

UNIVERSITY OF CALIFORNIA

Los Angeles

**Method Developments for CryoET to Resolve *in situ*
Biological Structures**

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy in Bioengineering

by

Hui Wang

2023

© Copyright by

Hui Wang

2023

ABSTRACT OF THE DISSERTATION

**Method Developments for cryoET to Resolve *in situ*
Biological Structures**

by

Hui Wang

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2023

Professor Hong Zhou, Co-Chair

Professor Song Li, Co-Chair

Moving over X-ray crystallography, single particle cryogenic electron microscopy (cryoEM) has emerged as a method for determining atomic structures of purified proteins in recent years, kicking up a “cryoEM revolution”. However, obtaining high-resolution, in-situ cellular structures that faithfully represent the native environment remains a formidable challenge. To address this need, cryogenic electron tomography (cryoET) has emerged as an invaluable tool for biomedical research. CryoET has the unique capability to provide detailed insights into the molecular mechanisms and biological functions of macromolecular assemblies within their native cellular context. This dissertation embarks on a journey through the world of cryoET combined with subtomogram averaging (STA) to unravel the mysteries of cellular organelles.

The first part of this work showcases the power of cryoET and STA in studying the entire

flagellar structure of *Trypanosoma brucei*. This includes the elucidation of the conserved "9+2" axonemal 96 nm subunit, paraflagellar rod (PFR), PFR and Axoneme connectors (PACs), and center pair complex (CPC) structure. This comprehensive exploration sheds light on the motility unit of *T. brucei*, crucial for the parasite's transmission and pathogenesis. Leveraging the RNAi knockdown technique, we delve into the functions of critical proteins like DRC11, unraveling their structural role in parasite motility. The dissertation further addresses major challenges in the cryoET and STA pipeline through the introduction of two innovative software packages. Firstly, IsoNet tackles the persistent issue of "missing wedge" artifacts. In cryoET data collection, obtaining high-tilt angle images becomes impractical due to the increasing effective sample thickness. This missing information results in wedge-shaped artifacts in Fourier space, causing elongation artifacts along the Z-axis. IsoNet employs deep neural network techniques to fill meaningful signals into the Fourier space's missing wedge area, significantly mitigating or eliminating these artifacts.

Secondly, TomoNet focuses on automating particle picking functions through a blend of optimized template matching and deep learning methods. This powerful combination aids in the identification and localization of particles, particularly within lattice-like structures. TomoNet not only simplifies the organization of complete tomography projects but also efficiently manages extensive tomogram datasets through its user-friendly graphical interface. To illustrate TomoNet's particle picking capability, we implemented it in the study of the sheath structure of *Methanospirillum hungatei*. Applying cryoET with STA, we attained a remarkable 7.9 Å resolution, unveiling the topological features and subunit organization responsible for the native cylindrical sheath tube. Furthermore, a near-atomic level structure prediction of the monomeric sheath protein provides intricate insights into the architecture of individual *M. hungatei* sheath

hoops and their assembly into the sheath layer. We propose a novel route for the synthesis and assembly of sheath fibrils, specifically their insertion into growing sheath sites at the junctions between individual cells.

This body of work exemplifies the immense potential of cryoET, STA, and innovative software solutions like IsoNet and TomoNet in elucidating complex biological structures and mechanisms.

The dissertation of Hui Wang is approved.

Kent L. Hill

Jeffery F. Miller

Song Li, Committee Co-Chair

Hong Zhou, Committee Co-Chair

University of California, Los Angeles

2023

Table of Contents

| | |
|---|-----------|
| ABSTRACT OF THE DISSERTATION | ii |
| List of Figures | 10 |
| List of Tables | 14 |
| Acknowledgements | 15 |
| Vita | 18 |
| Chapter 1: Introduction | 1 |
| 1.1 Overview of cryoET | 1 |
| 1.2 The application of cryoET in resolving the structure of <i>T. brucei</i> flagellum | 2 |
| 1.3 Current challenges in cryoET | 3 |
| 1.3.1 Missing Wedge Compensation | 3 |
| 1.3.2 Automatic Particle Picking | 4 |
| 1.4 Addressing challenges in cryoET by IsoNet and TomoNet | 5 |
| 1.4.1 IsoNet | 5 |
| 1.4.2 TomoNet | 5 |
| 1.5 Resolving high resolution Sheath structure of <i>M. hungatei</i> | 6 |
| 1.6 Thesis Outline | 7 |
| 1.7 References | 8 |
| Chapter 2: Cryo electron tomography with Volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen <i>Trypanosoma brucei</i> | 13 |
| 2.1 Abstract | 14 |
| 2.2 Introduction | 15 |
| 2.3 Results | 17 |
| 2.3.1 3D Structure of the trypanosome 96-nm axonemal repeat | 17 |
| 2.3.2 Axonemal dynein arrangement in <i>T. brucei</i> | 19 |
| 2.3.3 Extensive Inter-doublet connections in the <i>T. brucei</i> axoneme | 20 |
| 2.3.4 Doublet-specific features of the 96-nm repeat | 21 |
| 2.3.5 CMF22/DRC11 is part of the NDRC proximal lobe involved in binding the adjacent DMT | 22 |
| 2.3.6 Extensive, lineage-specific MIPs in <i>T. brucei</i> | 23 |
| 2.4 DISCUSSION | 25 |
| 2.5 Materials and Methods | 28 |
| 2.5.1 Preparation of demembranated flagellum skeletons for cryoET | 28 |
| 2.5.2 CryoET sample preparation and tilt-series acquisition | 29 |
| 2.5.3 Data processing..... | 31 |
| 2.5.4 3D visualization | 34 |
| 2.5.5 Trypanosome motility videos..... | 34 |
| 2.6 Author contributions | 34 |

| | |
|---|------------|
| 2.7 Acknowledgements | 34 |
| 2.8 Data availability..... | 35 |
| 2.9 Figures | 35 |
| 2.10 Supplemental Figures..... | 51 |
| 2.11 Table | 60 |
| 2.12 References..... | 61 |
| <i>Chapter 3: Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells.....</i> | 72 |
| 3.1 Abstract | 73 |
| 3.2 Introduction | 74 |
| 3.3 Results | 76 |
| 3.3.1 Resolving <i>T. brucei</i> flagellum components with different periodicities | 76 |
| 3.3.2 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 | 77 |
| 3.3.3 Contiguous overlapping layers in the proximal zone and flexible linkages in the intermediate zone | 80 |
| 3.3.4 PAC structures bridge different repeats of the PFR and axoneme | 82 |
| 3.3.5 Structure of the <i>T. brucei</i> central pair complex | 83 |
| 3.4 Discussion | 84 |
| 3.5 Materials and Methods..... | 88 |
| 3.5.1 Sample preparation and cryoET..... | 88 |
| 3.5.2 Missing-wedge compensation | 88 |
| 3.5.3 Sub-tomographic averaging..... | 89 |
| 3.5.4 Sequence alignment and secondary structure prediction of major PFR proteins | 92 |
| 3.5.5 3D visualization | 92 |
| 3.6 Acknowledgments | 93 |
| 3.7 Author contributions | 93 |
| 3.8 Conflict of interest | 93 |
| 3.9 Data Availability | 93 |
| 3.10 Figures..... | 93 |
| 3.11 Supplemental Figures..... | 107 |
| 3.12 References..... | 113 |
| <i>Chapter 4: Isotropic Reconstruction of Electron Tomograms with Deep Learning</i> | 125 |
| 4.1 Abstract | 126 |
| 4.2 Introduction | 127 |
| 4.3 Results | 129 |
| 4.3.1 Workflow of IsoNet..... | 129 |
| 4.3.2 Benchmarking with simulated data | 132 |
| 4.3.3 Application to virus tomograms..... | 133 |
| 4.3.4 Application to tomograms of cellular organelles..... | 135 |

| | |
|--|-------------------|
| 4.3.5 Applications to tomograms of cells..... | 136 |
| 4.4 Discussion | 138 |
| 4.5 Materials and Methods..... | 140 |
| 4.5.1 Software implementation | 140 |
| 4.5.2 Dataset preparation | 141 |
| 4.5.3 Deconvolve CTF..... | 141 |
| 4.5.4 Generate mask | 142 |
| 4.5.5 Extract subtomograms | 143 |
| 4.5.6 Refine | 144 |
| 4.5.7 Predict..... | 147 |
| 4.5.8 Benchmarking with simulated data | 148 |
| 4.5.9 Processing tomograms of HIV virus | 149 |
| 4.5.10 Processing tomograms of the Eukaryotic flagella | 149 |
| 4.5.11 Processing tomograms of hippocampal neurons..... | 150 |
| 4.5.12 3D visualization | 151 |
| 4.6 Acknowledgments | 151 |
| 4.7 Author contributions | 152 |
| 4.8 Conflict of interest | 152 |
| 4.9 Data Availability | 152 |
| 4.10 Figures..... | 152 |
| 4.11 Supplemental Figures..... | 161 |
| 4.12 References..... | 169 |
| <i>Chapter 5: A streamlined cryoET software pipeline with automatic particle picking on flexible lattices</i> | <i>175</i> |
| 5.1 Abstract | 176 |
| 5.2 Impact Statement | 176 |
| 5.3 Introduction | 177 |
| 5.4 Results | 180 |
| 5.4.1 Overall design of TomoNet | 180 |
| 5.4.2 “Auto Expansion” particle picking..... | 180 |
| 5.4.3 Automatic particle picking by deep learning | 181 |
| 5.4.4 3D Classification using TomoNet | 182 |
| 5.4.5 Application to in situ viral protein arrays: the matrix protein lattice in HIV VLPs | 183 |
| 5.4.6 Application to focused ion beam (FIB)-milled cellular sample: the surface layer lattice of prokaryotic cell | 184 |
| 5.4.7 Application to in vitro assembled arrays: nuclear egress complex (NEC) lattice | 184 |
| 5.5 Discussion | 185 |
| 5.6 Methods..... | 187 |
| 5.7 Availability | 194 |
| 5.8 Acknowledgement..... | 194 |
| 5.9 Authorship contributions | 194 |

| | |
|---|------------|
| 5.10 Competing Interests Statement..... | 194 |
| 5.11 Figures..... | 194 |
| 5.12 Supplemental Figures..... | 201 |
| 5.13 References..... | 204 |
| Chapter 6: Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein | 211 |
| 6.1 Abstract | 212 |
| 6.2 Introduction | 213 |
| 6.3 Results | 215 |
| 6.3.1 Organization of the <i>M. hungatei</i> sheath by cellular cryoET..... | 215 |
| 6.3.2 SH contains a motif similar to the β -arch kernel found in amyloid fibrils..... | 217 |
| 6.3.3 SH interactions within a β -hoop and between adjacent β -hoops | 218 |
| 6.3.4 In situ structure of an immature cell | 221 |
| 6.4 DISCUSSION..... | 222 |
| 6.5 Materials and Methods..... | 225 |
| 6.5.1 <i>M. hungatei</i> cell growth and sample preparation | 225 |
| 6.5.2 CryoET and tomogram reconstruction..... | 226 |
| 6.5.3 Subtomogram averaging..... | 227 |
| 6.5.4 Modeling and 3D visualization | 228 |
| 6.6 Author contributions | 228 |
| 6.7 Acknowledgements | 228 |
| 6.8 Data availability..... | 228 |
| 6.9 Figures | 229 |
| 6.10 Supplemental Figures..... | 236 |
| 6.11 Table | 241 |
| 6.12 References..... | 243 |
| Chapter 7: Conclusion | 250 |

List of Figures

Chapter 1

Figure 1- 1. Typical workflow of cryoET as illustrated by *T. brucei* flagellum.....2

Chapter 2

Figure 2- 1. Phylogenetic tree of eukaryotes36

Figure 2- 2. Intact demembranated flagella from BSF *T. brucei*.....37

Figure 2- 3. The 3D ultrastructure of the 96-nm repeat from intact axonemes of BSF *T. brucei* .39

Figure 2- 4. *In situ* structure of outer arm dyneins and novel OAD-alpha inter-doublet connector in BSF *T. brucei*.....41

Figure 2- 5. Comparison of 96-nm axonemal repeat structures across species.43

Figure 2- 6. Doublet-specific structures of the BSF *T. brucei* 96-nm repeat.....44

Figure 2- 7. Comparison between averaged 96-nm repeats of wild-type and CMF22/DRC11 knockdown PCF *T. brucei*45

Figure 2- 8. TbMIP3 and ponticulus in the B-tubule of BSF *T. brucei*.....46

Figure 2- 9. The RingMIP and Ring Associated MIP (RAM) in the A-tubule of BSF *T. brucei*..47

Figure 2- 10. The snake MIP connects the A-tubule and the B-tubule of BSF *T. brucei*.....49

Figure 2- 11. Schematic overview of the trypanosome axoneme.51

Supplemental Figure 2- 1. Cross sections of the ten tomograms used to obtain the entire averaged BSF 96-nm axonemal repeat structure. [related to Fig. 3].....51

Supplemental Figure 2- 2. Gold Standard Fourier shell correlation (FSC) and ResMAP analyses. [related to Fig. 3].....52

Supplemental Figure 2- 3. Extra densities outside protofilaments b7b8 and massive density at the base of RS3 in BSF *T. brucei*. [related to Fig. 3].....54

| | |
|--|-----|
| Supplemental Figure 2- 4. Sub-tomogram averages of the 96-nm repeat of individual DMTs of BSF <i>T. brucei</i> . [related to Fig. 6] | 56 |
| Supplemental Figure 2- 5. The Spine MIP is a contiguous structure, spanning 48 nm and contacting adjacent MIPs in BSF <i>T. brucei</i> . [related to Fig. 8] | 57 |
| Supplemental Figure 2- 6. MIPs in the A-tubule of BSF <i>T. brucei</i> . [related to Fig. 9] | 58 |
| Supplemental Figure 2- 7. Illustration of the principles of <i>autoPicker</i> . [related to Fig. 3] | 59 |
| Chapter 3 | |
| Figure 3- 1. CryoET of <i>T. brucei</i> flagellum in its bloodstream form. | 94 |
| Figure 3- 2. Structure of the paracrystalline PFR distal zone. | 96 |
| Figure 3- 3. Paracrystalline network of the PFR distal zone and structural interpretation of subunits | 99 |
| Figure 3- 4. Structure of the proximal zone and its structural continuity with the intermediate and distal zones | 100 |
| Figure 3- 5. Structures of PFR-axoneme connectors (PACs) | 102 |
| Figure 3- 6. Structure of the <i>T. brucei</i> central pair complex (CPC) | 104 |
| Figure 3- 7. Model for SSN planes supporting non-planar helical wave of the <i>T. brucei</i> flagellum | 106 |
| Supplemental Figure 3- 1. Density slices through a representative tomogram. | 108 |
| Supplemental Figure 3- 2. Density slices of a tomogram showing connections between CPC and RS | 109 |
| Supplemental Figure 3- 3. PAC contacts with the axoneme. | 110 |
| Supplemental Figure 3- 4. Resolution evaluation of sub-tomographic averages. | 111 |

Chapter 4

| | |
|--|-----|
| Figure 4- 1. Principle and workflow of IsoNet. | 153 |
| Figure 4- 2. IsoNet reveals lattice defects in immature HIV capsid. | 156 |
| Figure 4- 3. Architecture of the PFR revealed after missing-wedge correction. | 157 |
| Figure 4- 4. IsoNet recovers missing information in the tomograms of neuronal synapses. | 158 |
| Figure 4- 5. IsoNet reveals various types clathrin coats in a synapse. | 160 |
| Supplementary Figure 4- 1. | 161 |
| Supplementary Figure 4- 2. Rotation schemes. | 162 |
| Supplementary Figure 4- 3. The architecture of neural network based on U-net. | 163 |
| Supplementary Figure 4- 4. Iteratively filling the missing-wedge region. | 164 |
| Supplementary Figure 4- 5. XZ slices of microtubule doublets in axoneme. | 165 |
| Supplementary Figure 4- 7. Orthogonal views of the tomogram containing clathrin cages reconstructed with the SIRT algorithm in IMOD. | 167 |
| Supplementary Figure 4- 8. 3D views for the shape of the heptagon containing clathrin cage. | 168 |
| Chapter 5 | |
| Figure 5- 1. Illustration of TomoNet's comprehensive pipeline for cryoET and STA. | 195 |
| Figure 5- 2. Illustration of the first two iterations of "Auto Expansion" particle picking. | 196 |
| Figure 5- 3. Illustration of "AI AutoPicking" process consisting of three steps. | 197 |
| Figure 5- 4. TomoNet application on HIV VLPs. | 199 |
| Figure 5- 5. TomoNet application on S-layer. | 199 |
| Figure 5- 6. TomoNet application on NEC lattice. | 200 |
| Supplementary Figure 5- 1. A screenshot of the TomoNet GUI. | 201 |
| Supplementary Figure 5- 2. Final map resolution of the HIV Gag hexamer. | 202 |
| Supplementary Figure 5- 3. Comparative visualization of the lattices obtained from TomoNet | |

and Relion tutorial.....203

Chapter 6

Figure 6- 1. In situ structure of *M. hungatei* sheath layer.....230

Figure 6- 2. Amyloid properties of SH monomer.....232

Figure 6- 3. Atomic model of sheath β -hoop.....233

Figure 6- 4. In situ structure of an immature cell.234

Figure 6- 5. Illustration of proposed sheath biogenesis of *M. hungatei* cell.....235

Supplemental Figure 6- 8. Raw cryoET tomogram and quality evaluation for subtomogram averages.....236

Supplemental Figure 6- 9. Negative stain TEM image of a region of a “ghost” *M. hungatei* cell showing overlapped hoop and flattened sheath segment.....237

Supplemental Figure 6- 10. Inner connection between β -sheet 1 and 2.238

Supplemental Figure 6- 11. Atomic model of the 4- β -ring hoop assembly and subtomogram average of β -hoop with extra densities near the plug end.....239

Supplemental Figure 6- 12. Surface charge of β -hoop.240

Supplemental Figure 6- 13. Hydrophobicity of β -hoop.....241

List of Tables

Chapter 2

| | |
|--|----|
| Table 2- 1. MIPs of BSF <i>T. brucei</i> | 60 |
|--|----|

Chapter 6

| | |
|--|-----|
| Table 6- 1. CryoET data collection and processing statistics | 242 |
|--|-----|

Acknowledgements

I extend my profound gratitude to Dr. Zhou, my mentor, whose unwavering support and guidance have been my steadfast companions throughout this five-year journey. His invaluable insights and wise counsel not only provided essential direction but also instilled in me the confidence to navigate the labyrinth of research. In the beginning of my PhD's journey at UCLA, I was uncertain about my scientific direction nor my life's path. Conversations with Dr. Zhou were enlightening, and his wisdom bolstered my self-belief. Under his mentorship, I achieved significant milestones, including publications, scholarships, and the opportunity to mentor exceptional UCLA undergraduates. Dr. Zhou also significantly enhanced my critical thinking abilities, making it an invaluable aspect of my academic journey. Even when faced with daunting projects, conversations with Dr. Zhou illuminated the path, assuring me that success was attainable.

I would like to express my heartfelt appreciation to my doctoral committee members Dr. Z. Hong Zhou, Dr. Kent L. Hill, Dr. Song Li, and Dr. Jeff F. Miller. Their support has been a constant source of strength throughout these years, providing me with a reliable compass. In particular, Professor Kent Hill has played a role akin to a second mentor at UCLA. He exhibited remarkable patience, offering clear explanations of fundamental biological concepts when I was still a novice in the field. His commitment to ensuring my understanding of these concepts was truly remarkable. With his background in biology, he provided me with a valuable perspective and insights that enriched my understanding of various research topics.

Special thanks go out to the past and present members of the Zhou Lab for their support, camaraderie, and companionship during my time at UCLA. Dr. Jiayan Zhang deserves my deepest appreciation for her patience and guidance in teaching me the intricacies of cryoET. We

accomplished numerous remarkable projects together, such as our work on the parasite trypanosome flagellum and archaeal-related research. Dr. YunTao Liu, one of my close friends, consistently shared brilliant ideas, including the inception of the IsoNet project. Our conversations were always inspiring, and his kindness extended to all our scientific discussions. I would also like to thank Dr. Xian Xia for generously sharing her expertise in high-resolution cryoEM. Her contributions significantly enhanced my knowledge of wet lab experiments. Angela Agnew, with whom I collaborated on PFR-related research, deserves my gratitude. I am also thankful to Qing Lou for her help during our collaboration on Toxoplasma cryoET research. Among the undergraduate students, Shiqing Liao, Xueting Zhou, and Star Yu deserve my appreciation for their dedication, especially Shiqing, who spent three to four years actively participating and contributing to nearly all my projects.

I would also like to extend my gratitude to Dr. Peng Ge, Ivo Atanasov, Wong Hoi Hui, and Dr. David Strugatsky for providing extensive technical support and guidance. Their assistance with electron microscopy, even during weekends, was invaluable. They patiently taught me how to prepare samples, adjust electron microscope parameters, and collect data.

I am deeply grateful to the UCLA Bioengineering Department for providing departmental fellowships and to the MIMG Department for offering me the opportunity to serve as a teaching assistant for a semester and for nominating me for the Schering Award. I am sincerely appreciative of these prestigious scholarships, which provided international students like me with the opportunity to explore the beauty of science.

