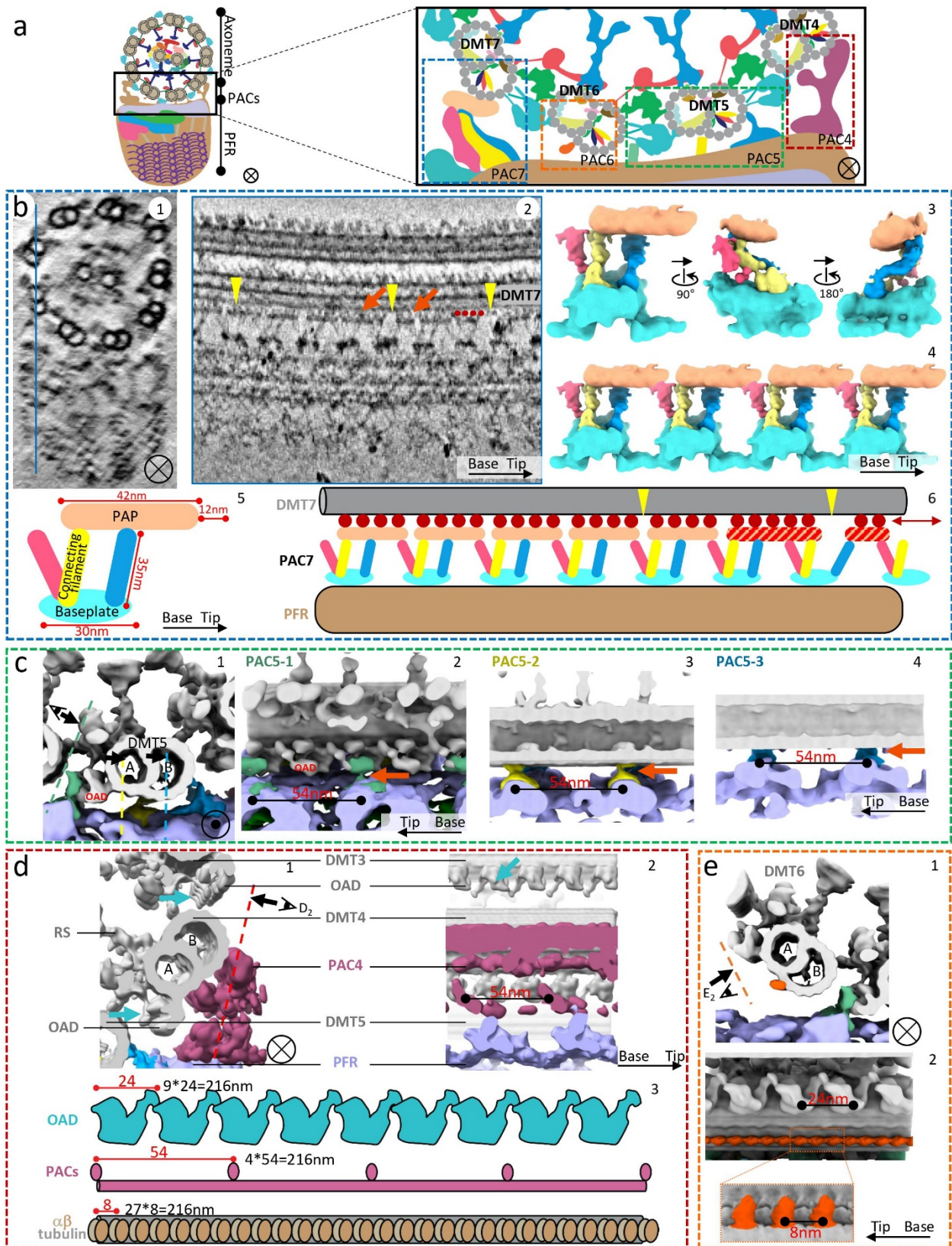
**Figure 5- 5. Structures of PFR-axoneme connectors (PACs)**

(a) A schematic with an enlarged inset illustrating the locations of the four PACs in the *T. brucei* flagellum. The colored boxes enclosing each PAC approximate the regions used for obtaining sub-tomographic averages detailed in (b-e).

(b) Details and dynamics of the PAC7 structure. A 10nm-thick longitudinal density slice (b_2) from the location indicated by the blue line in B_1 shows that PAPs have variable sizes (orange arrows), are connected to DMT7 by globular densities (red dots) and are separated by variable gaps (yellow arrowheads). Three surface views of the averaged PAC7 structure (b_3) show its PAP, baseplate, and connecting filaments. Four PAC7 structures are montaged (b_4) according to the locations in b_2 to illustrate connectivity. Schematics in b_5 and b_6 depict dimensions of PAC7 components (b_5), illustrate how PAPs interact with DMT7 through globular densities (red dots) (b_6), and that sliding (double-headed arrow) might produce the heterogeneity in size and spacing observed for PAPs (b_6).