My heartfelt thanks go to my collaborators, Prof. Kent Hill, Prof. Robert Gunsalus, and Prof. Ekaterina Heldwein, for their unending support and mentorship. Working with them has been a delight, and they warmly embraced me into the world of *T. brucei*, *M. Hungatei*, and HSV-1

nuclear egress complex.

Lastly, I would like to express my gratitude to my family - my wife, Jiayan Zhang, and my son, Jeff Wang, who have been my pillars of strength. Their support and ability to uplift me during challenging times have been a constant source of motivation. Jiayan, as my soulmate, is always proud of me and provides me with confidence. I would also like to thank my parents, ChengDing Wang and Mei He, as well as my in-laws, ChangLai Zhang and CongLian Xiang, for their love and support.

Vita

Education

- PhD student, Bioengineering, University of California, Los Angeles, USA
2018 – 2023
- Master's degree in computer science, The University of Texas at Dallas, USA
2016 – 2018
- Bachelor's degree in software engineering, Tianjin University, China
2016 – 2018

Publications

- Imhof S, Zhang J, **Wang H**, Bui KH, Nguyen H, Atanasov I, Hui WH, Yang SK, Zhou ZH, Hill KL. Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*. *Elife*. (2019) Nov 11;8:e52058.
- **Wang H***, Zhang J*, Imhof S, Zhou X, Liao S, Atanasov I, Hui WH, Hill KL, Zhou ZH. Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell discovery*. (2021) Jul 13;7(1):1-7.
- Liu YT, Zhang H, **Wang H**, Tao CL, Bi GQ, Zhou ZH. Isotropic reconstruction for electron tomography with deep learning. *Nature communications*. (2022) Oct 29;13(1):6482.
- Shimogawa M., Wijono S., **Wang H.**, Zhang J., Sha J., Szombathy N., ... & Hill K. (2023). FAP106 is an interaction hub for assembling microtubule inner proteins at the cilium inner junction. *Nature Communications*, 14(1), 5225.
- **Wang H**, Liao S, Yu X, Zhou ZH, TomoNet: an integrated workflow GUI software for cryo-electron tomography including particle picking for lattice like structures. In submission to *Biological Imaging*.
- **Wang H**, Zhang J, Toso D, Liao S, Sedighian F, Gunsalus R, Zhou ZH, Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein, *Nature communications*. (2023) In press.
- Draganova EB, **Wang H**, Wu M, Liao S, Vu A, Gonzalez-Del Pino G, Zhou ZH, Roller RJ, Heldwein K. The universal suppressor mutation in the HSV-1 nuclear egress

complex restores membrane budding defects by stabilizing the oligomeric lattice.
bioRxiv. (2023) & PLOS Pathogens, in revision.

*: These authors contribute equally to this research.

Presentations and Posters

- “High resolution 3D structure of the trypanosome paraflagellar rod”, UCLA Parasitology Seminar, Mar. 2021
- “Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*”, the 5th Annual SoCal CryoEM Symposium, Oct. 2020
- “An amyloid-like protein polymerizes into sheath during archaeal cell growth”, 7th SoCal Cryo-EM Symposium, 2022
- “Isotropic Reconstruction of Electron Tomograms with Deep Learning”, 4th International Symposium on Cryo-3D Image Analysis, 2022

Awards and Honors

- UCLA Bioengineering departmental fellowship 2018 – 2019
- UCLA MIMG Schering Award 2022 – 2023

Teaching

University of California, Los Angeles (UCLA)

- MIMG105: Biological Microscopy 2019 Fall

The University of Texas at Dallas (UTD)

- CS6363: Design and Analysis of Computer Algorithms 2018 Spring
- CS4349: Advanced Algorithm Design and Analysis 2017 Fall

Chapter 1: Introduction

1.1 Overview of cryoET

The world of structural biology has witnessed a transformative revolution in recent decades with the advent of cutting-edge imaging techniques. Among these, cryogenic electron tomography (cryoET) stands out as a powerful tool that has pushed the boundaries of our understanding of cellular and molecular structures¹⁻⁵. As we delve into the intricacies of life at the nanoscale, cryoET offers a unique and invaluable perspective, allowing researchers to explore the 3D architecture of biological specimens with unprecedented clarity and precision⁶.

The cryoET technique, which combines cryogenic sample preparation with transmission electron microscopy (TEM) and computational tomographic reconstruction, has opened up new avenues for visualizing complex biological structures in their native, hydrated states. This method has proven to be especially instrumental in elucidating the organization and interactions of macromolecules^{1,7,8}, subcellular organelles^{9,10}, and cellular assemblies¹¹⁻¹⁴. From uncovering the structural details of virus particles to unraveling the mysteries of cellular complexes, cryoET has emerged as a transformative technique with broad applications in the life sciences¹³.

The cryoET workflow (Fig. 1) begins with the critical step of sample preparation, rapidly freezing specimens in vitreous ice to maintain their structural integrity. Data acquisition follows, where a series of 2D projection images is captured from various angles. The tilt series further back projected to perform tomographic reconstruction, where complex algorithms align and merge the tilt series images into a three-dimensional volume. This reconstructed volume provides a wealth of information and can be further analyzed and segmented to reveal internal structures. Subtomogram averaging (STA) is further used to investigate not only the structure of protein complex of interest, but also the higher resolution structure of their subunits¹⁵.

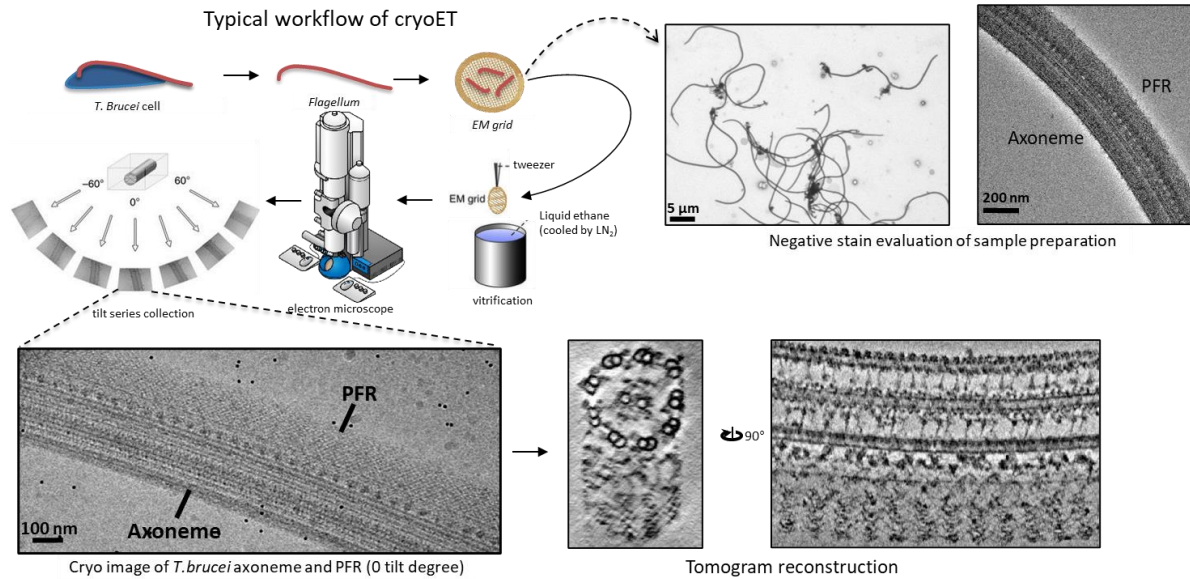


Figure 1- 1. Typical workflow of cryoET as illustrated by *T. brucei* flagellum.

1.2 The application of cryoET in resolving the structure of *T. brucei* flagellum

Flagella and cilia are essential hair-like structures protruding from the surface of eukaryotic cells, playing pivotal roles in motility and signaling¹⁶. While the 96-nm axonemal repeat, with its dynein motors and accessory structures, underpins the motility of these structures¹⁷, detailed 3D structural analyses have been limited, especially within the Excavata lineage, which includes medically significant pathogens. Additionally, the intricate paraflagellar rod (PFR) in *Trypanosoma brucei*, crucial for the parasite's unique non-planar helical motility and signaling, remained a mystery^{18,19}. Our study aimed to bridge these gaps, highlighting the application of cryoET as a powerful tool.

Employing cutting-edge cryoET, we unveiled the intricate 3D structure of both the *T. brucei* flagellum and its PFR. In the flagellum, we discovered novel lineage-specific structures, including unique inter-doublet linkages and microtubule inner proteins (MIPs). Notably, we established the

critical role of the DRC11/CMF22 subunit in binding adjacent doublet microtubules^{9,20}. These findings provide insights into the specialized adaptations that enable *T. brucei* to meet its distinct motility requirements in diverse host environments.

In the case of the PFR, our research unraveled the complex interface between the axonemal tubulin dimer and the PFR. We focused on the distal zone with STA, where we identified a unique "scissors stack network" (SSN) connected by helix-rich wires, creating a paracrystalline array with substantial empty space¹⁰. Our study also elucidated a contiguous pathway for signal transmission from the PFR to the axoneme, shedding light on the unique motility of *T. brucei*. Based on the model we established from the STA result of PFR distal zone, we also proposed a model for how SSN planes supporting the non-planar helical wave of the *T. brucei* flagellum and the distal zone served as biological spring structure that preserve the axoneme integrity. In this study, we have used a combination of cryoET and STA 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.

Our comprehensive research not only deepens our understanding of flagellum and PFR function but also demonstrates the potential of cryoET in unraveling complex structural enigmas in the field of flagellar biology and beyond.

1.3 Current challenges in cryoET

While cryoET combined with STA offers remarkable capabilities, it also comes with its own set of challenges, and two of them stand out: one is the cryoET inherited "missing wedge" effect and the second is the lack of automated tools, particularly in the context of automatic particle picking.

1.3.1 Missing Wedge Compensation

One of the fundamental challenges in cryoET is the presence of a so-called "missing wedge" in the acquired data. The missing wedge arises due to the limitations of data collection angles in a traditional electron microscope. In Fourier space, these missing views lead to a devoid of information in two continuous, opposing wedge-shaped regions, commonly referred to as the "missing wedge". This missing-wedge causes severe artifacts in the 3D reconstruction of cellular cryoET, manifesting as, e.g., oval-shaped synaptic vesicles²¹. Hence, the combination of low signal-to-noise ratio (SNR) in the reconstructed tomograms and the presence of missing-wedge artifacts creates a barrier to directly interpreting the 3D reconstructed densities. This interpretation is crucial for realizing the potential of cryoET in resolving molecular organization in its native environment.

1.3.2 Automatic Particle Picking

This workflow for cryoET and STA typically comprises several key components, and Users often need to navigate between several specialized software packages to optimize their results which is not convenient and often demands a certain level of computational proficiency. In particular, the method for particle picking varies on a case-by-case basis, dictated by the characteristics of in situ cellular samples. With an increasing demand for automation to enhance efficiency with minimal manual intervention, template matching has emerged as a popular method for automatic particle picking, relying on a user-provided reference map^{22,23}. Simultaneously, deep neural network demonstrates significant promise for cryoET automatic particle picking, thanks to its capacity to analyze three-dimensional feature maps and autonomously identify prominent features within specific samples²⁴⁻²⁶. This approach typically operates in a template-free mode and sometimes even obviates the need for human annotation²⁷.

In general, the lack of efficient automated particle picking tools in cryoET results in time-consuming, manual particle selection processes, which not only slow down data acquisition but also introduce subjective biases.

1.4 Addressing challenges in cryoET by IsoNet and TomoNet

1.4.1 IsoNet

IsoNet, a convolutional neural network (CNN)-based software system, was purpose-built for the isotropic reconstruction of electron tomography²⁸. IsoNet is optimized to operate effectively with low-resolution tomograms, featuring a pixel size of 10 Å. Its architecture involves training a deep CNN, which iteratively restores meaningful details to compensate for the missing wedge, drawing upon insights gained from the original tomogram. By applying IsoNet to a range of tomograms representing viral, organelle, and cellular samples, we have showcased its exceptional performance in unveiling novel structures. These include the identification of lattice defects within immature human immunodeficiency virus capsids, the revelation of the intricate scissors-stack-network architecture of the paraflagellar rod, and the discovery of heptagon-containing clathrin cages within neuronal synapses. The tomograms produced by IsoNet, featuring isotropic resolution, offer a significant advantage for the direct interpretation and segmentation of 3D cellular structures, as well as the potential to efficiently select hundreds of thousands of subtomogram particles for future high-resolution cryoET investigations.

1.4.2 TomoNet

TomoNet stands as a user-friendly GUI pipeline software package meticulously crafted to streamline and simplify the management of cryoET workflows. At its core is a "geometry-based" template matching approach founded on the concept of "auto expansion," which proves to be a versatile particle picking solution, particularly effective in structures featuring lattice-like

arrangements. This approach significantly diminishes the need for manual inputs, and its results seamlessly integrate into Relion for subsequent high-resolution 3D classification and refinement. Furthermore, we introduce a machine learning-based solution for automatic particle picking, which requires only 1-3 tomograms with known particle locations as ground truth, adding to its adaptability. Importantly, this method extends beyond lattice-like particles, making it a versatile and generalizable training model for a broader range of particle types. To demonstrate the prowess of TomoNet's particle picking methodology, we subjected it to the scrutiny of three distinct datasets, revealing its exceptional accuracy and efficiency in the identification of particles across diverse configurations.

1.5 Resolving high resolution Sheath structure of *M. hungatei*

The study of *Methanospirillum hungatei*'s proteinaceous sheath layer holds profound significance in the realm of cellular biology. This outermost layer, a defining feature of these cylindrical archaeal cells, has remained a source of intrigue due to its diverse characteristics and the pivotal role it plays in enabling gas exchange and cell growth^{29,30}. For decades, understanding the molecular composition and structural organization of this enigmatic sheath has been a challenge, fraught with questions about its assembly and maintenance³¹⁻³⁵. However, recent breakthroughs have been achieved with the application of innovative tools, IsoNet and TomoNet, paving the way for a comprehensive exploration of this essential cellular component.

Employing cryoET in conjunction with STA and harnessing the capabilities of IsoNet and TomoNet, we embark on a journey to unravel the molecular intricacies of *M. hungatei*'s cylindrical sheath layer. Approximately 450 SH monomers elegantly encircle the cylindrical cell, forming a remarkable ring structure termed the β -ring. This configuration consists of two colossal β -sheets, reminiscent of the " β -arch kernel" concept typically associated with amyloids. These β -rings, each

composed of 2 to 6 β -rings, stack axially, giving rise to an extended cylindrical sheath that can span hundreds of micrometers. Furthermore, our examination of SH structures within immature cells allows us to propose a pathway for the synthesis, assembly, and insertion of nascent β -hoops into the growing sheath. This discovery answers the long-standing question of how proteins traverse multiple barriers to assemble into the largest known structure composed entirely of a single protein—the external sheath of archaeal cells.

1.6 Thesis Outline

This dissertation is organized into seven chapters, each contributing to a comprehensive exploration of cryoET and its applications. Chapter 1 serves as an extensive introduction, providing a background on cryoET combined with STA for in situ macromolecular structure resolution. Chapters 2 and 3 delve into specific applications of cryoET and STA on *T. brucei* flagellar structure. Chapter 2 explores the novel structural foundations of the 96-nm axonemal repeat in *T. brucei* using cryoET with a Volta phase plate. Chapter 3 offers insights into the structure of the trypanosome PFR, shedding light on the non-planar motility of eukaryotic cells. Chapters 4 and 5 are dedicated to introducing the innovative tools, Isonet and TomoNet, and explaining how they address the current challenges in cryoET, respectively. In Chapter 6, the tools IsoNet and TomoNet are applied to successfully resolve the high-resolution structure of the *M. Hungatei* sheath layer. Lastly, Chapter 7 serves as a comprehensive summary, bringing together the findings and insights from Chapters 2 to 6 and offering a cohesive conclusion to the thesis, emphasizing the significant contributions made in advancing the field of cryoET and STA.

1.7 References

- 1 Wan, W. & Briggs, J. A. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol* **579**, 329-367, doi:10.1016/bs.mie.2016.04.014 (2016).
- 2 Chen, M. *et al.* A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature Methods* **16**, 1161-1168, doi:10.1038/s41592-019-0591-8 (2019).
- 3 Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Current Opinion in Structural Biology* **58**, 249-258, doi:<https://doi.org/10.1016/j.sbi.2019.05.021> (2019).
- 4 Castaño-Díez, D. & Zanetti, G. In situ structure determination by subtomogram averaging. *Curr Opin Struct Biol* **58**, 68-75, doi:10.1016/j.sbi.2019.05.011 (2019).
- 5 Hong, Y., Song, Y., Zhang, Z. & Li, S. Cryo-Electron Tomography: The Resolution Revolution and a Surge of In Situ Virological Discoveries. *Annu Rev Biophys* **52**, 339-360, doi:10.1146/annurev-biophys-092022-100958 (2023).
- 6 Beck, M. & Baumeister, W. Cryo-Electron Tomography: Can it Reveal the Molecular Sociology of Cells in Atomic Detail? *Trends Cell Biol* **26**, 825-837, doi:10.1016/j.tcb.2016.08.006 (2016).
- 7 Leigh, K. E. *et al.* Subtomogram averaging from cryo-electron tomograms. *Methods Cell Biol* **152**, 217-259, doi:10.1016/bs.mcb.2019.04.003 (2019).
- 8 Xue, L. *et al.* Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205-211, doi:10.1038/s41586-022-05255-2 (2022).
- 9 Imhof, S. *et al.* Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*. *eLife* **8**, e52058, doi:10.7554/eLife.52058 (2019).

- 10 Zhang, J. *et al.* Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell Discov* **7**, 51, doi:10.1038/s41421-021-00281-2 (2021).
- 11 Schur, F. K. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508, doi:10.1126/science.aaf9620 (2016).
- 12 von Kügelgen, A. *et al.* In situ structure of an intact lipopolysaccharide-bound bacterial surface layer. *Cell* **180**, 348-358. e315 (2020).
- 13 Huang, Y., Zhang, Y. & Ni, T. Towards in situ high-resolution imaging of viruses and macromolecular complexes using cryo-electron tomography. *Journal of Structural Biology* **215**, 108000, doi:<https://doi.org/10.1016/j.jsb.2023.108000> (2023).
- 14 Sibert, B. S. *et al.* Workflow for High-resolution Sub-volume Averaging from Heterogenous Viral and Virus-like Assemblies. *Microscopy and Microanalysis* **29**, 943-944, doi:10.1093/micmic/ozad067.470 (2023).
- 15 Metskas, L. A., Wilfong, R. & Jensen, G. J. Subtomogram averaging for biophysical analysis and supramolecular context. *Journal of Structural Biology: X* **6**, 100076, doi:<https://doi.org/10.1016/j.yjsbx.2022.100076> (2022).
- 16 Smith, E. F. & Rohatgi, R. Cilia 2010: the surprise organelle of the decade. *Sci Signal* **4**, mr1, doi:10.1126/scisignal.4155mr1 (2011).
- 17 Khan, S. & Scholey, J. M. Assembly, Functions and Evolution of Archaelia, Flagella and Cilia. *Curr Biol* **28**, R278-r292, doi:10.1016/j.cub.2018.01.085 (2018).
- 18 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, doi:10.1073/pnas.1103634108 (2011).

- 19 Alves, A. A. *et al.* Control of assembly of extra-axonemal structures: the paraflagellar rod of trypanosomes. *J Cell Sci* **133**, doi:10.1242/jcs.242271 (2020).
- 20 Bower, R. *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* **24**, 1134-1152, doi:10.1091/mbc.e12-11-0801 (2013).
- 21 Tao, C. L. *et al.* Differentiation and Characterization of Excitatory and Inhibitory Synapses by Cryo-electron Tomography and Correlative Microscopy. *J Neurosci* **38**, 1493-1510, doi:10.1523/jneurosci.1548-17.2017 (2018).
- 22 Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* **15**, 955-961, doi:10.1038/s41592-018-0167-z (2018).
- 23 Böhm, J. *et al.* Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences* **97**, 14245-14250, doi:doi:10.1073/pnas.230282097 (2000).
- 24 de Teresa-Trueba, I. *et al.* Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods* **20**, 284-294, doi:10.1038/s41592-022-01746-2 (2023).
- 25 Moebel, E. *et al.* Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods* **18**, 1386-1394, doi:10.1038/s41592-021-01275-4 (2021).
- 26 Hao, Y. *et al.* VP-Detector: A 3D multi-scale dense convolutional neural network for macromolecule localization and classification in cryo-electron tomograms. *Computer*

- Methods and Programs in Biomedicine* **221**, 106871,
doi:<https://doi.org/10.1016/j.cmpb.2022.106871> (2022).
- 27 Rice, G. *et al.* TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods* **20**, 871-880, doi:10.1038/s41592-023-01878-z (2023).
- 28 Liu, Y.-T. *et al.* Isotropic reconstruction for electron tomography with deep learning. *Nature Communications* **13**, 6482, doi:10.1038/s41467-022-33957-8 (2022).
- 29 Toso, D. B., Henstra, A. M., Gunsalus, R. P. & Zhou, Z. H. Structural, mass and elemental analyses of storage granules in methanogenic archaeal cells. *Environ Microbiol* **13**, 2587-2599, doi:10.1111/j.1462-2920.2011.02531.x (2011).
- 30 Poweleit, N. *et al.* CryoEM structure of the Methanospirillum hungatei archaeellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nature Microbiology* **2**, 16222, doi:10.1038/nmicrobiol.2016.222 (2016).
- 31 Zeikus, J. G. & Bowen, V. G. Fine structure of Methanospirillum hungatii. *J Bacteriol* **121**, 373-380, doi:10.1128/jb.121.1.373-380.1975 (1975).
- 32 Shaw, P. J., Hills, G. J., Henwood, J. A., Harris, J. E. & Archer, D. B. Three-dimensional architecture of the cell sheath and septa of Methanospirillum hungatei. *J Bacteriol* **161**, 750-757, doi:10.1128/jb.161.2.750-757.1985 (1985).
- 33 Southam, G. & Beveridge, T. J. Characterization of novel, phenol-soluble polypeptides which confer rigidity to the sheath of Methanospirillum hungatei GP1. *J Bacteriol* **174**, 935-946, doi:10.1128/jb.174.3.935-946.1992 (1992).