(c) Zoomed in surface views of the PAC5 region showing details and interconnections. Each panel is a montage of five sub-tomographic averages: the 96nm averaged axoneme (grey), 54nm averaged PFR proximal zone (light purple), PAC5-1 (green), 5-2 (yellow), and 5-3 (blue).

(d) Two orthogonal zoomed in surface views (d_1 and d_2) and schematics (d_3) of the PAC4 region, illustrating periodicities of OAD, PACs, and the α/β-tubulin dimer. Panels d_1 and d_2 are each a montage of three sub-tomographic averages: the 96nm averaged axoneme (grey) and the 54nm averaged PFR proximal zone (light purple) and PAC4 (dark purple).

(e) Zoomed in surface views of the DMT6 region showing details. Panel e1 is a montage of four sub-tomographic averages: the 96nm averaged axoneme (grey), the 8nm averaged DMT6-MOP-B1,2 (orange), and the 54nm averaged PFR proximal zone (light purple) and PAC5-1 (green). Panel e_2 is a montage of two sub-tomographic averages: the 96nm averaged axoneme and the 8nm averaged DMT6-MOP-B1,2 (orange).
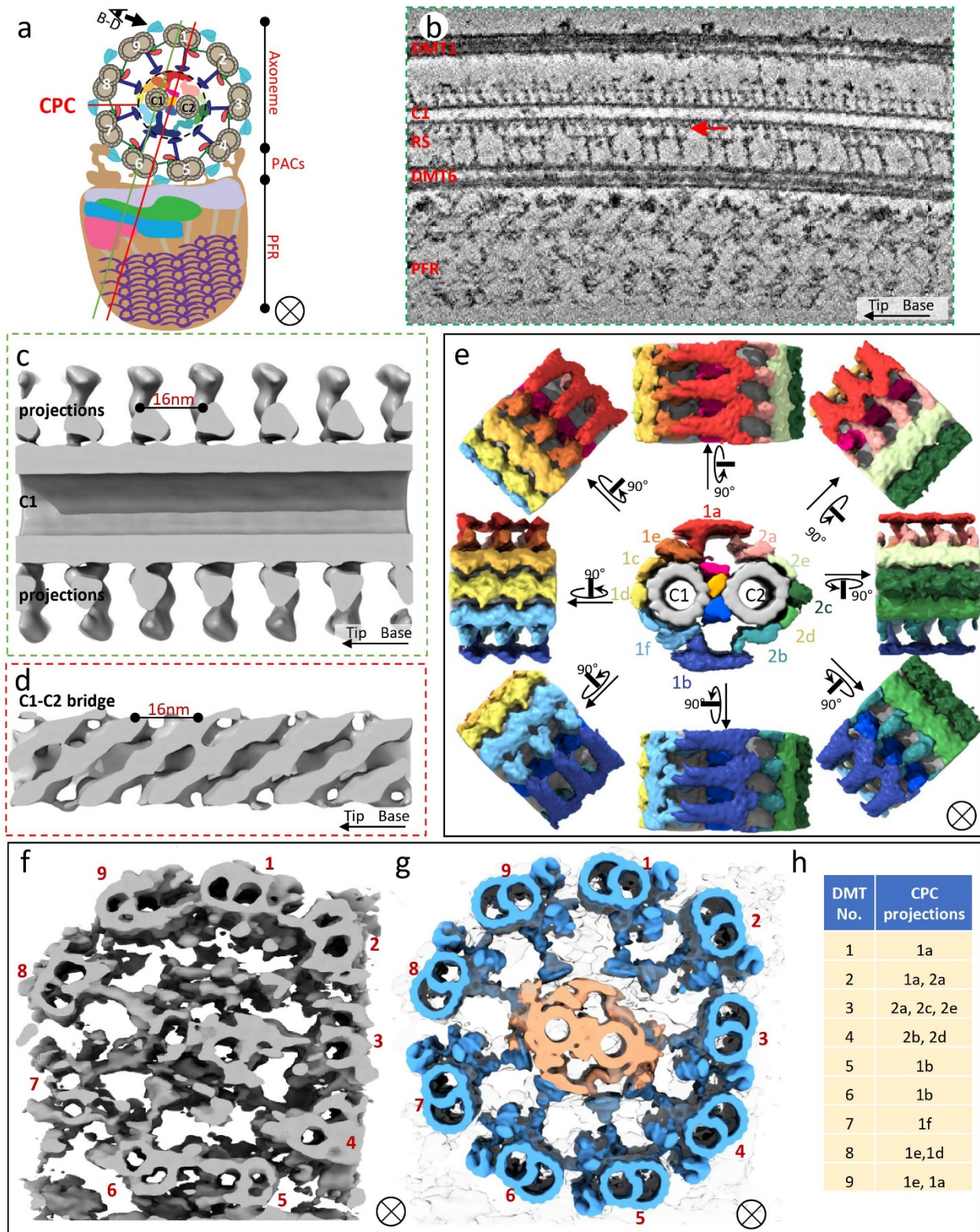
Figure 5- 6. Structure of the *T. brucei* central pair complex (CPC)

(a) A schematic illustrating the location and relative orientation of the CPC with respect to the

surrounding 9 DMTs. The section planes are indicated for panels b and c (green line), and panel d (red line).

(b) A 10nm-thick density slice through a tomogram along the plane marked by the green line in (a). Connections between the RS and CPC are visible (red arrow).

(c) Surface view of the CPC 96nm sub-tomographic average sectioned along plane marked by the green line in (a), showing C1 projections, which repeat at 16nm intervals.

(d) Surface view of the CPC 96nm sub-tomographic average sectioned along the plane marked by the red line in (a), showing the bridge structure between the C1 and C2 microtubules of CPC, which repeats at 16nm intervals.

(e) Cross sectional (middle) and rotational (surrounding) surface views of the CPC 16nm sub-tomographic average, showing the 1a-f and 2a-e projections, C1 and C2 microtubules (grey), and tripartite bridge structure (pink, light orange and blue).

(f-g) Connections between CPC and RS. Surface view of a single tomogram (f), fitting (g) of 96nm averaged CPC and individual DMT-RS structures within the tomogram shown in f.

(h) Summary of contacts observed between CPC and RS from each DMT (Based on Fig. 6g and Supplementary Fig. S2).

$$H(t) = A \cdot sin(\omega t) \cdot \vec{\imath}$$
$$+A \cdot cos(\omega t) \cdot \vec{\jmath}$$
$$+v \cdot t \cdot \vec{k}$$
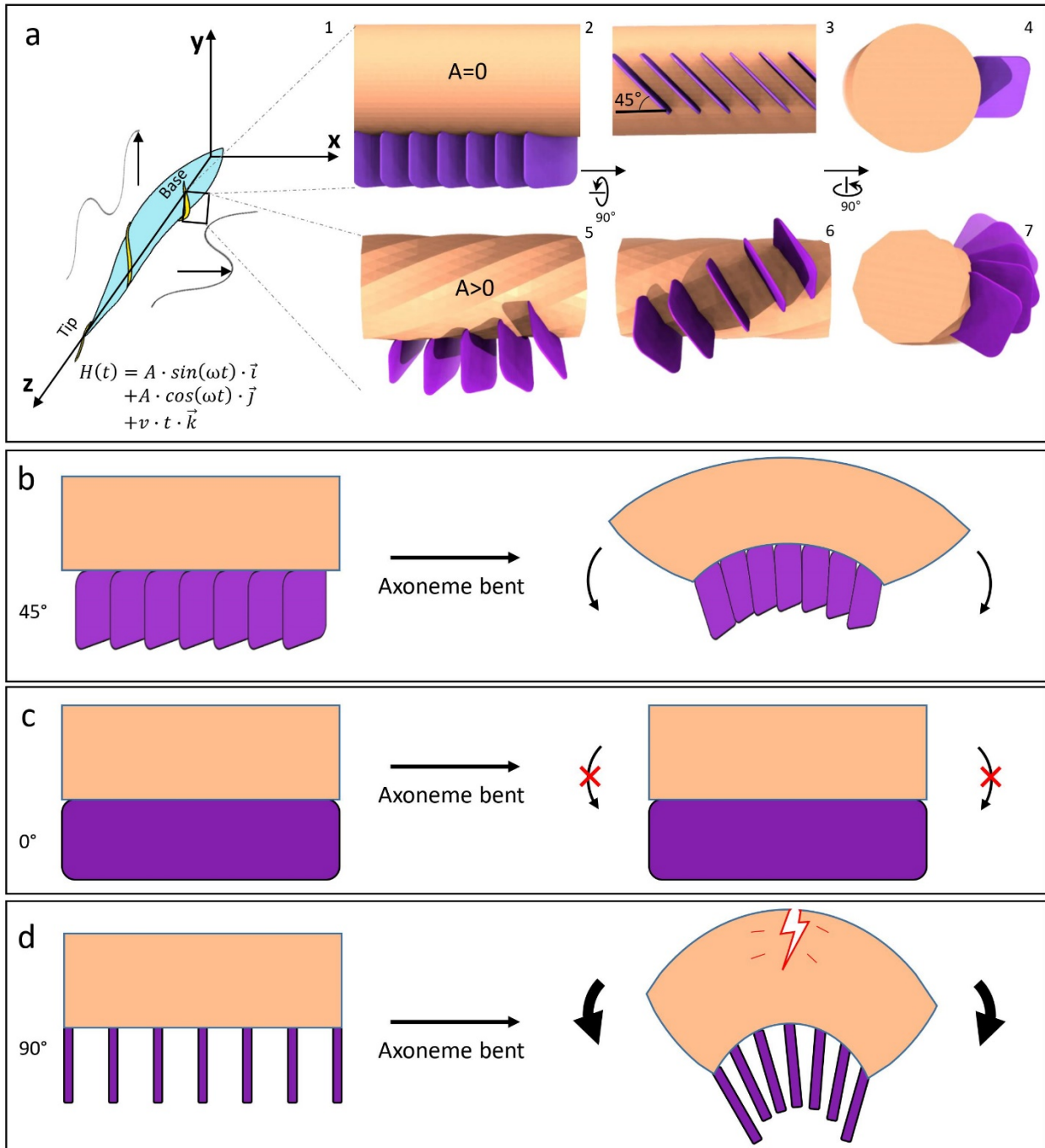
Figure 5- 7. Model for SSN planes supporting non-planar helical wave of the *T. brucei* flagellum