- 34 Southam, G. & Beveridge, T. J. Dissolution and immunochemical analysis of the sheath of the archaeobacterium *Methanospirillum hungatei* GP1. *J Bacteriol* **173**, 6213-6222, doi:10.1128/jb.173.19.6213-6222.1991 (1991).
- 35 Stewart, M., Beveridge, T. J. & Sprott, G. D. Crystalline order to high resolution in the sheath of *Methanospirillum hungatei*: a cross-beta structure. *J Mol Biol* **183**, 509-515, doi:10.1016/0022-2836(85)90019-1 (1985).

Chapter 2: 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 Bui⁵, Hoangkim Nguyen^{1,6}, Ivo Atanosov³, Wong H. Hui³, Shun Kai Yang⁵, 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

2.1 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.

2.2 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.

2.3 Results

2.3.1 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 OAD α 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.

2.3.2 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](#)).

2.3.3 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.

2.3.4 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 connector, 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.

2.3.5 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.

2.3.6 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.

2.4 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](#), [11](#), [14](#)), yet direct tests of this idea have been limited.

The 96-nm spacing of the axoneme is controlled by a molecular ruler ([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](#)), where the 24-nm repeat of OADs is dictated by something other than the FAP59/172 ruler ([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).

2.5 Materials and Methods

2.5.1 Preparation of demembranated flagellum skeletons for cryoET

BSF single marker (BSSM) and PCF (29-13) ([76](#)) *T. brucei* cells were cultured as described ([77](#), [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 μ l Extraction buffer (20mM HEPES pH: 7.4, 1mM MgCl₂, 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/ μ l) was added and incubated at room temperature for 15 minutes. In order to solubilize the subpellicular microtubules, 1mM CaCl₂ (2 μ l of 100 mM CaCl₂) 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 200 μ l of the upper fraction of the buffer-sucrose interface were collected and washed twice in 200 μ l Extraction buffer, centrifugation at 4,000 rpm 1500g at 4°C for 10 minutes, then resuspended in 40 μ l 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 μ m (Fig. 2B-F). PCF samples were examined by light microscopy to ensure uniform length distribution.

2.5.2 CryoET sample preparation and tilt-series acquisition

BSF or PCF samples in the amount of 40 μ l 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 μ l 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 $+60^\circ$ with 2° increments. The cumulative electron dosage was limited to $100\sim 110\text{ e}^-/\text{\AA}^2$ per tilt series. All $4\text{k}\times 4\text{k}$ frames were recorded on a Gatan K2 Summit direct electron detector in counting mode with the dose rate of $8\text{-}10\text{ e}^-/\text{pixel}/\text{s}$. For each tilt angle, a movie consisting of 7 to 8 frames was recorded. For the PCF samples, the nominal magnification was $\times 26,000$, giving rise to a calibrated pixel size of 6.102 \AA . The defocus value was targeted at $-4\text{ }\mu\text{m}$. 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 60 s to achieve a phase shift of $\sim 54^\circ$. Tilt series were recorded at a nominal magnification of $53,000\times$ (corresponding to a calibrated pixel size of 2.553 \AA) and a targeted defocus value of $-0.6\text{ }\mu\text{m}$. 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).

2.5.3 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 ($p_i, i=1\dots9$) 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:

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

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):

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

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. Step 1: 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.

2.5.4 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). *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).

2.5.5 Trypanosome motility videos

Motility videos of BSF cells were obtained exactly as described in (90). Motility videos of PCF cells were obtained exactly as described in (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). WT and mutant PCF videos correspond to this knockdown line cultured in the absence (WT) or presence (mutant) of tetracycline to induce RNAi.

2.6 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.

2.7 Acknowledgements

We thank Changlu Tao for technical assistance in SerialEM operation, Robert Minahan and Masahiro Yabe for help in data processing, Neville Kisalu and Michelle Shimogawa for motility videos of BSF cells. We thank Michelle Shimogawa for critical reading of the manuscript. This research has been supported in part by grants from NIH (R01GM071940, AI052348). SNF (P300PA_174358 and P2BEP3_162094). HK was supported by NIH-NRSA fellowship GM007185. We acknowledge the use of instruments in the Electron Imaging Center for Nanomachines supported by UCLA and grants from NIH (S10RR23057, S10OD018111 and U24GM116792) and NSF (DMR-1548924 and DBI-1338135).

2.8 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.

2.9 Figures

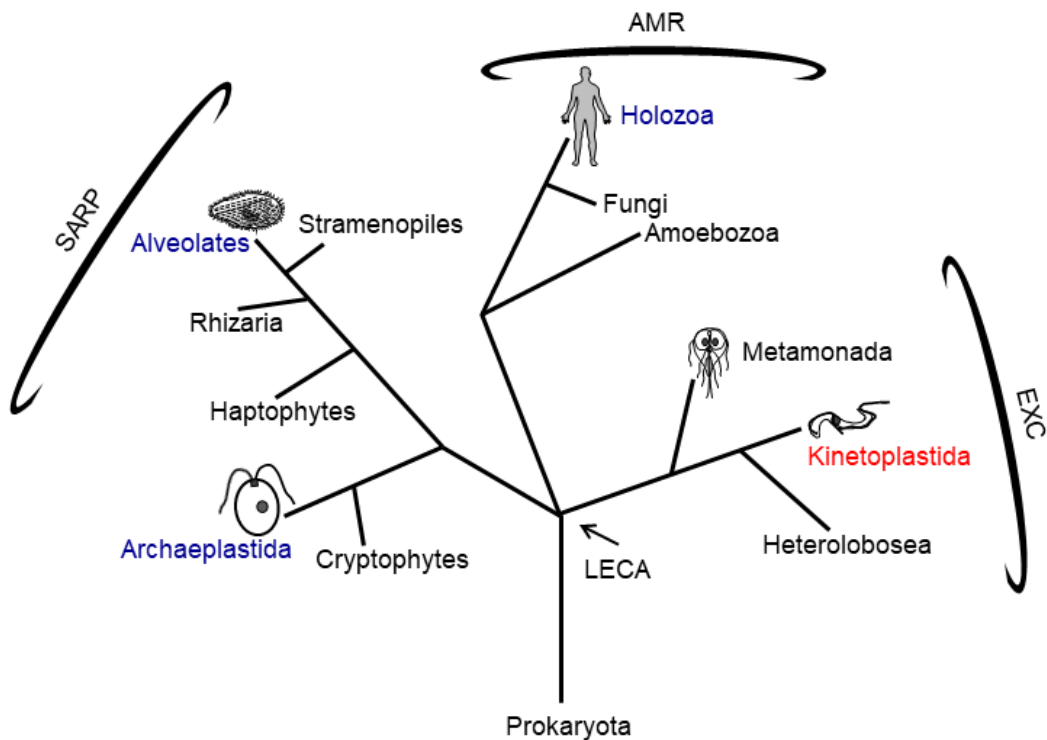


Figure 2- 1. Phylogenetic tree of eukaryotes

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.

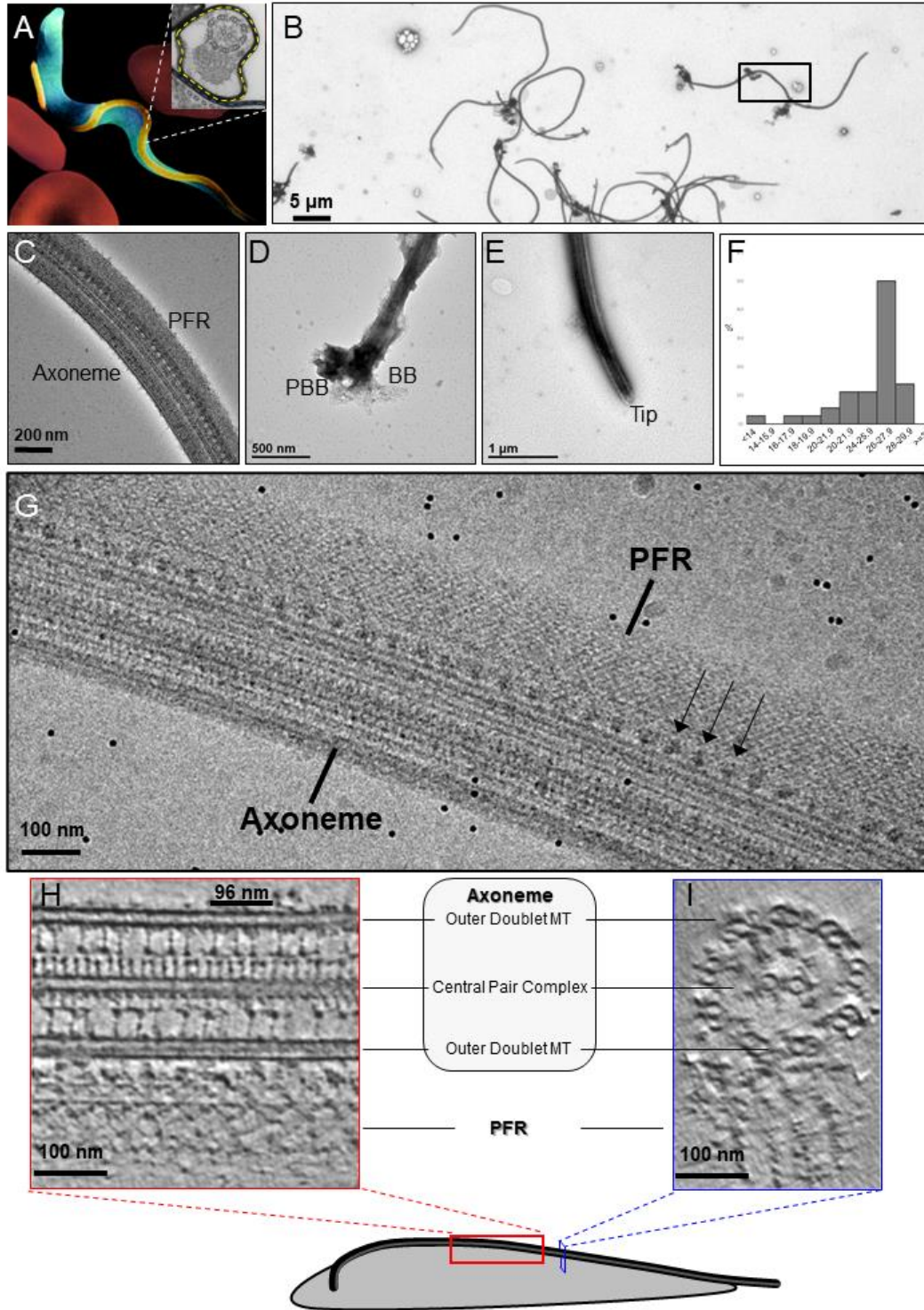


Figure 2- 2. Intact demembrated flagella from BSF *T. brucei*

(A) A representative scanning electron microscope image of a *T. brucei* parasite in blood from an 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.

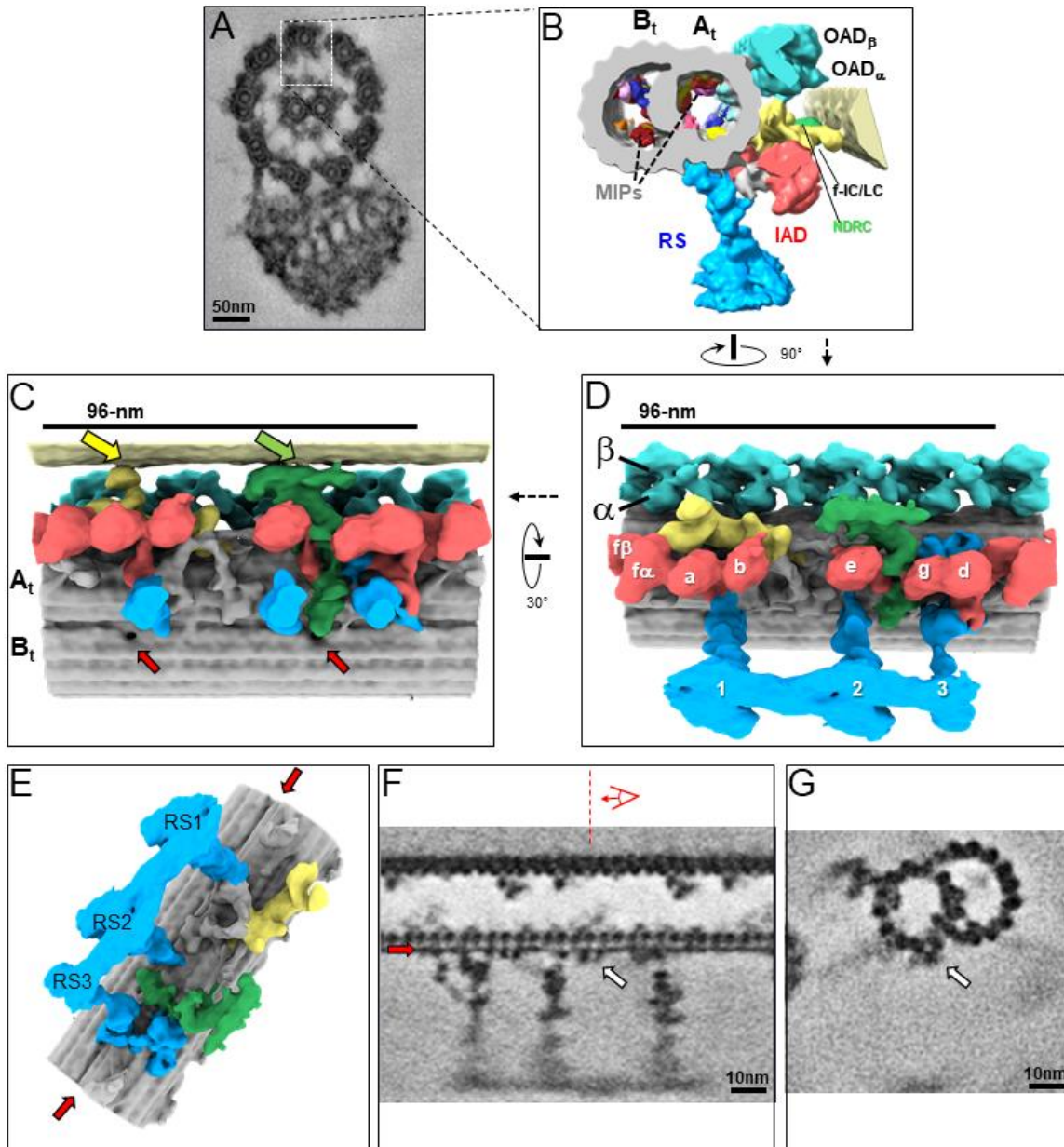


Figure 2- 3. The 3D ultrastructure of the 96-nm repeat from intact axonemes of BSF *T.*

brucei

(A) A representative cross-section of a demembrated 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 (\square) and beta (\square) 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.

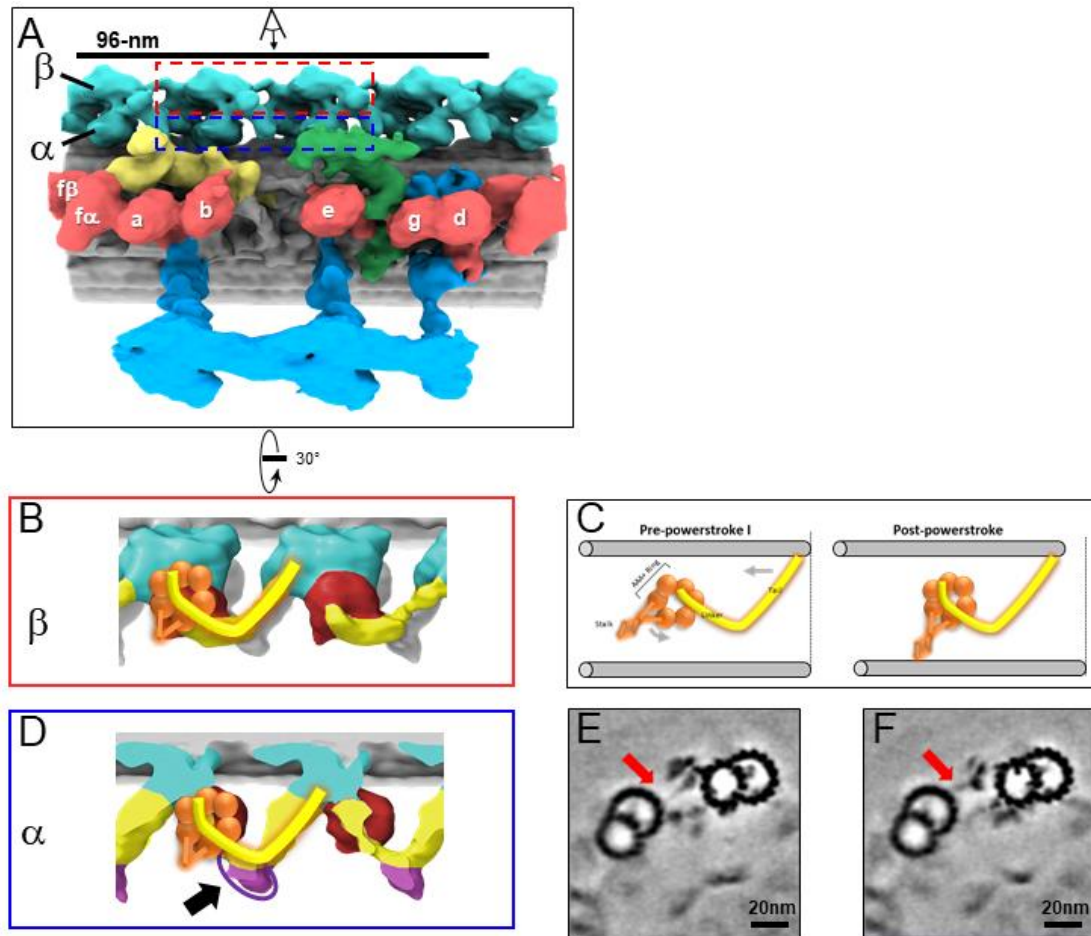


Figure 2- 4. *In situ* structure of outer arm dyneins and novel OAD-alpha inter-doublet connector in BSF *T. brucei*.

(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 β 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 OAD α dyneins. The linker and tail domains are colored in yellow and the AAA+ ring is colored in red. The arrow points to the OAD α 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 OAD α connector in (F), contacting the neighboring DMT.

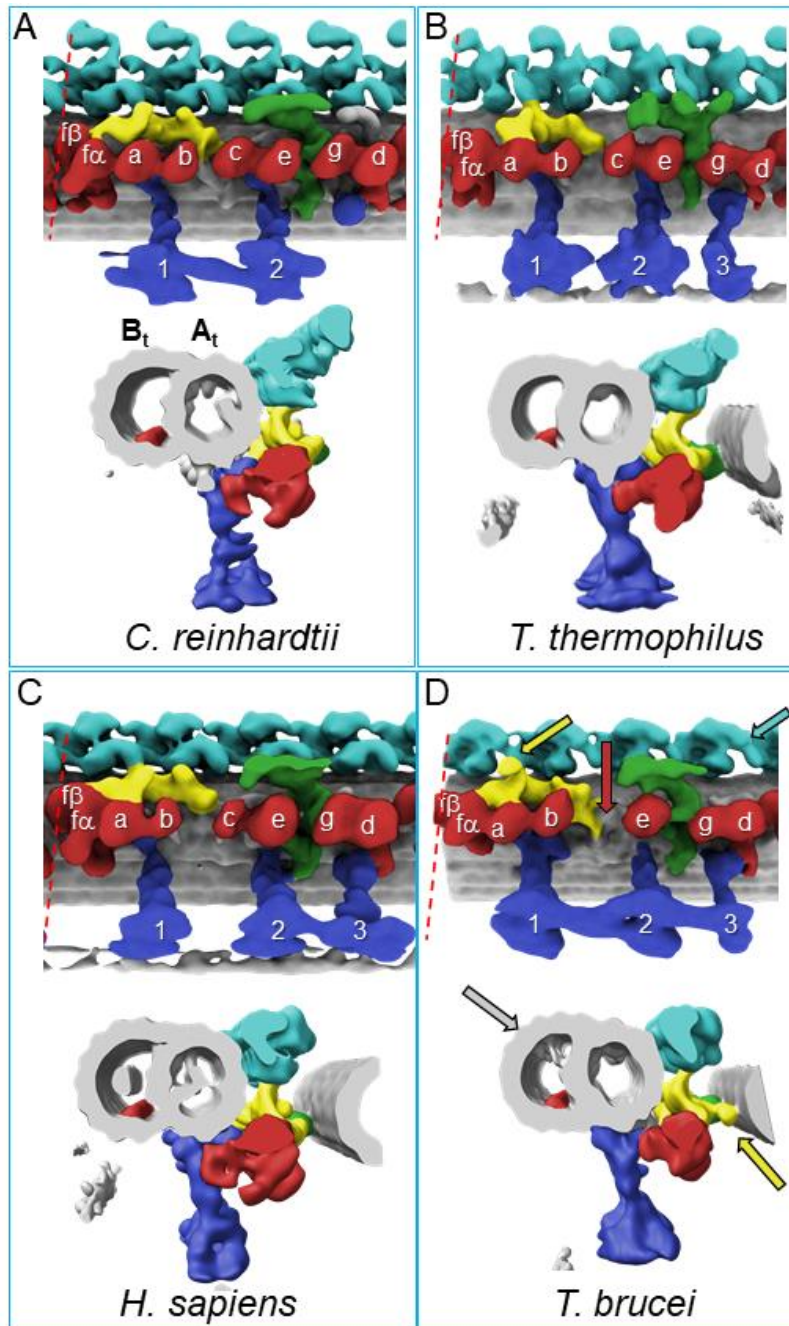


Figure 2- 5. Comparison of 96-nm axonemal repeat structures across species.

(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.

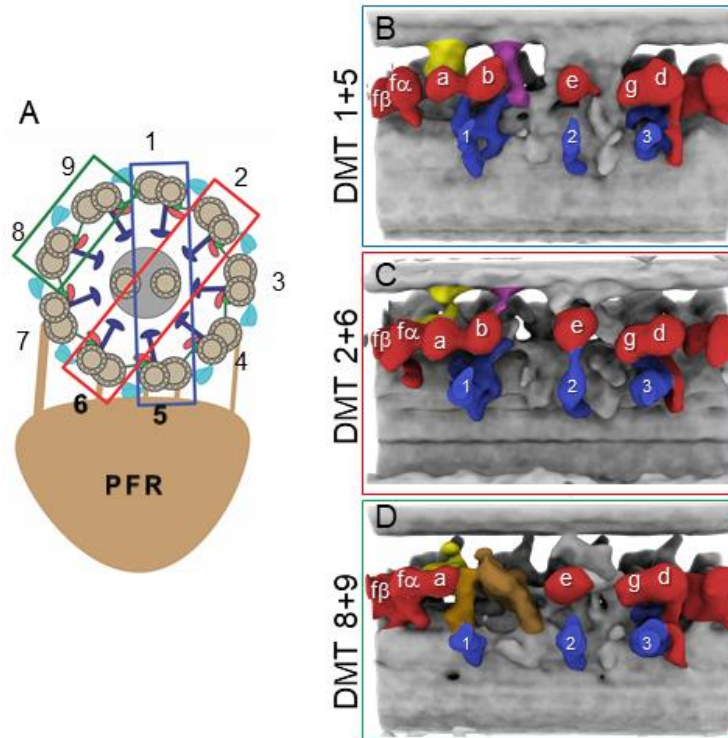


Figure 2- 6. Doublet-specific structures of the BSF *T. brucei* 96-nm repeat.

(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.

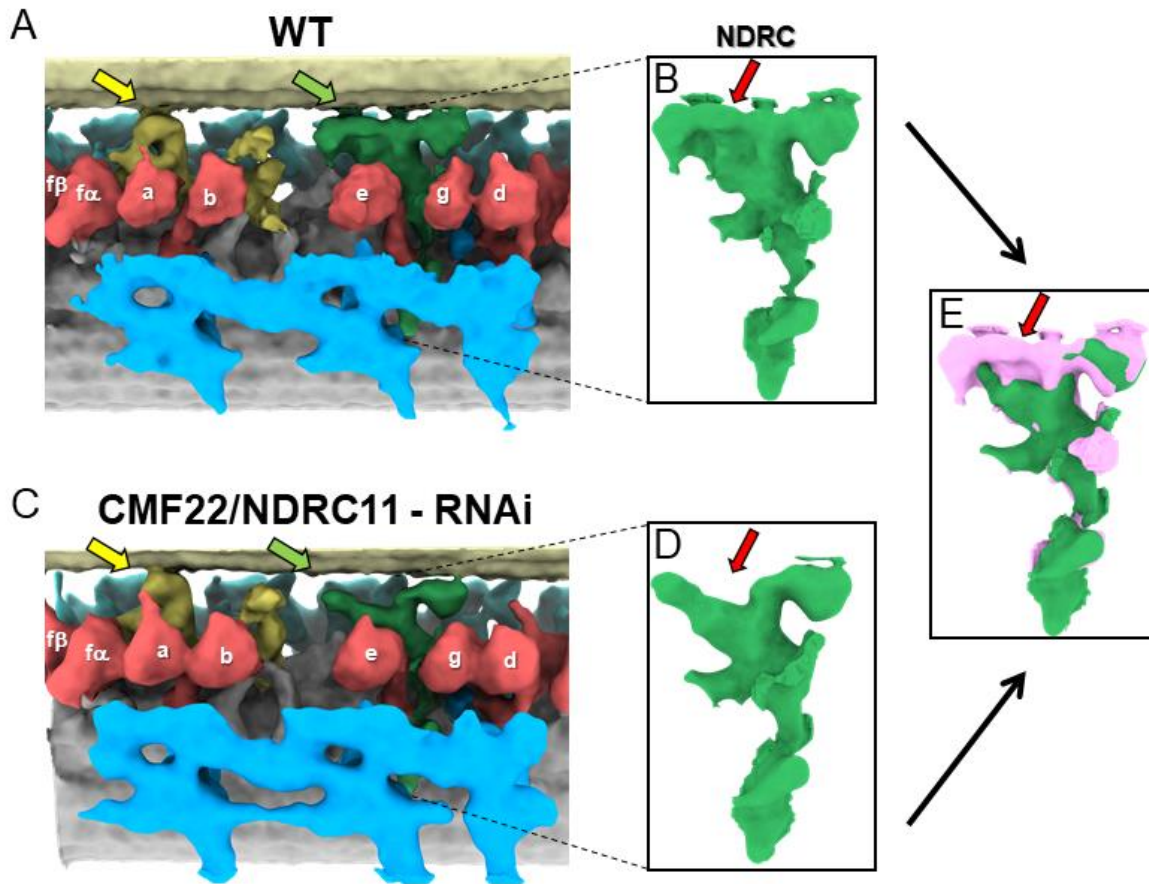


Figure 2- 7. Comparison between averaged 96-nm repeats of wild-type and CMF22/DRC11

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.

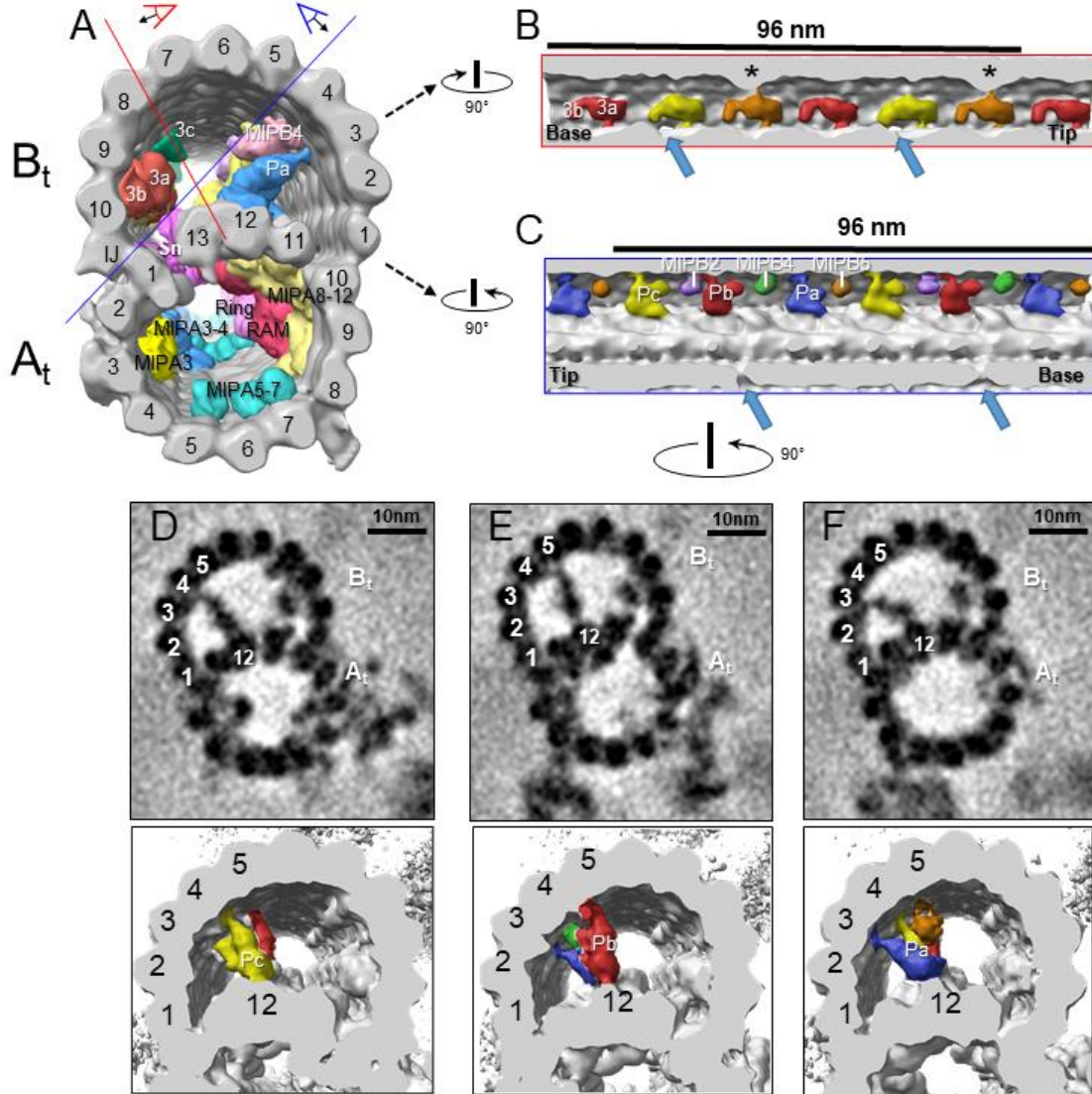


Figure 2- 8. TbMIP3 and ponticulus in the B-tubule of BSF *T. brucei*.

(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.

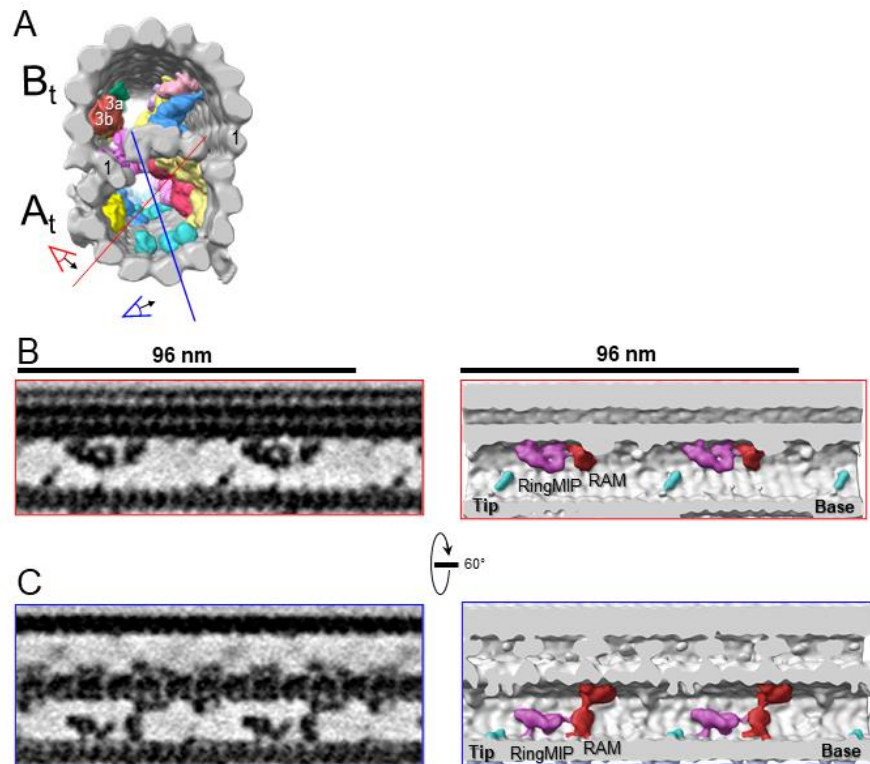


Figure 2- 9. The RingMIP and Ring Associated MIP (RAM) in the A-tubule of BSF *T.*

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.

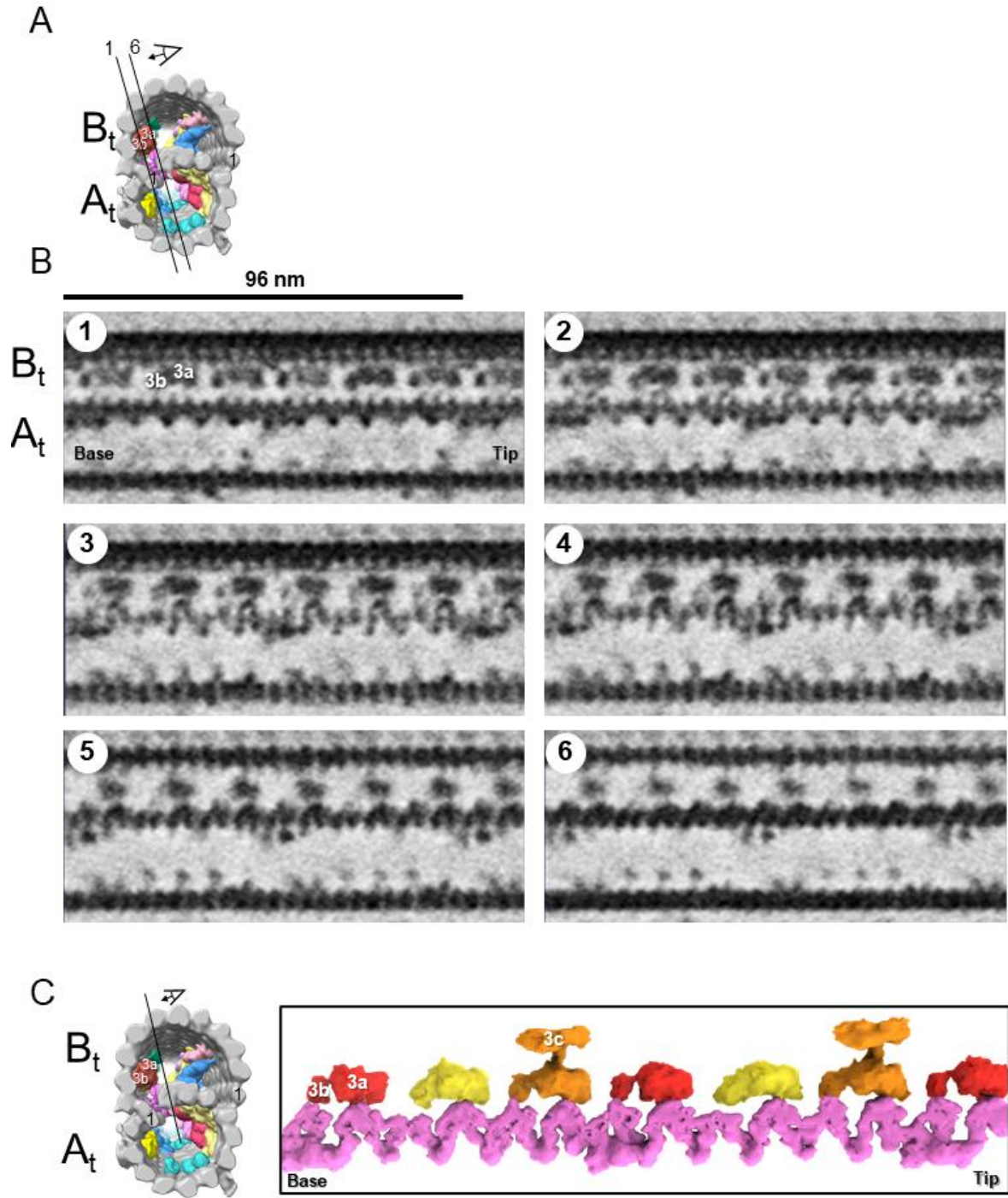


Figure 2- 10. The snake MIP connects the A-tubule and the B-tubule of BSF *T. brucei*.

(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. The right panel shows segmented TbMIP3 (red, yellow and orange, as described for Figure 8B) and Snake MIP (mauve). (See also Video 7.)

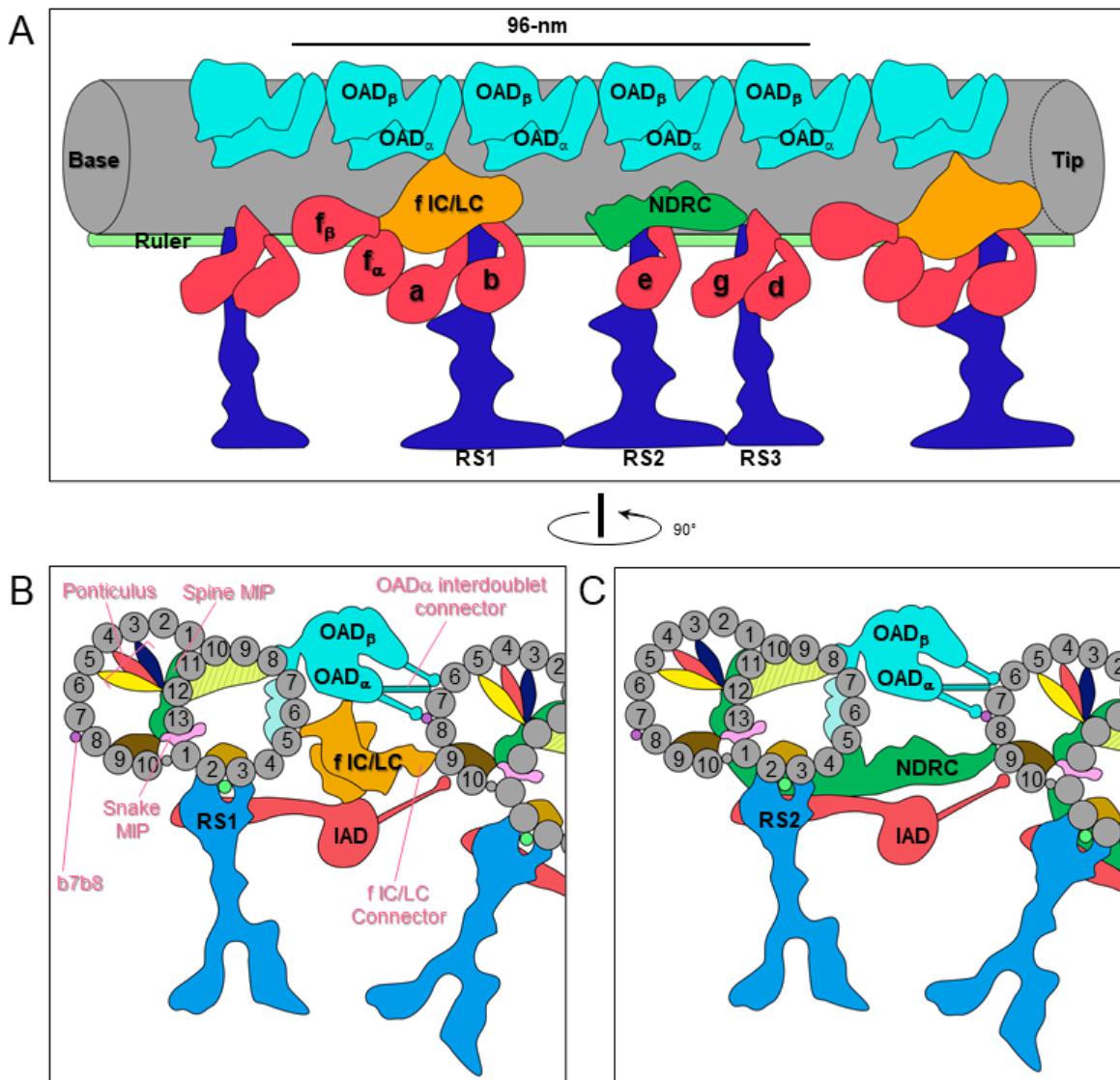
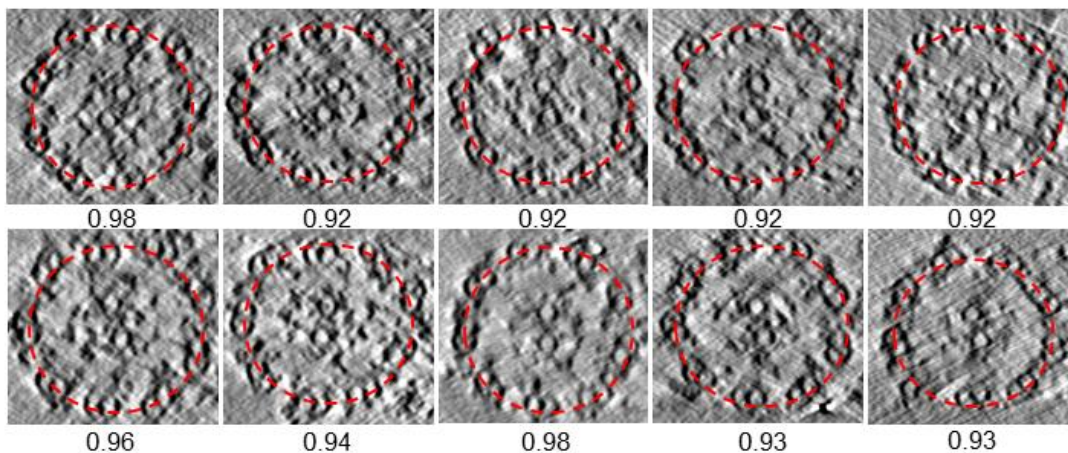


Figure 2- 11. Schematic overview of the trypanosome axoneme.