(a) Schematics depicting movement of the *T. brucei* cell with the flagellum's non-planar helical wave decomposed into x, y, and z components (left, a_1). $\vec{\imath}, \vec{\jmath}, \vec{k}$ are unit vectors along x, y, and z axis, respectively. The right panels (a_{2-7}) show the architectural relationship between the axoneme

(tan) and SSN planes of the PFR (purple) on the flagellum in the resting (amplitude A=0) and beating (A>0) states. Note that non-planar helical wave would introduce local twist to the flagellum (a_{5-7}). (b-d) Three different axoneme bending scenarios with SSN planes placed at either 45° (observed), or hypothetically, 0° or 90°, with respect to the axoneme axis.
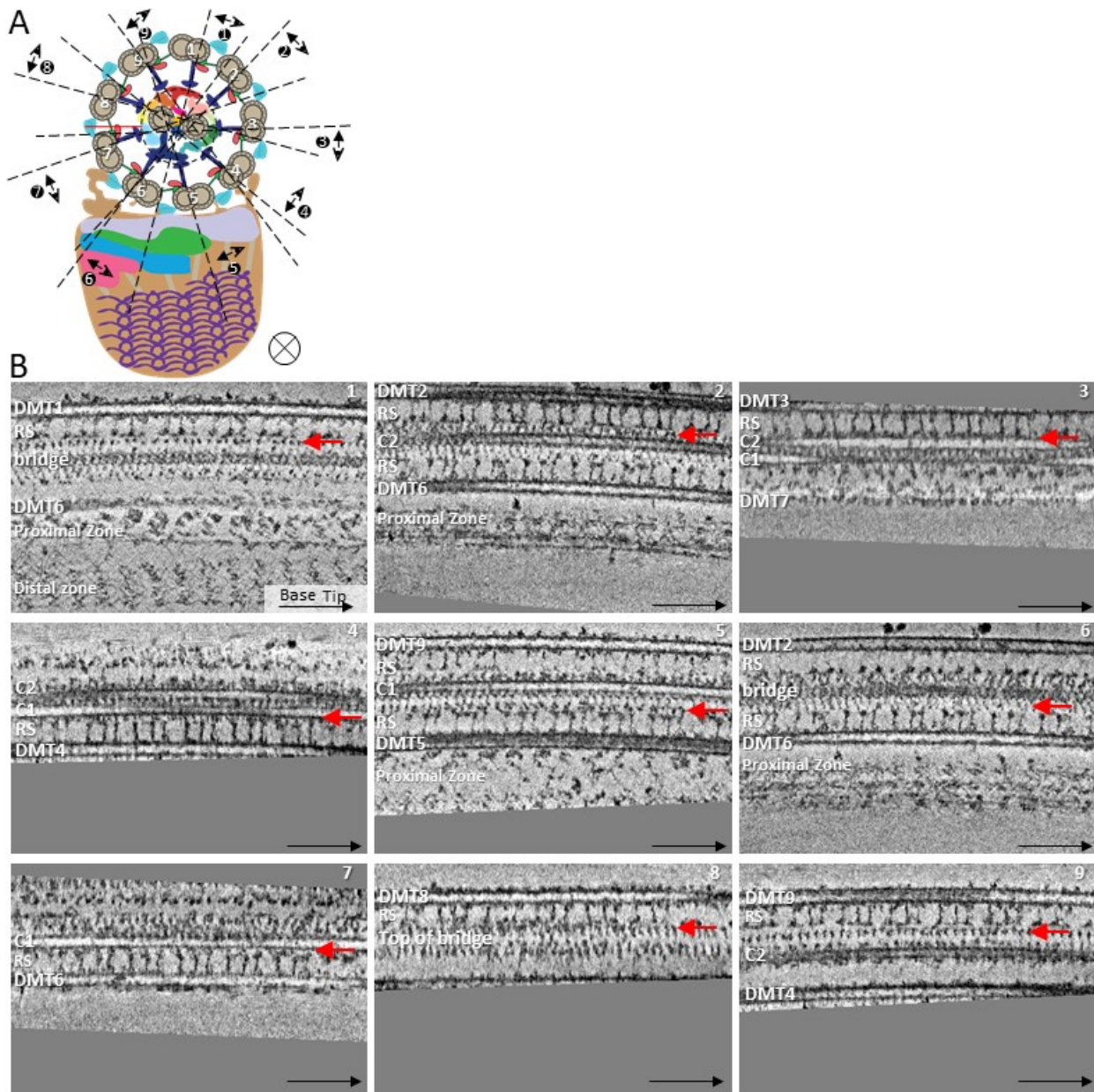
5.12 Supplemental Figures
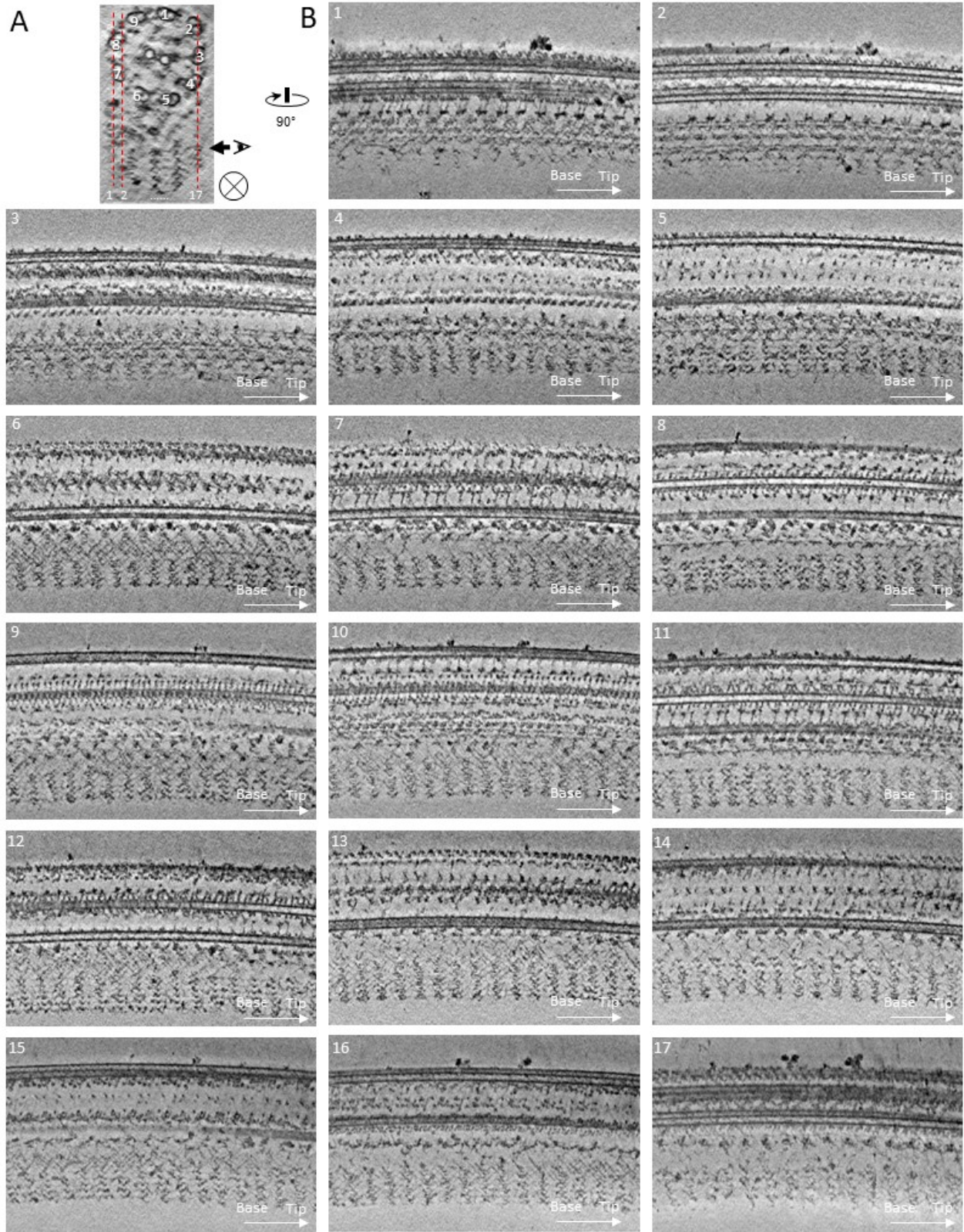
Supplemental Figure 5- 1. Density slices through a representative tomogram.