(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.

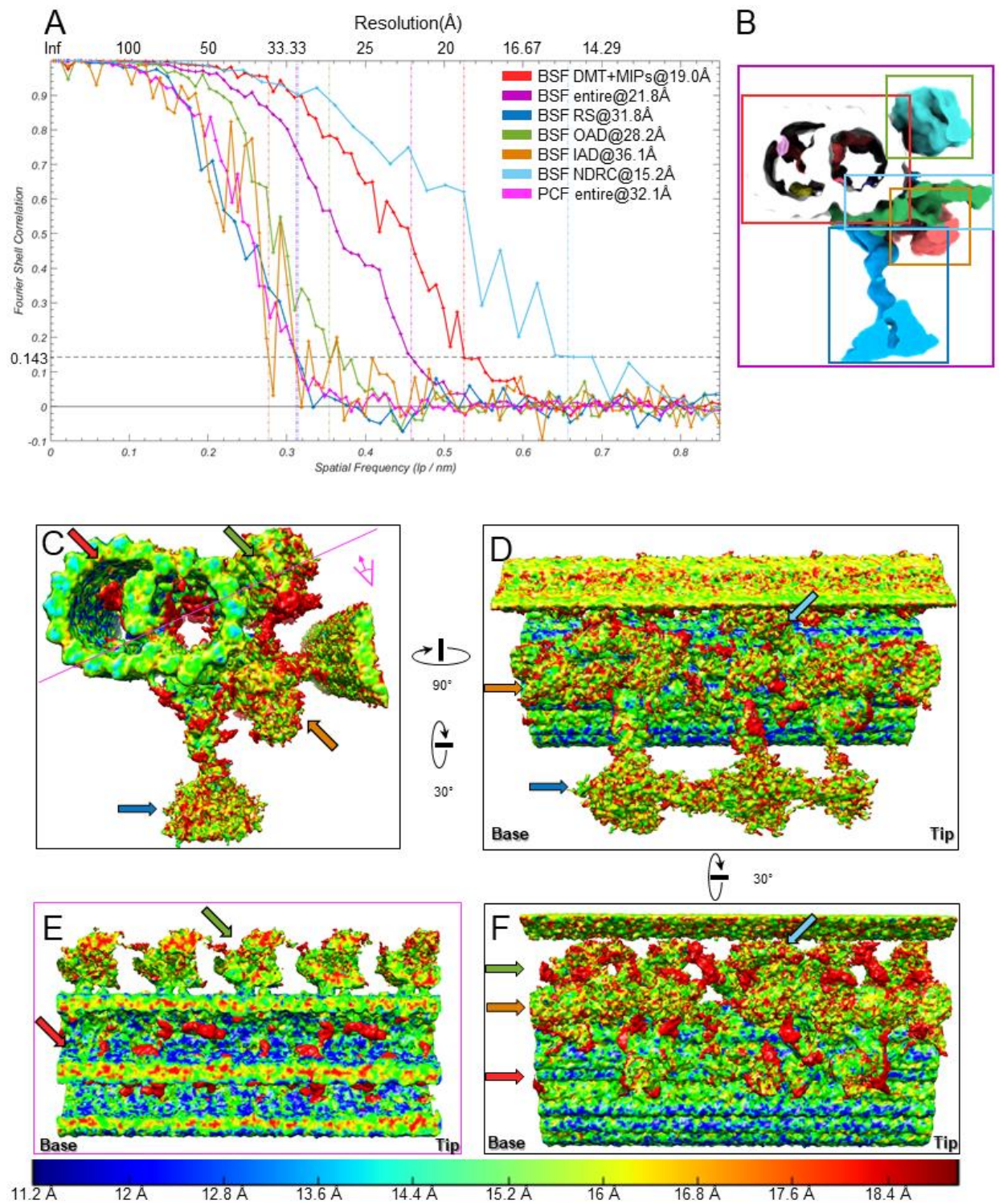
(B and C) Cartoon cross-section view of the axoneme (viewed from the proximal end) at roughly the position of RS1 (B) and RS2 (C). Protofilaments are numbered and structures are labelled as for panel A. Major trypanosome-specific structures described in text are labelled in pink. Note that additional *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.

2.10 Supplemental Figures



Supplemental Figure 2- 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.

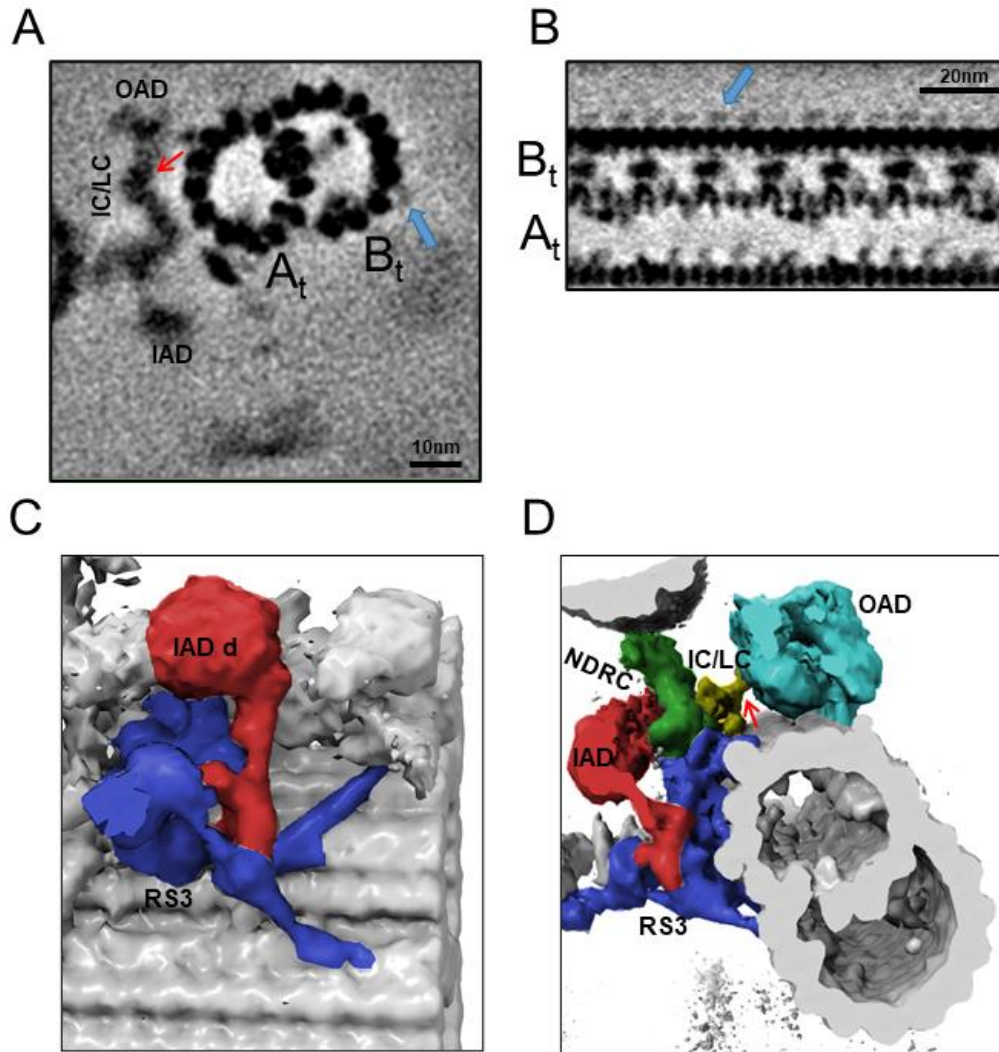


Supplemental Figure 2- 2. Gold Standard Fourier shell correlation (FSC) and ResMAP analyses. [related to Fig. 3]

(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. The 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 2- 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.

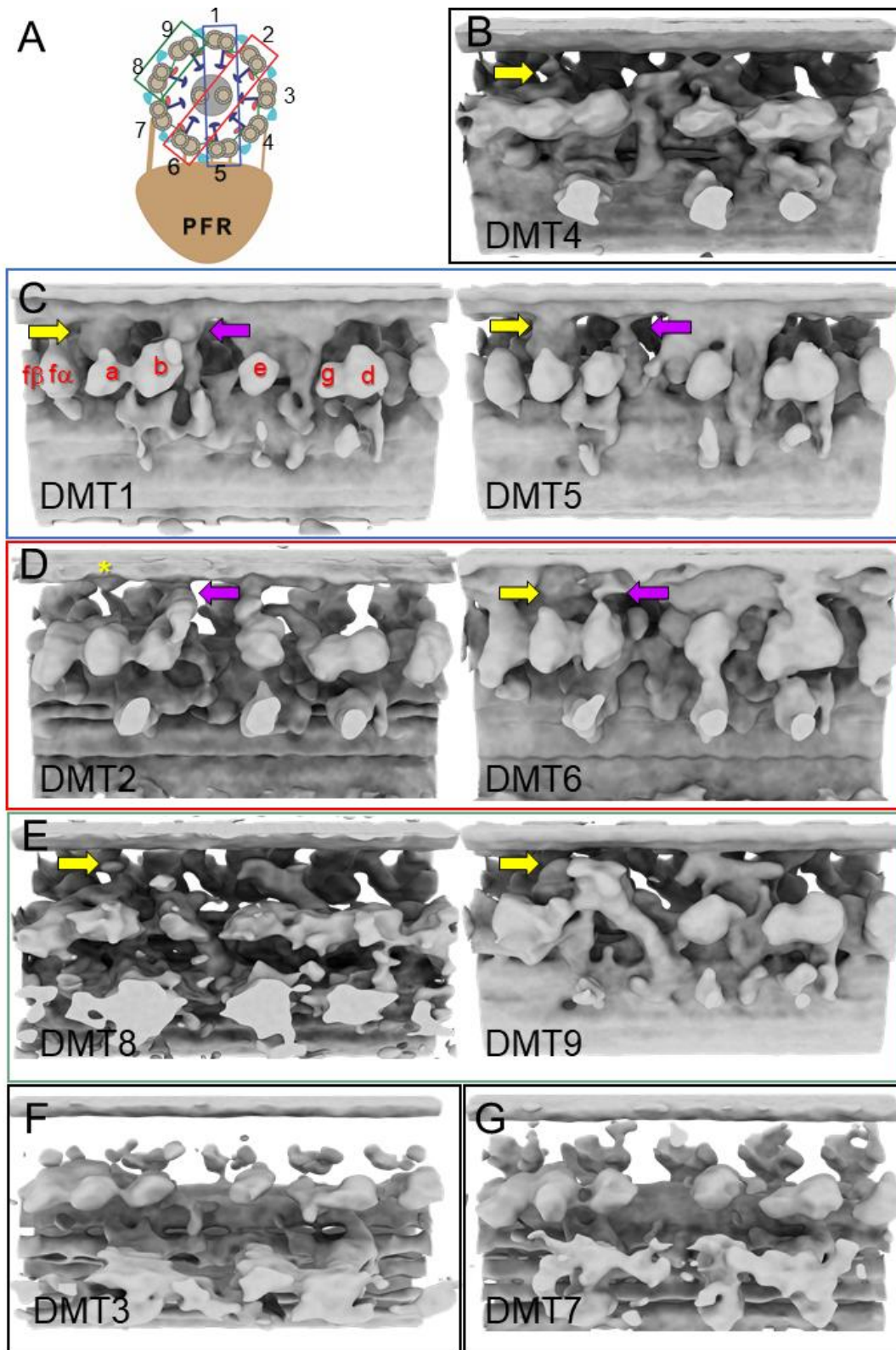
(B) Longitudinal section through the BSF averaged density map, with the proximal end of the axoneme at the left. A and B-tubules are labeled and blue arrow points to the novel density outside protofilaments b7 and b8.

(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.

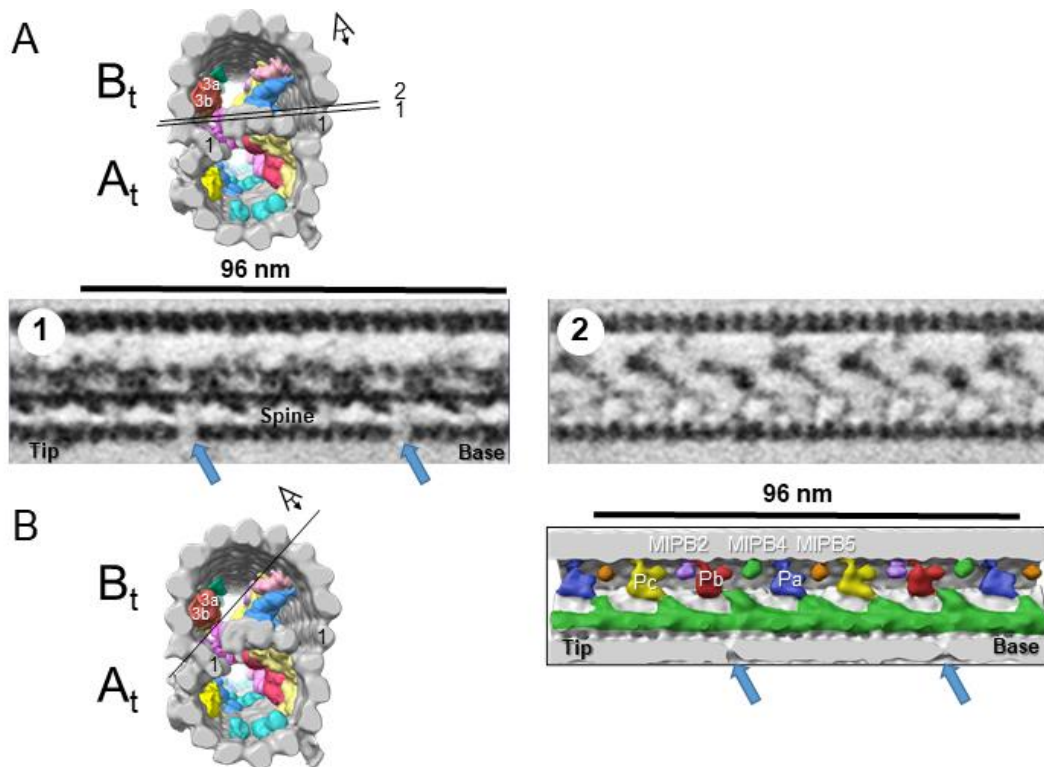


Supplemental Figure 2- 4. Sub-tomogram averages of the 96-nm repeat of individual DMTs

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.

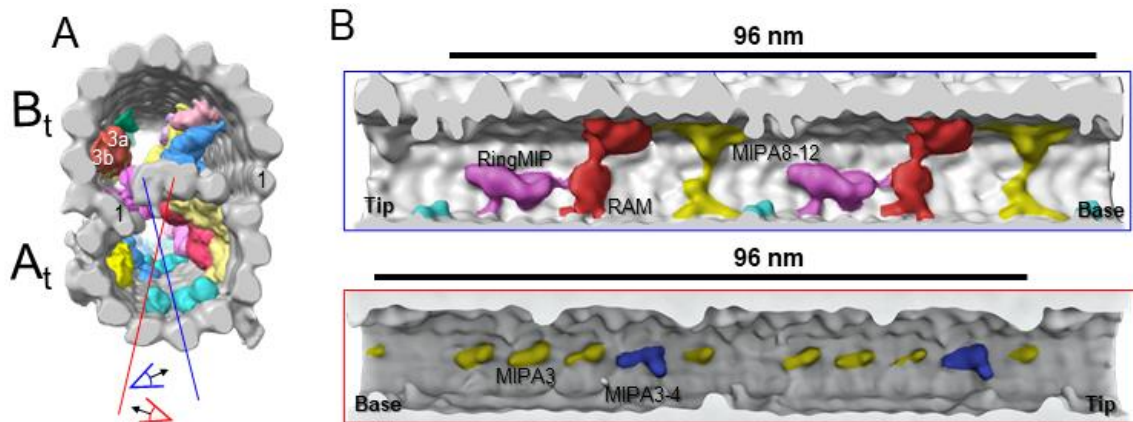


Supplemental Figure 2- 5. The Spine MIP is a contiguous structure, spanning 48 nm and contacting adjacent MIPs in BSF *T. brucei*. [related to Fig. 8]

(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).

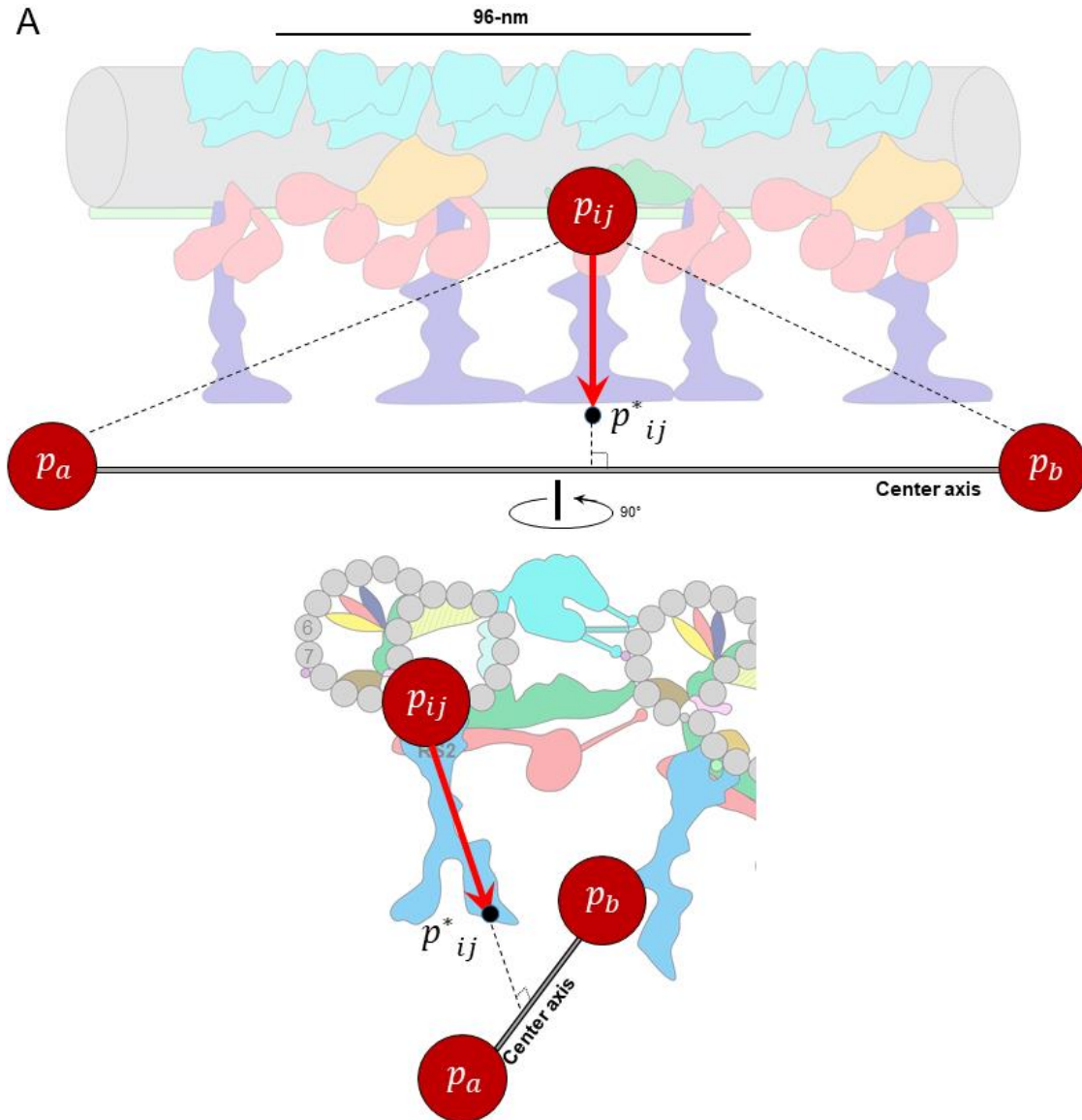


Supplemental Figure 2- 6. MIPs in the A-tubule of BSF *T. brucei*. [related to Fig. 9]

(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.

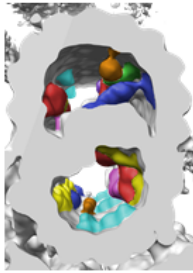


Supplemental Figure 2- 7. Illustration of the principles of *autoPicker*. [related to Fig. 3]

(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 perpendicular to $\overrightarrow{p_a p_b}$, and parallel to the plane defined by the

three points of p_a , p_b and p_{ij} .

2.11 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 2- 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.

2.12 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 Archaeella, 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, Morpew 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.

Chapter 3: Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells

Hui Wang^{1, 3, 4, *}, Jiayan Zhang^{1, 2, 3, *}, Simon Imhof¹, Xueting Zhou¹, Shiqing Liao¹, Ivo Atanasov³, Wong H Hui³, Kent L. Hill^{1, 2, 3, @} & Z. Hong Zhou^{1, 2, 3, @}

* co-first authors contribute equally to this project.

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

²Molecular Biology Institute, UCLA, Los Angeles, California 90095, USA

³California NanoSystems Institute, UCLA, Los Angeles, California 90095, USA

⁴Department 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)

3.1 Abstract

Eukaryotic flagella (synonymous with cilia) rely on a microtubule-based axoneme, together with accessory filaments to carry out 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

3.2 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 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 averaging⁵¹, 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

tensegrity⁵², rather than space-filling observed in other cellular structures⁵³, as the means to achieve competing needs of 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.

3.3 Results

3.3.1 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 recording²⁶ and relied on sub-tomographic averaging with wedge-mask differences⁵⁵. 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 convention⁵⁶ (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 distal³⁰, 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.

3.3.2 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 PFR⁴⁸⁻⁵⁰, 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 $\sim 28.5\text{\AA}$ 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 angles 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. 2h₂) nor orthogonal to dimension \vec{a} . Thus, the unit cell is neither orthorhombic⁴⁸, nor helical⁴⁴. Rather, when examining from the top of the axoneme down to the PFR (Figs. 2f, h₂), dimension \vec{b} is oriented approximately 45° from the axoneme axis, which may account for the $\sim 45^\circ$ 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 PFR⁵⁰. Density scissors stack vertically upon one another along dimension \vec{a} (Figs. 2c, h₃), 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 20 μm 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 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 identified⁴², 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 ultrastructure⁶³, indicating they are not required for assembly of the main PFR structural elements.

3.3.3 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 flagella⁴⁹ (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.

3.3.4 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. 5b₂-b₅): 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. 5b₂, b₆) and structural details of these filaments become smeared in the sub-tomographic average of PAC7 (Figs. 5b₃, b₄). 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. 5b₂, b₆), 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, l), which connect to OAD, and protofilaments B1, 2 and B5 of DMT5, respectively (Fig. 5c₁; Supplementary Fig. S3). These results support an earlier report that PFR may influence motility through direct interaction with dynein⁴⁹. 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. 5d₁; Supplementary Fig. S3). PAC4 has a periodicity of 54nm along the flagellum (Fig. 5d₂), 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. 5d₃), 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.

3.3.5 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 abnormalities⁶⁷ 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 bridge⁶⁹ 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* CPC⁶⁹ (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.

3.4 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 filaments²³. However, helical waves are common among microbes, including important pathogens^{5, 8, 36, 71}, and even occur in human sperm⁷². Non-planar helical motion may contribute to microbial pathogenesis¹², as it is recognized to facilitate propulsion through viscous environments⁷³, 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 $H(t) = A \cdot \sin(\omega t) \cdot \vec{i} + A \cdot \cos(\omega t) \cdot \vec{j} + v \cdot t \cdot \vec{k}$ where A , ω , v are amplitude, rotational speed and forward speed respectively, and \vec{i} , \vec{j} , \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 resistance²¹ 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 assembly⁸¹⁻⁸³, 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. 5c₁, c₂)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}.

3.5 Materials and Methods

3.5.1 Sample preparation and cryoET

T. brucei bloodstream form single marker (BSSM) cells⁸⁷ were used, and details for culturing, flagella isolation, and cryoET were described previously²⁶. Briefly, demembrated 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 package⁹⁰. The contrast transfer function (CTF) was determined by *ctffind4*⁹¹ and corrected with the *ctfphaseflip* program⁹² 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.

3.5.2 Missing-wedge compensation

Demembrated *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.

3.5.3 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 *stalknit* 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 following: 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).

3.5.4 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.

3.5.5 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.

3.6 Acknowledgments

This research has been supported in part by grants from NIH (AI052348 to KLH and GM071940/DE028583 ZHZ) and fellowships from Swiss National Foundation (P300PA_174358 and P2BEP3_162094 to SI). We acknowledge the use of instruments in the Electron Imaging Center for Nanomachines supported by UCLA and grants from NIH (S10RR23057, S10OD018111 and U24GM116792) and NSF (DMR-1548924 and DBI-1338135). JZ was supported by UCLA Whitcome Pre-doctoral Fellowship (JZ), and the UCLA Dissertation Year Fellowship (JZ).

3.7 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.

3.8 Conflict of interest

The authors declare that they have no conflict of interest.

3.9 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.

3.10 Figures

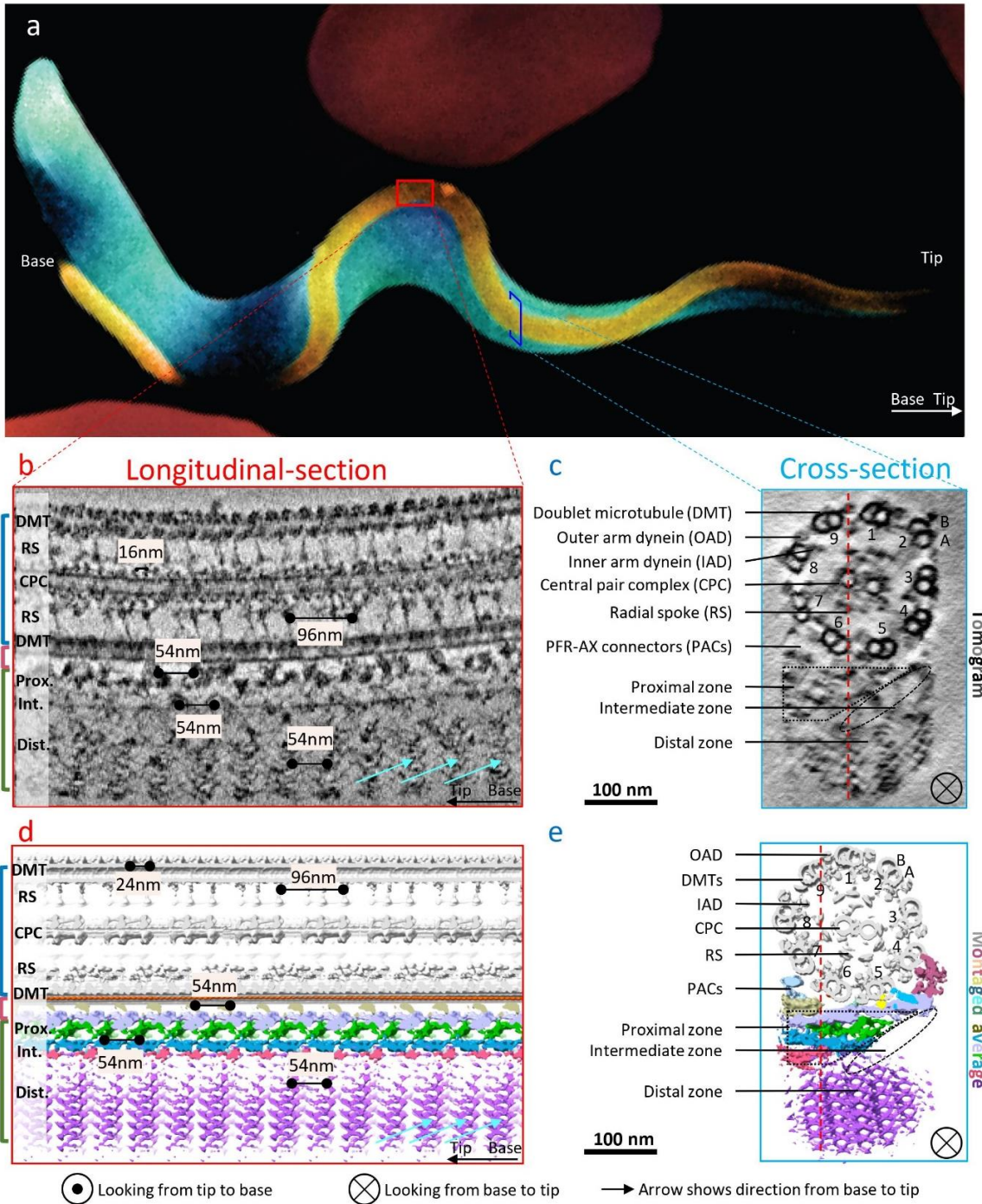


Figure 3- 1. CryoET of *T. brucei* flagellum in its bloodstream form.

(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.

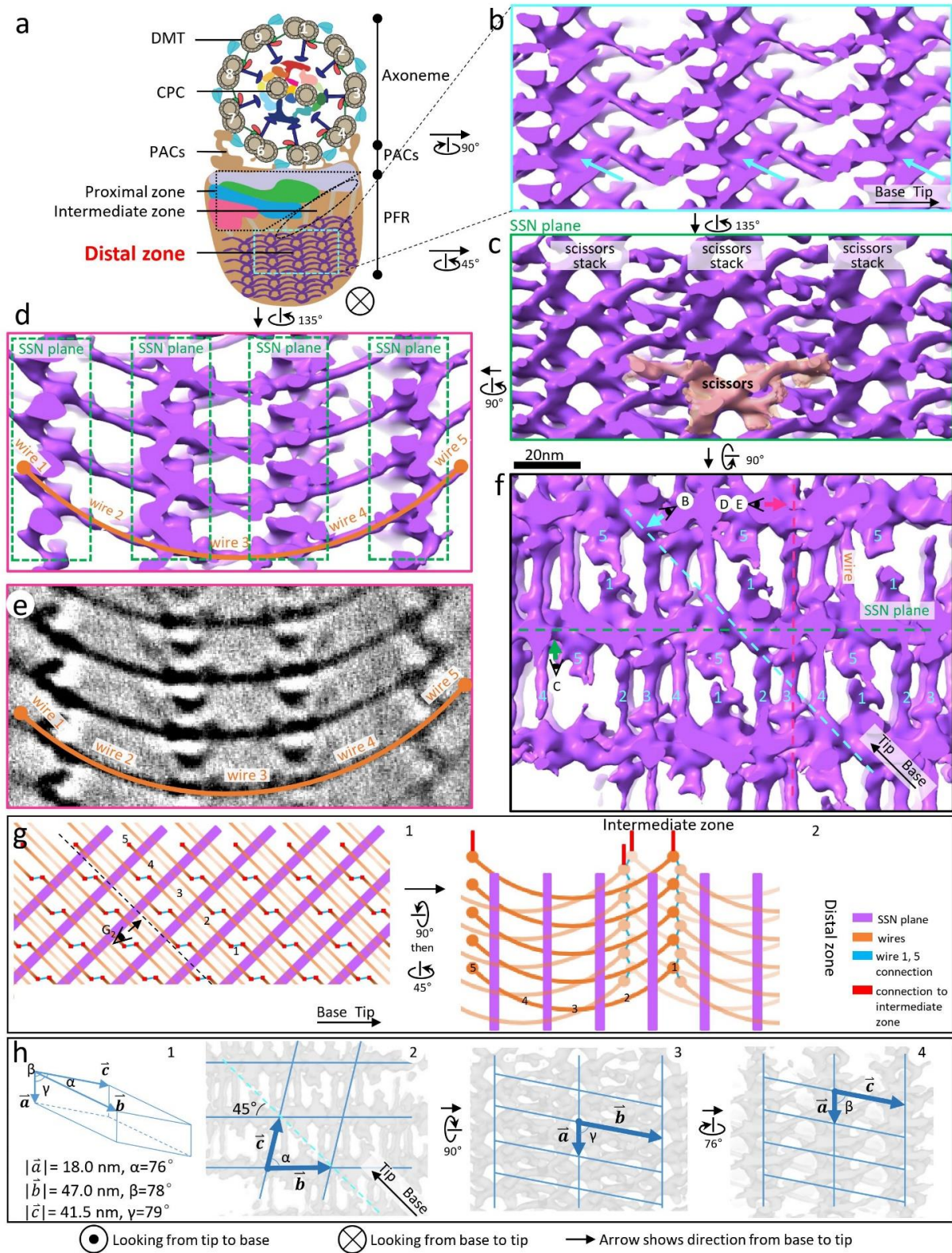


Figure 3- 2. Structure of the paracrystalline PFR distal zone.

(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 3- 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.

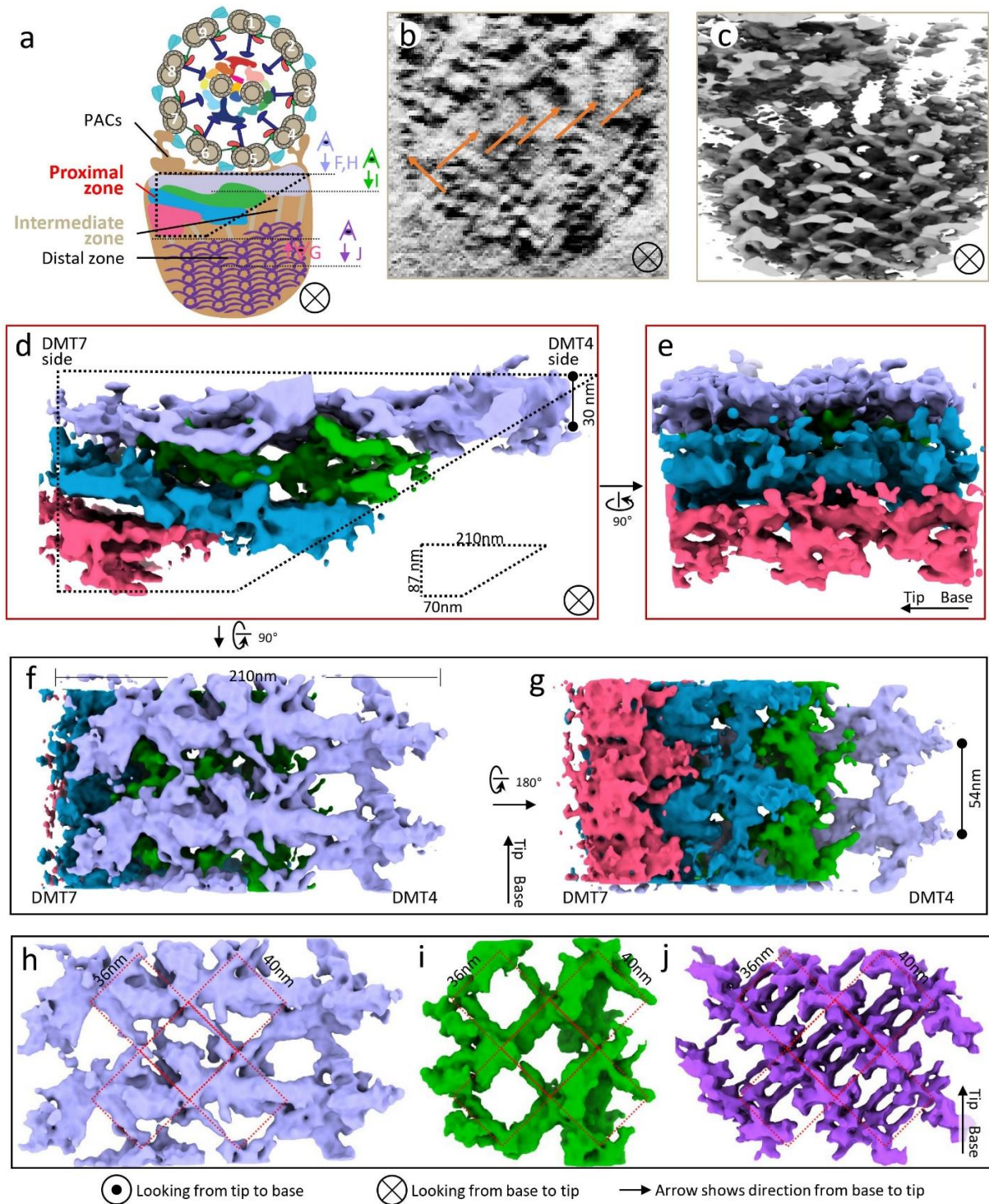


Figure 3- 4. Structure of the proximal zone and its structural continuity with the intermediate and distal zones

(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.

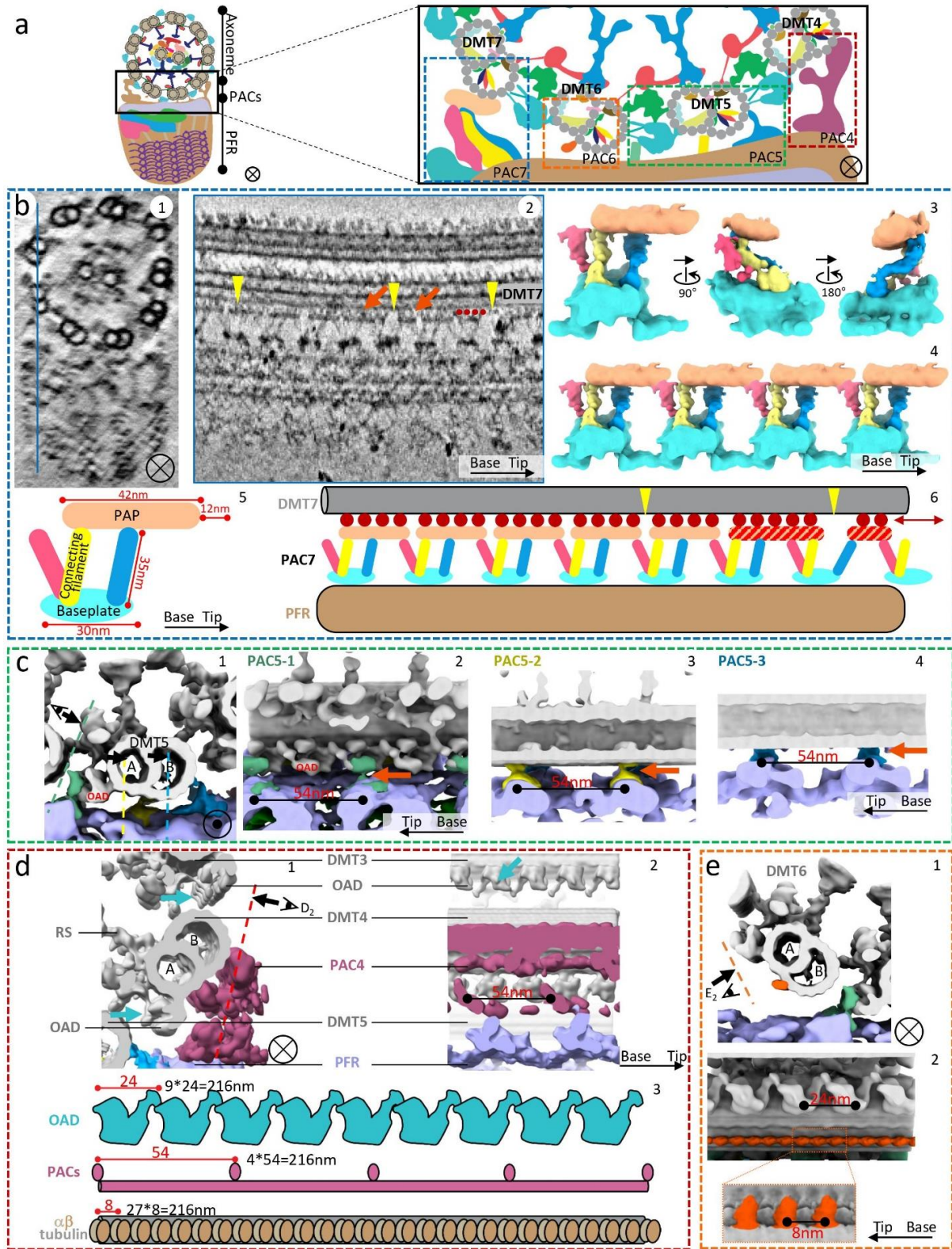


Figure 3- 5. Structures of PFR-axoneme connectors (PACs)

(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 e_1 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).