(a) Cross section views with DMT numbered.

(b) Seventeen evenly spaced longitudinal sections as indicated by red lines in a.

Supplemental Figure 5- 2. Density slices of a tomogram showing connections between CPC and RS.

(a, b) Schematic (a) with nine dashed lines indicating the section planes and viewing directions of longitudinal density slices (b1-b9). Arrows indicate connections between the CPC and RS from the indicated DMTs.

PACs	Protofilament / OAD
PAC4	A9, 10; B1, 3
PAC5	OAD | B1, 2 |B5
DMT6-MOP	B1, 2
PAC7	B2-4
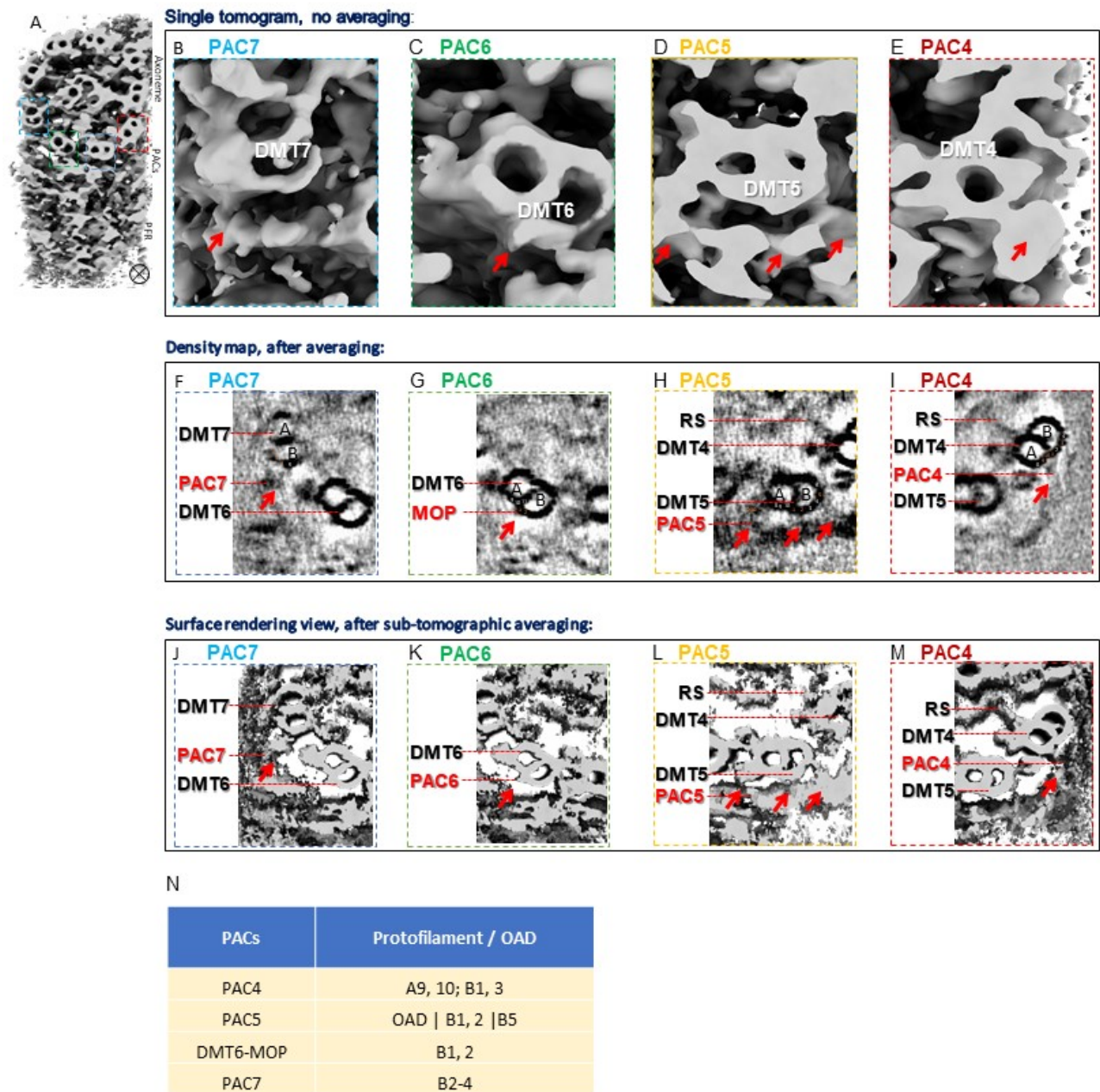
**Supplemental Figure 5- 3. PAC contacts with the axoneme.**

(a) Surface view of a single tomogram with boxes indicating PAC4-7 regions shown in zoomed-in views in be.

(b-e) Surface view of the PAC4-7 regions boxed in panel a from an individual tomogram.