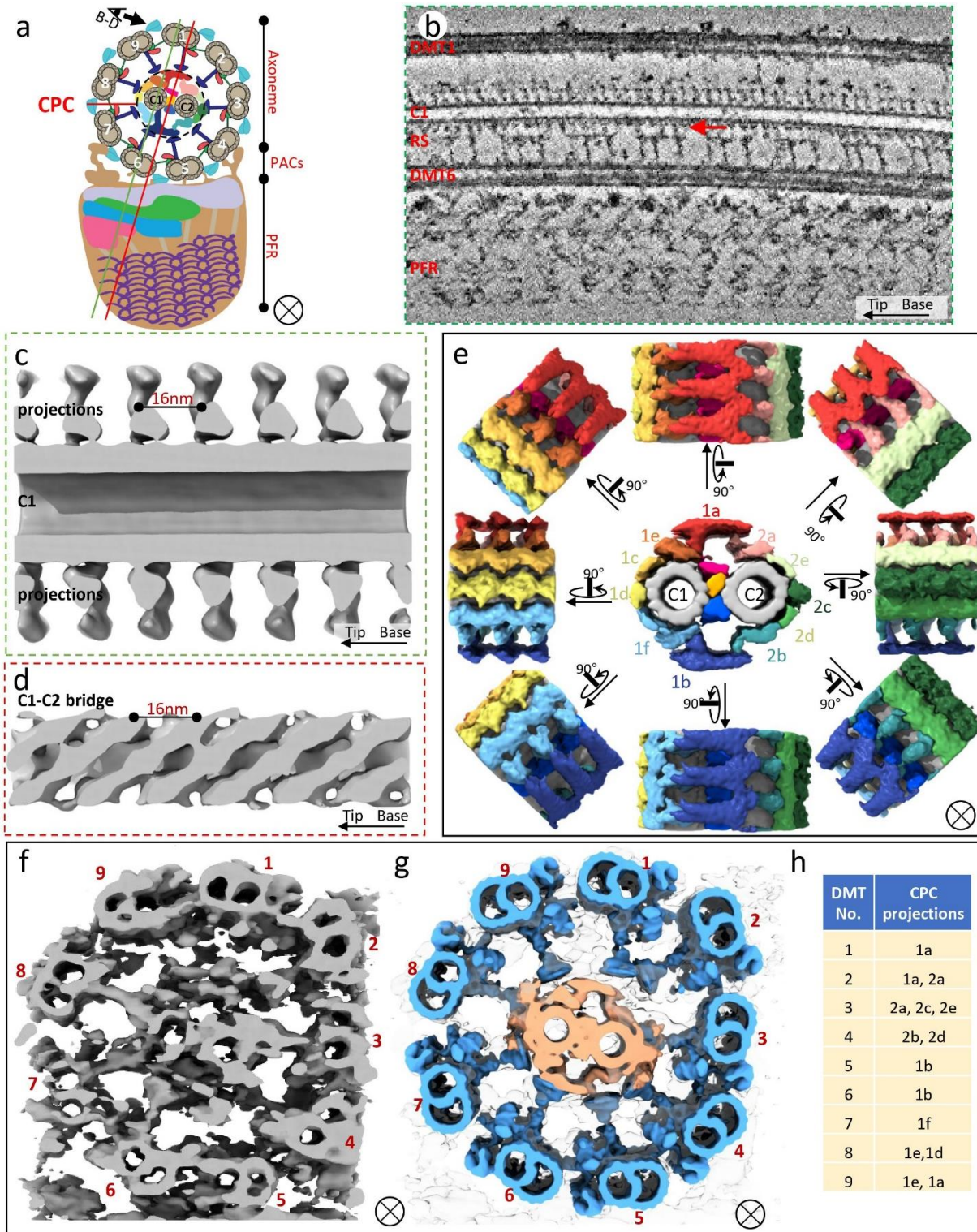


Figure 3- 6. Structure of the *T. brucei* central pair complex (CPC)

(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).

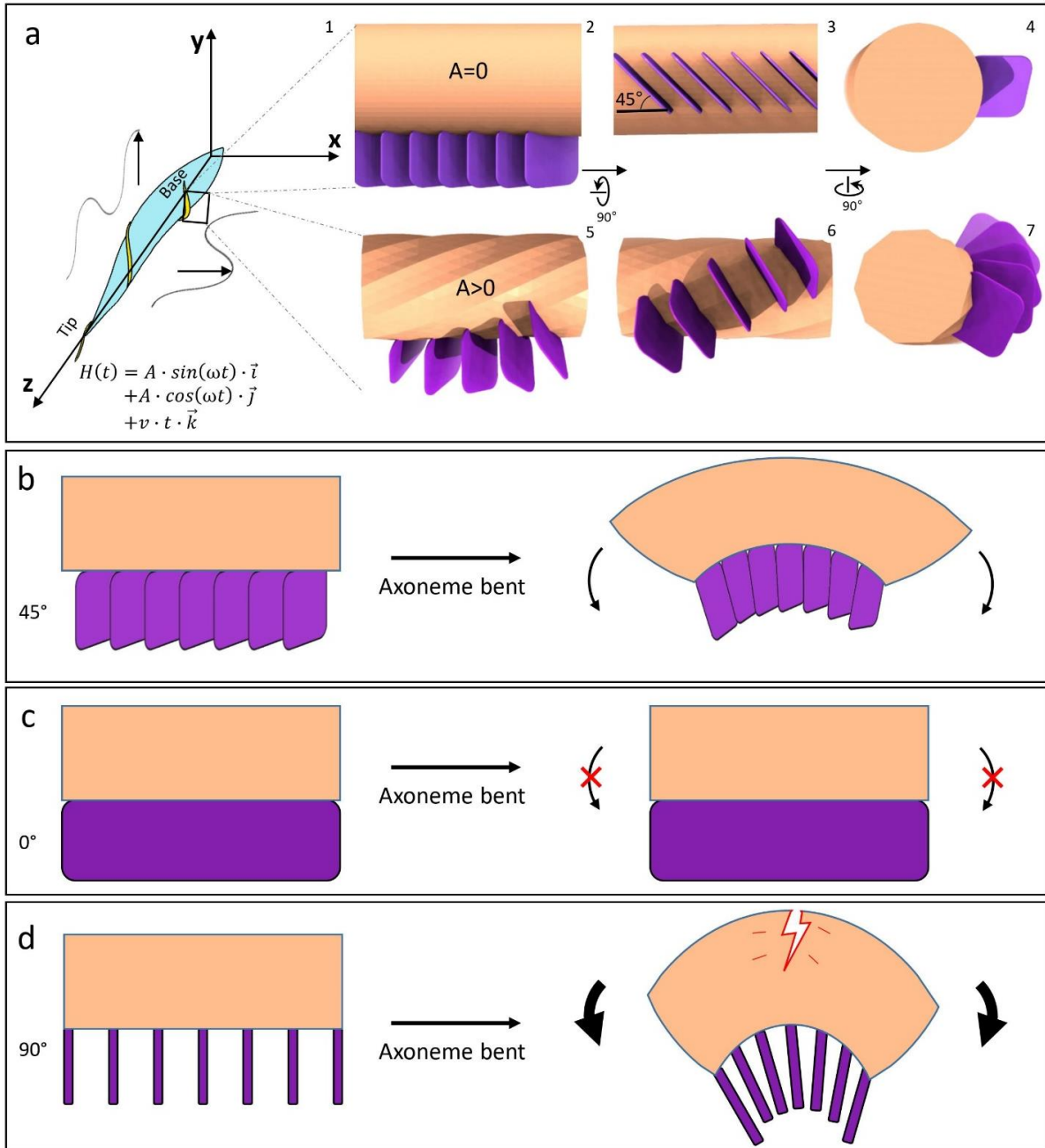
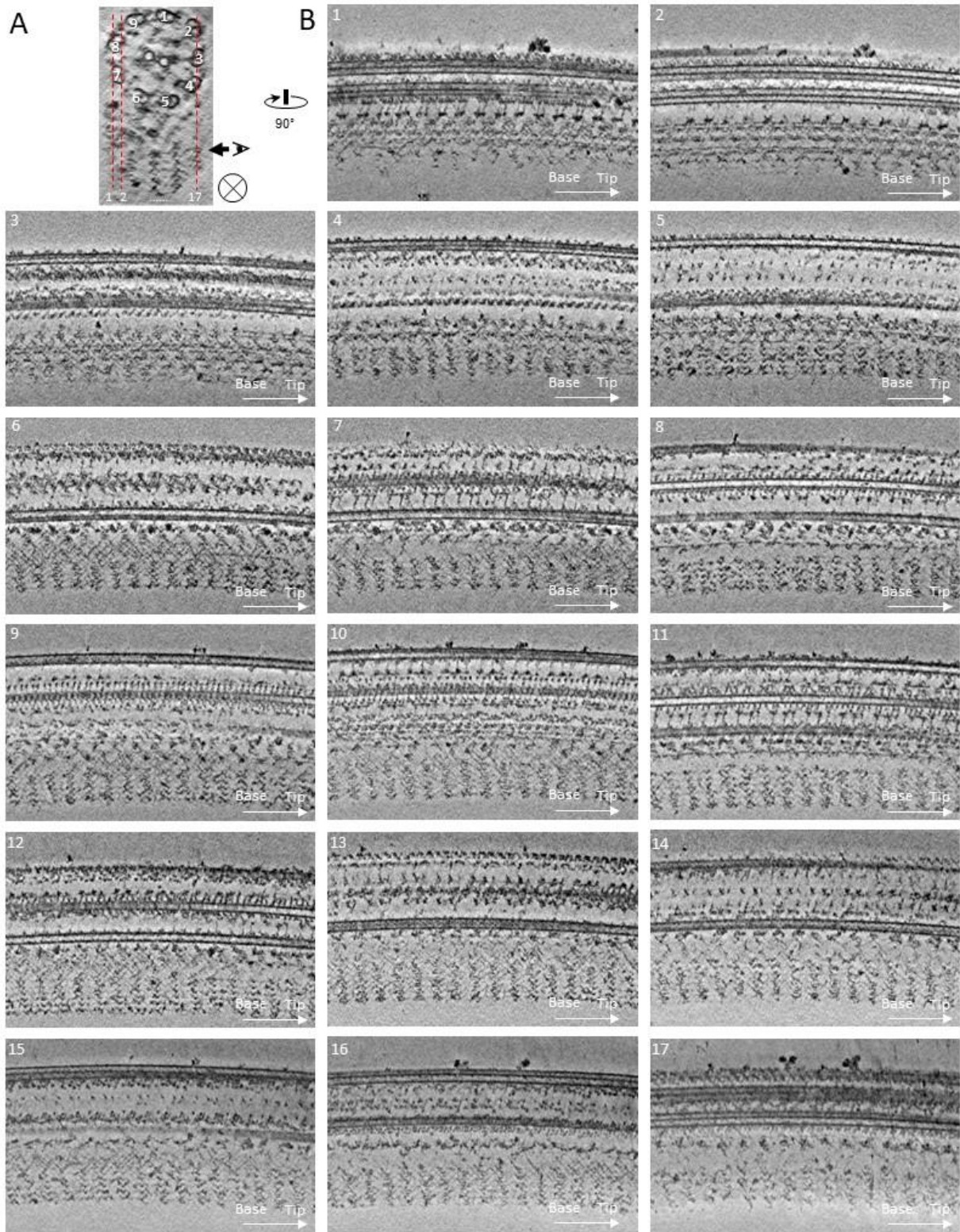


Figure 3- 7. Model for SSN planes supporting non-planar helical wave of the *T. brucei* flagellum

(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, a₁). $\vec{i}, \vec{j}, \vec{k}$ are unit vectors along x, y, and z axis, respectively. The right panels (a₂₋₇) 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 (a₅₋₇). (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.

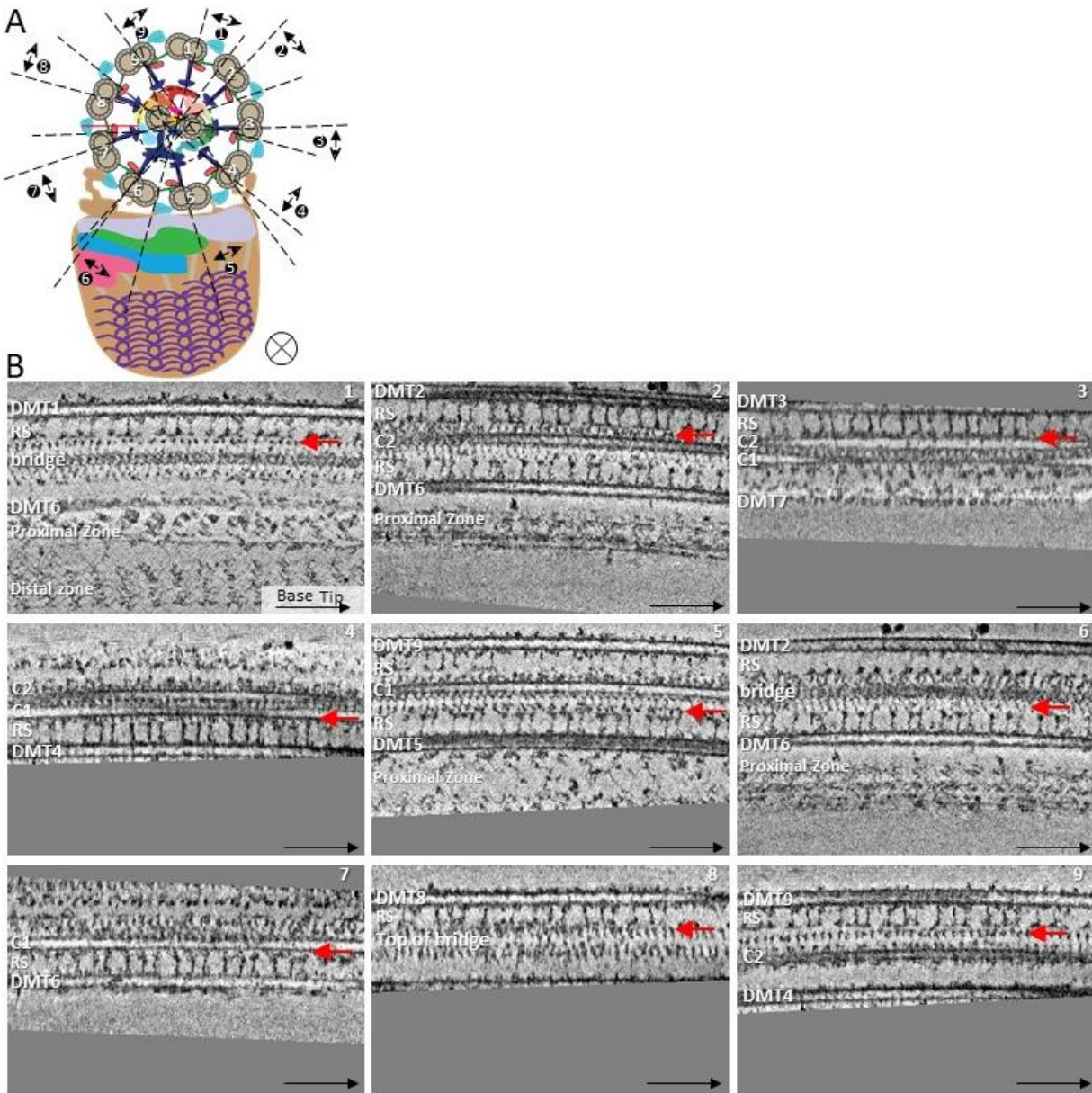
3.11 Supplemental Figures



Supplemental Figure 3- 1. Density slices through a representative tomogram.

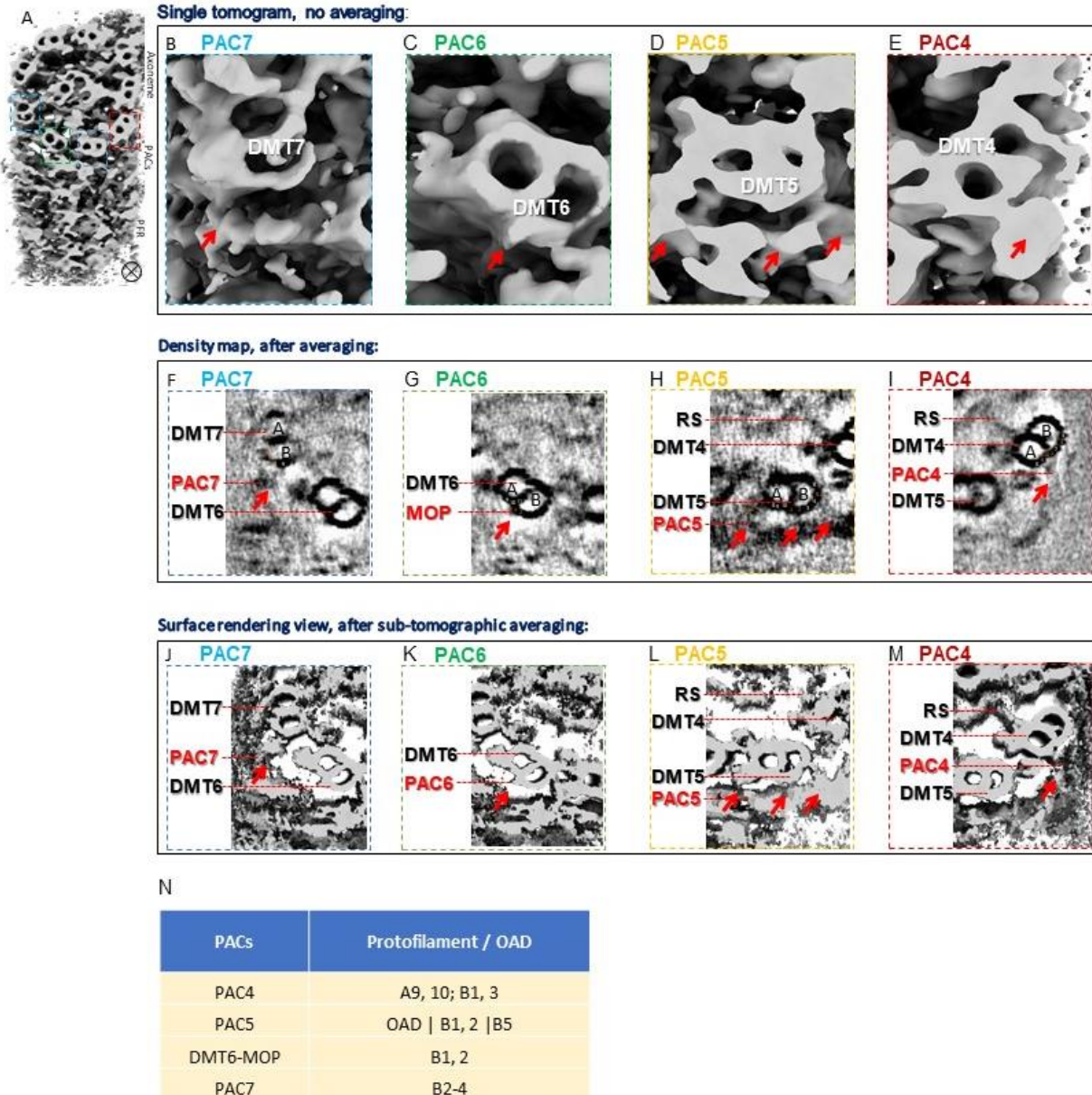
(a) Cross section views with DMT numbered.

(b) Seventeen evenly spaced longitudinal sections as indicated by red lines in a.



Supplemental Figure 3- 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.



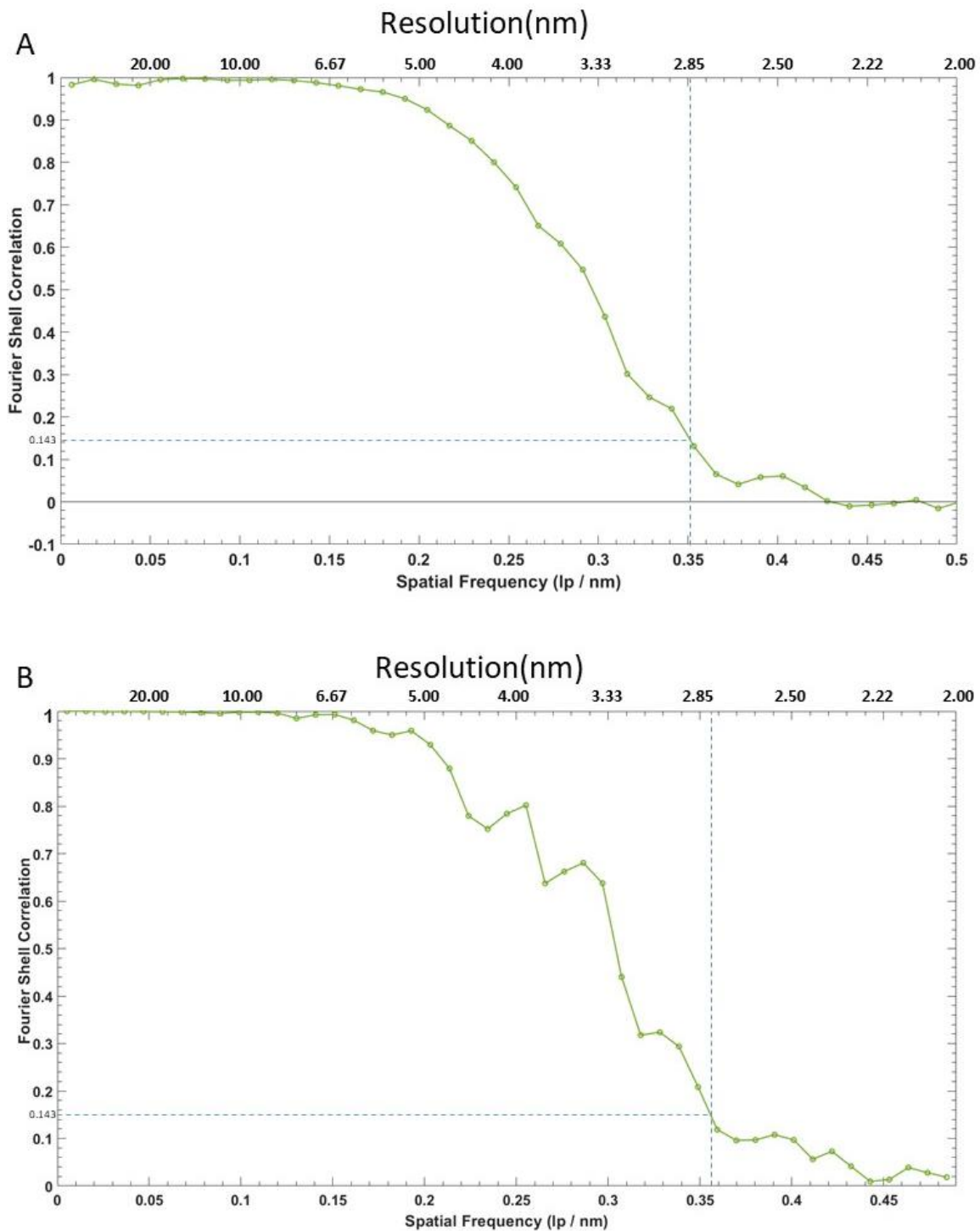
Supplemental Figure 3- 3. PAC contacts with the axoneme.