(f-m) Density slices (f-i) and surface views (j-m) of a sub-tomographic average showing regions corresponding to those in b through e, respectively. Protofilaments of the A and B-tubules are

218

numbered according to convention.

(n) Summary of contacts between PACs and the axoneme (see also Fig. 5).
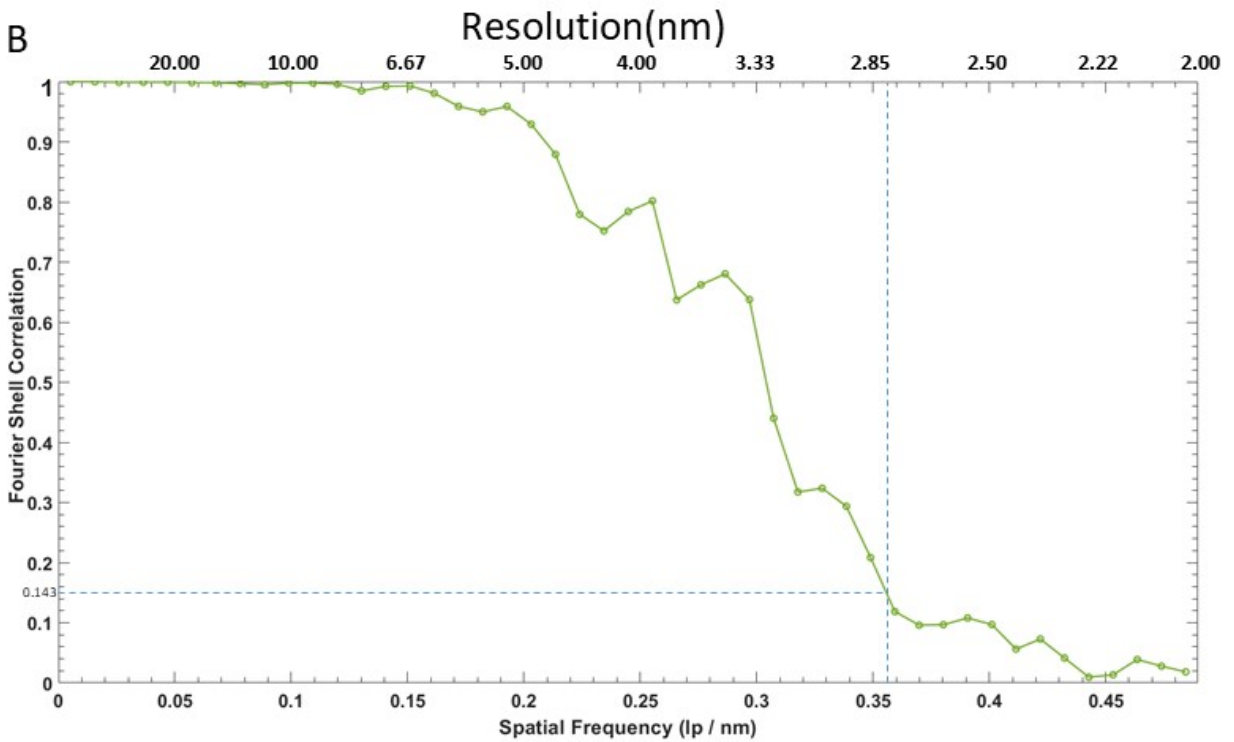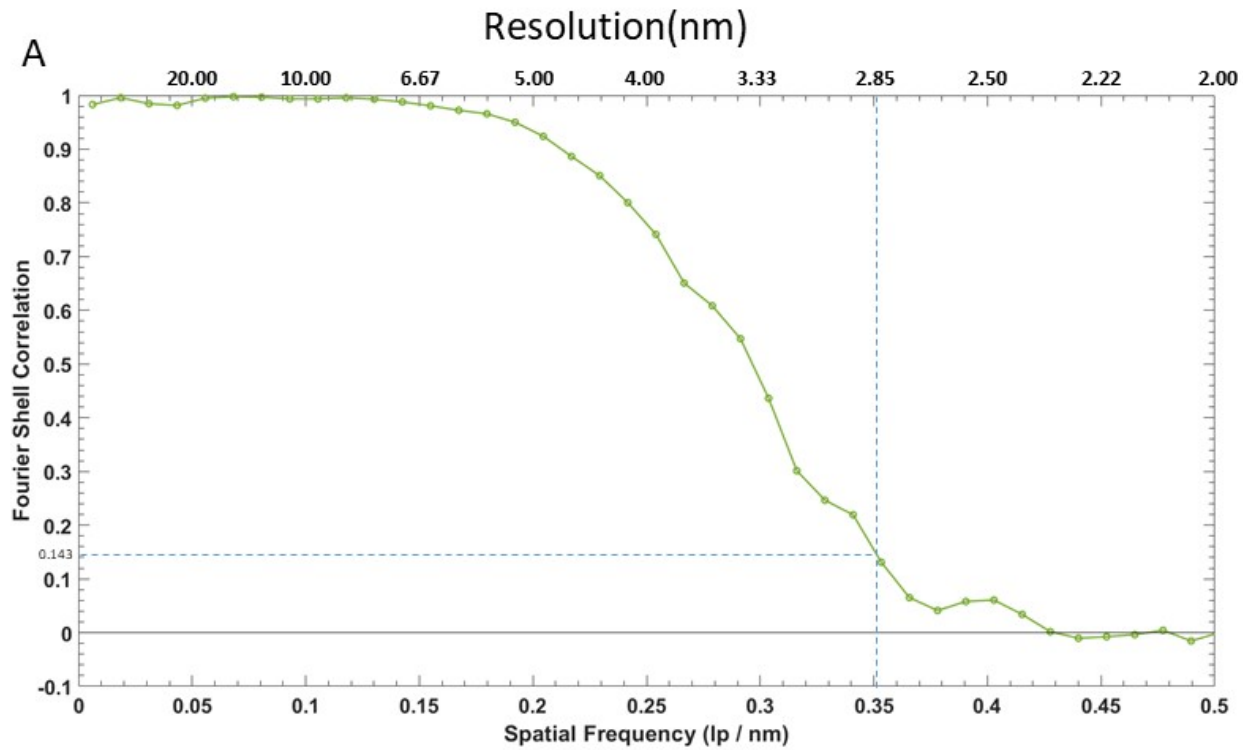
Supplemental Figure 5- 4. Resolution evaluation of sub-tomographic averages.

(a, b) FSC coefficients as a function of spatial frequency for the final sub-tomographic averages of the distal zone of PFR (a) and the CPC (b). The effected resolution, as indicated on the top of each plot, is based on the FSC at the 0.143 coefficient cutoff.

Chapter 6: Conclusion

To resolve cryoET structures essential for infection, from viruses to protozoa, we have firstly established the validity of our approach of cryoET with sub-tomogram averaging using a relatively simpler model organism, virus, then applied the same approach to investigate the flagellum structures of a more complex protozoa, *T. brucei*. We have captured the novel structural foundation of the 96-nm axonemal repeats and the trypanosome paraflagellar rod which shed lights into the non-planar motility of eukaryotic cells.

Firstly, we demonstrated the utility of cryoET with energy filtering and the cutting-edge technologies of VPP and direct electron-counting detection to capture metastable prefusion viral fusion proteins. We established the validity of our approach by obtaining cryoET *in situ* structures of the VSV glycoprotein G trimer (171 kD) in prefusion and postfusion conformations, which agree with the known crystal structures of purified G trimers in both conformations. From the native environment of herpesvirus HCMV virions, we identified gB trimers (303kD) in two distinct conformations in tomograms and resolved their structures at up to 21 Å resolution. We further captured the prefusion gB in complex with an "L"-shaped density attributed to the gH/gL complex. Our resolved structures of HCMV glycoproteins in multi-functional states provide insights on the mechanism of cellular membrane fusion when the HCMV infection begins.

To further understand the herpesvirus infection, we applied the approach of the cryoET and sub-tomogram averaging to understand how NEC—mediator of the viral capsid budding into the cytoplasm—interacts with the viral capsid and how proper

curvature of the coat is achieved to enable budding. We found that binding of a capsid protein, UL25, promotes the formation of a pentagonal rather than hexagonal NEC arrangement. Our results suggest that during nuclear budding, interactions between the UL25 bound to the pentagonal capsid vertices and the NEC introduce pentagonal insertions into the hexagonal NEC array to yield an NEC coat of the appropriate size and curvature, leading to productive budding and egress of UL25-decorated capsids.

Eukaryotic flagella rely on a microtubule-based axoneme, together with accessory filaments to carryout motility and signaling functions to infect the host. We then applied the cryoET technology to a more complex cellular circumstance, to investigate the flagellum structures in *T. brucei*, a protozoan parasite in the Excavata lineage that causes African trypanosomiasis. In the flagellum, the axoneme is accompanied by a PFR that supports non-planar motility and signaling necessary for infection, transmission and pathogenesis. We resolved the 3D 96-nm axonemal repeat and discovered several lineage-specific structures, including novel inter-doublet linkages and MIPs. We also obtained 3D structures of the PFR, PACs, and the CPC. Together, our findings fill the previously critical gap in understanding structural foundations of eukaryotic flagella, provide insights into flagellum-driven, non-planar helical motility of *T. brucei* and have broad implications ranging from cell motility and tensegrity in biology to engineering principles in bionics.

The body of my thesis work described here resolved high-resolution structures, essential for infection, from viruses to protozoa, by a validated approach of cryoET and sub-tomogram.