(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 numbered according to convention.

(n) Summary of contacts between PACs and the axoneme (see also Fig. 5).



Supplemental Figure 3- 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.

3.12 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 reactions 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. & Mastrorade, 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. Schlaeppli, 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).

Chapter 4: Isotropic Reconstruction of Electron Tomograms with Deep Learning

Yun-Tao Liu^{1,2,3,8}, Heng Zhang^{1,4,8}, Hui Wang^{2,3,5}, Chang-Lu Tao^{1,6}, Guo-Qiang Bi^{1,6,7,*} and Z. Hong Zhou^{2,3,5,*}

¹Center for Integrative Imaging, Hefei National Laboratory for Physical Sciences at the Microscale, and School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230026, China;

²California NanoSystems Institute, University of California, Los Angeles, CA, USA;

³Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, CA, USA;

⁴Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

⁵Department of Bioengineering, UCLA, Los Angeles, USA;

⁶Interdisciplinary Center for Brain Information, Brain Cognition and Brain Disease Institute, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

⁷Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

⁸These authors contributed equally: Yun-Tao Liu, Heng Zhang.

*Correspondence should be addressed to:

Z.H.Z: Email: Hong.Zhou@UCLA.edu; phone: 1-310-694-7527

Or G.-Q. B: gqbi@ustc.edu.cn; +86-551-63602891

4.1 Abstract

Cryogenic electron tomography (cryoET) allows visualization of cellular structures *in situ*. However, anisotropic resolution arising from the intrinsic “missing-wedge” problem has presented major challenges in visualization and interpretation of tomograms. Here, we have developed IsoNet, a deep learning-based software package that iteratively reconstructs the missing-wedge information and increases signal-to-noise ratio, using the knowledge learned from raw tomograms. Without the need for sub-tomogram averaging, IsoNet generates tomograms with significantly reduced resolution anisotropy. Applications of IsoNet to three representative types of cryoET data demonstrate greatly improved structural interpretability: resolving lattice defects in immature HIV particles, establishing architecture of the paraflagellar rod in Eukaryotic flagella, and identifying heptagon-containing clathrin cages inside a neuronal synapse of cultured cells. Therefore, by overcoming two fundamental limitations of cryoET, IsoNet enables functional interpretation of cellular tomograms without sub-tomogram averaging. Its application to high-resolution cellular tomograms should also help identify differently oriented complexes of the same kind for near-atomic resolution sub-tomogram averaging.

4.2 Introduction

The advent of single-particle cryoEM has made it routine to determine structures of isolated macromolecular complexes at 2-4 Å resolution by averaging hundreds of thousands of particles, enabling atomic modeling. The biological functions of these complexes, however, are carried out through their interactions and often depend on their spatial arrangements within cells or sub-cellular organelles^{1,2}. Examples abound, ranging from pleomorphic viruses, to cellular organelles, to large-scale cellular structures like synapses between neurons. Many viruses, notably those involved in devastating pandemics such as SARS-CoV-2, influenza viruses, and human immunodeficiency viruses (HIV), are pleomorphic in the organizations of their proteins and genomes. Cellular organelles, such as axonemes containing microtubule doublets surrounding a central pair³, though largely conserved in their core elements across different species, rely on their non-conserved and variable attachment of peripheral components that define their characteristic species-specific functions⁴. In neurons, organizations of molecules, rather than molecules alone, inside the synapse might underlie synaptic plasticity that is generally regarded as the cellular basis of learning and memory^{5,6}. Such organizational information, or “molecular sociology”, unfortunately is lost in single-particle cryoEM.

To reveal such molecular sociology across viruses or inside cells, cryogenic electron tomography (cryoET) has become the tool of choice. This technique requires collecting a series images of the sample at different tilt angles, called “tilt series”. Due to radiation damage, limited electron dosage must be further fractionated throughout the tilt series, resulting in low signal-to-noise ratio (SNR) for the cryo tomogram. Furthermore, as tilting increases the effective thickness of the sample, the tilt range for cryoET is usually restricted to about $\pm 70^\circ$. The missing views at higher tilt angles result in anisotropic resolution of the reconstructed 3D tomograms, with the resolution along the Z-axis the lowest (Supplementary Fig. 1). In Fourier space, these missing views lead to devoid of information in two continuous, opposing wedge-shaped regions,

commonly referred to as the “missing-wedge”, along the tilt axis. This missing-wedge causes severe artifacts in 3D reconstruction of cellular cryoET, manifesting as, e.g., oval-shaped synaptic vesicles⁷ (Supplementary Fig. 1). Thus, together with the low SNR in the reconstructed tomograms, the presence of missing-wedge artifacts prohibits direct interpretation of the reconstructed densities in 3D, which is key to the promise of cryoET to resolve molecule organization *in situ*.

Previous attempts have been made to partially recover information in the missing-wedge⁸⁻¹⁰ with *a priori* assumptions (e.g., density positivity and solvent flatness) to constraint the structural features in reconstructed tomograms. However, such assumptions have limited information content (or “entropy”) and may not always hold true, given the complexity of biological systems. Alternatively, dual-axis tomography relies on imaging the same sample with two perpendicular tilt axes, reducing the two missing-wedges to two missing pyramids; thus it has the potential to alleviate artifacts in resulting tomograms¹¹. However, acquisition and alignment of dual-axis tilt series are substantially more complicated than that of single-axis tilt series and could waste the already limited electron dose used for tilt series acquisition¹². Consequently, dual-axis tomography, while implemented in high-end instruments such as the Thermo-Fisher Titan Krios, has not been practically attractive. Indeed, to date, no structure with better than nanometer resolution was obtained from dual-axis tomography.

Deep neural networks are known to learn relationships of complex data that are non-linear or have high dimensionality. In the field of computer vision, convolutional neural network (CNN) has been applied to various tasks, such as object recognition, image segmentation, and classification, often achieving high performance. In cryoEM field, CNN-based neural networks are applied to particle picking tasks and outperform conventional methods such as the Laplacian of Gaussian approach¹³. CNN is also introduced to cryoEM reconstruction to analyze heterogeneity of protein complexes with remarkable performance¹⁴. However, whether CNN can also recover missing-wedge information in cryoET has not been explored.

Here, we have developed a CNN-based software system, called *IsoNet*, for isotropic reconstruction of electron tomograms. IsoNet trains deep CNN that iteratively restores meaningful contents to compensate missing-wedge, using the information learned from the original tomogram. The resolution at Z-axis reaches about 30Å resolution as measured by the gold-standard Fourier shell correlation (FSC) criterion. By applying IsoNet to processing tomograms representing viral, organelle, and cellular samples, we demonstrate its superior performance in resolving novel structures of lattice defects in immature human immune-deficiency virus capsid, the scissors-stack-network architecture of the paraflagellar rod, and heptagon containing clathrin cage inside a neuronal synapse. The resulting tomograms with isotropic resolution from IsoNet should help direct interpretation and segmentation of 3D structure in cells and 3D picking hundreds of thousands of sub-tomogram particles for future near-atomic resolution cryoET studies.

4.3 Results

4.3.1 Workflow of *IsoNet*

In spite of anisotropic resolution, tomograms generated by cryoET reconstruction contain rich information with structural features such as plasma membranes, organelles, and protein complexes. Thus, it is possible to recover the missing information by merging information from similar features present in the same tomograms but at different orientations relative to each other. An example of filling such missing information is through subtomogram averaging, which aligns and averages structures of particles that are identified to be identical but at different orientations in the tomogram. IsoNet is designed to expand this technique to reconstruct missing-wedge information by training the neural network targeting the subtomograms at different rotations for both regular and polymorphous structures.

The workflow of IsoNet contains five steps (Fig. 1a). Among them, three are major and required: *Extract*, *Refine* and *Predict*; and the other two are optional: *Deconvolve CTF* and

Generate Mask. Each of these steps can be performed with one command of IsoNet in Linux terminal. Among the 5 steps, *Refine* and *Predict* relies on graphical processing unit (GPU) that provides superior processing power. The input of IsoNet is either from a single or multiple tomograms. Based on the principle of machine learning, more tomograms will generate more reliable results but takes longer processing time. In practice, the typical number of tomograms for IsoNet is from one to five. The tomogram(s) can be reconstructed by either weighted back projection (WBP) or iterative methods, such as simultaneous iterative reconstructive technique (SIRT).

We implemented IsoNet in Python using Linux as the native operating system. The package takes advantage of the Keras interface of well-established Tensorflow platform¹⁵ and can be run standard-alone, independent of other software packages. The package can be run either through command line or through a graphical user interface (GUI) (Fig. 1b), thus meeting to the needs of both seasoned and novice cryoET investigators. The GUI contains three tabs to facilitate navigation. In each tab, information of the tomograms and the parameters for each command can be specified. By clicking “*Deconvolve*”, “*Generate Mask*”, “*Extract*”, “*Refine*” and “*Predict*” buttons, user can execute the corresponding command. The “only print command” option prints out the corresponding command for each step which can be executed on other computers or submitted to computer clusters.

Deconvolve CTF and Generate Mask steps. These two optional steps are performed on the input tomograms prior to the subtomogram extraction in *Extract* step (Fig. 1a). The *Deconvolve CTF* step has two purposes: to enhance low-resolution information and compensate for the contrast transfer function (CTF) in the tomograms acquired at certain underfocus conditions. Due to the presence of zeros in CTF, we used a Weiner filter for CTF compensation, as implemented in Warp¹⁶. The *Generate Mask* step uses statistical methods to detect “empty” areas in the tomograms (including vacuum above and below the sample and those only containing ice or carbon) to be excluded from the subsequent analysis. Both steps

could improve performance and efficiency of neural network training.

Extract step. This step allows randomly cropping subtomograms in the original tomograms or the region-of-interest of the tomograms defined by masks. The maximum sizes of subtomograms depend on the memory of graphics processing units (GPU), and 64^3 or 96^3 voxels are often used. The extracted subtomograms can be split into random halves to train the neural network independently (Fig. 1a), allowing users to perform 3D gold-standard FSC^{17,18} to determine the resolution of IsoNet reconstructed tomograms over different angular directions, particularly on Z-axis.

Refine Step. Central to the IsoNet workflow is the *Refine* step, which iteratively trains neural networks to perform missing-wedge correction and denoising (Fig. 1c). Training of the neural network requires paired subtomograms as the “inputs” and the “targets.” The “targets” for IsoNet are the extracted subtomograms rotated at different orientations. In total, 20 different orientations are defined in IsoNet, generating 20 “target” subtomograms for each extracted subtomogram (Supplementary Fig. 2). For each “target” subtomogram, the missing-wedge is computationally imposed in Fourier space to generate the corresponding “input” subtomograms (Fig. 1c). After generating the paired dataset, we train a neural network to map the “input” to the “target”, enabling the network to recover the imposed missing-wedge artifacts. The neural network used in IsoNet adopts U-Net architecture¹⁹, containing an encoder path that extracts low-dimensional representation retaining essential properties, a decoder path to reconstruct from the encoded representation, and skip-connections between encoder and decoder to preserve high-resolution information (Supplementary Fig. 3).

However, the “target” in the data pairs described above are not ideal subtomograms. These subtomograms, though rotated, still miss information in other directions. To recover that information and make “target” subtomograms resembling “ground truth”, we adopt an iterative approach: In the first iteration, we train the network with subtomograms generated from the *Extract* step and obtain the IsoNet-predicted subtomograms. Then, the gained information in the

missing-wedge region in the Fourier space of the predicted subtomograms was added to the original subtomograms, generating the first-iteration missing-wedge corrected subtomograms (Fig. 1c). To further improve miss-wedge correction with more iterations, the corrected subtomograms from the previous iteration are used for the paired data generation in the next iteration because they are closer to missing-wedge-free 3D volumes than the extracted original subtomograms. The trained network from the previous iteration is then refined with the newly generated data pairs. Through multiple iterations, the missing-wedge information is gradually added to the subtomograms (Fig. 1c and Supplementary Fig. 4). Usually, after 10-20 iterations, the refinement converges when the mean square error no longer decreases. The result of this *Refine* step is a trained network that will be applied to the full tomograms and produce the isotropic reconstruction in the *Predict* step (Fig. 1a).

Within the *Refine* step of IsoNet, we also implemented a denoising module based on the noisier-input strategy^{20,21}. When this optional denoising module is enabled in the *Refine* step, in each iteration, 3D noise volumes are reconstructed by the back-projection algorithm from a series of 2D images containing only Gaussian noise. Those 3D noise volumes are then added to “input” subtomograms, with the “target” subtomograms staying the same. With this strategy, IsoNet can be robustly trained with these noisier “input” subtomograms to eliminate the added noise and improve the SNR of final isotropic reconstructions (Fig. 1c and Supplementary Fig. 4).

Predict step. This step performs missing-wedge correction by applying networks obtained in *Refine* step to the tomograms of interest. This *Predict* step runs much faster than the *Refine* step. The tomograms used for *Predict* step are typically (preferably because there are no concerns of bias) the same or a subset of the tomograms used to train the network. Nonetheless, users can in theory apply the trained network to tomograms of other similar samples.

4.3.2 Benchmarking with simulated data

We first perform IsoNet reconstruction on simulated subtomograms using the public available atomic models. Two scenarios have been considered: apoferritin²² for the first test because it has been widely used as a benchmarking specimen in high-resolution cryoEM and ribosome²³ as the second test due to its asymmetric shape and primarily nucleic acid content. For both tests, density maps are simulated from the atomic models using *molmap* function in Chimera²⁴ and filtered to 8Å resolution (Figs. 1d and e). The simulated maps were then rotated in 10 random directions and imposed missing-wedge in Fourier space, resulting in simulated subtomograms with missing-wedge artifacts (leftmost columns in both Figs. 1d and e).

As evident in both tests with simulated subtomograms, features such as alpha-helices perpendicular to the Z-direction are smeared out in those simulated subtomograms due to the missing-wedge artifact. IsoNet was then used to process those simulated subtomograms. As expected, the missing information was recovered during this iterative refinement process (Figs. 1d and e). After 7 iterations, all the alpha helices are visible and identical to the ground truth structures in the first test. The cubic symmetry of apoferritin gradually emerged even though we did not impose symmetry during the processing using IsoNet. In the second test, the distortion in the shapes of ribosome is reduced during the *Refine* step, with the major and minor grooves of the RNA become distinguishable (Fig. 1e). These results indicate that IsoNet performed well with simulated round/symmetric protein complex as well as asymmetric complex containing both protein and nucleic acid.

4.3.3 Application to virus tomograms

To further demonstrate the superior performance of IsoNet in real-world examples, we perform the IsoNet reconstruction with the well-characterized cryoET datasets of virus-like particles (VLP) of immature HIV-1, which is publicly available from the Electron Microscopy Pilot Image Archive^{25,26} (EMPIAR-10164).

After reconstructed with IsoNet, gold beads in the tomogram appear spherical (Fig. 2a),

as they should, instead of the “X” shape due to the missing-wedge problem. Notably, the top and the bottom of the VLP can now be observed in the IsoNet generated tomogram. When examined in the Fourier space, the missing-wedge region on the XZ slices was filled with values compared to the Fourier transform of the original tomogram devoid of the information (Fig. 2a). To quantify the resolution of the filled information, we split the extract subtomograms into two random subsets, trained two neural networks using those two subsets independently, and then performed 3D FSC calculation¹⁷. The resolution on the XY plane is higher than other planes (Fig. 2b), with the resolution along the X and Y axis reaching the Nyquist resolution, showing our network preserves the high-resolution information of the original tomograms. The Z-axis resolution of the isotropic resolution is about 30Å (Fig. 2b), which is the lowest resolution in all directions. This result demonstrates that our isotropic reconstruction can faithfully reconstruct the missing-wedge information at least 30Å resolution.

Importantly, our isotropic 3D reconstruction shows that the quality of the structure is similar across all directions, allowing biological structures to be interpreted adequately (Fig. 2c and Supplementary Video 1). We resolved those broken viruses, sheared along top and bottom planes of the tomograms (Fig. 2c and Supplementary Video 2), indicating that the air-water interfaces caused deformation of the capsid, as well-recognized in the cryoEM field²⁷. The denatured Gag proteins, which are subunits of capsids, at the air-water interfaces are mostly featureless.

The spherical viruses that were fully embedded in ice are made of hexagonal lattices (Fig. 2c), whereas no pentagon subunit is observed, consistent with the subtomogram averaging results of immature HIV particles²⁶. Lattice defects are incorporated onto the hexagonal lattices, making gaps between patches of the lattices (Fig. 2c). These defects and slight curvature on the hexagonal lattices could enable the formation of the spherical shape without pentagons. On lattice edges, small density protrusions extending from the hexagons were observed (Fig. 2d), indicating the complete hexagons are not assembly units of HIV. In

concert with this observation, a recent study shows the Gag dimers are the basic assembly units of the HIV-1 particle²⁸. These protrusions could be Gag dimers and are prone to structural changes during proteolytic cleavage²⁸. Those 3D details on HIV lattices can only be directly visualized after processed by IsoNet. Taken together, the above observations demonstrate that IsoNet can effectively compensate for the missing-wedge problem for relatively thin but heterogeneous structures, such as the immature HIV particles, and reach about 30Å Z-axis resolution.

4.3.4 Application to tomograms of cellular organelles

Next, we tested the performance of IsoNet on resolving structures within cellular organelles by processing tomograms of flagella of *Trypanosoma. Brucei* using IsoNet. The missing-wedge compensated tomogram shows relatively uniform or isotropic structures, in all three dimensions (Figs. 3a and b). The overall contrast is better than the original tomogram partially due to the denoising of the network. One noticeable missing-wedge artifact is that it is difficult to recognize the well-established 9 (outer doublets) + 2 (central-pair singlets) microtubule arrangement in the cross-section view (i.e., XZ view in Fig. 3a). This arrangement can be readily visible in the result generated by IsoNet (Fig. 3b). The missing-wedge effect is also reflected by the broken and oval-shaped microtubules and severe artifacts in XZ and YZ planes in the original tomogram reconstructed with SIRT algorithm (Fig. 3a). In tomograms generated by IsoNet, the microtubules become complete and circular-shaped with some visible tubulin subunits (Fig. 3b and Supplementary Fig. 5). Binding to the microtubules, the arrays of outer (red arrows in Fig. 3b) and inner (blue arrows in Fig. 3b) arm dynein proteins are now clearly distinguishable in the IsoNet generated tomogram. And radial spokes connecting the outer doublets to the central pair can be distinguished in all three orthogonal slices (Fig. 3a-b).

On one side of 9+2 microtubules lies paraflagellar rod (PFR). The structure of PFR is obscure in the tomogram reconstructed by SIRT (Fig. 3c), which has given rise to the long-

lasting debate of the PFR organization²⁹⁻³¹. The IsoNet generated tomograms showed a much clearer picture of PFR. PFR density consists of parallelly arranged planes, and the angle between those planes and the direction of the axoneme is 45°. Within these planes, scissors-like densities stack upon each other, forming a scissors-stack-network (Figs. 3d). This highly organized mesh structure could serve as a biological spring to assist the movement of the flagella. This unique PFR structure observed here is consistent with the organization resolved through tedious efforts of sub-tomogram averaging of thousands of sub-tomograms³². The above observations demonstrate that IsoNet can compensate for the missing-wedge problem for nonspherical cellular organelles, such as those in the Eukaryotic flagella, and unveil structure with meticulous details without the need of sub-tomogram averaging.

4.3.5 Applications to tomograms of cells

To evaluate IsoNet's performance for much larger and more complex structures in cells, we applied IsoNet to tomograms of synapses in cultured hippocampal neurons⁷. Hippocampal synapses are key devices in brain circuits for information processing and storage. They are about 200-1000 nm in size, rich in proteins, lipid membranes, vesicles, mitochondria, and other organelles^{6,7,33}. These samples are thicker⁷ (300-500 nm) than the above-described flagella and virus samples, thus are representative of low SNR tomograms. The intrinsic molecular crowding and structural complexity of the synapse also present difficulties for missing-wedge correction. Arguably, synaptic cryoET tomograms are among the most challenging datasets for any analysis algorithm. Indeed, the missing wedge of cellular tomograms can not be fully recovered using previous softwares (Supplementary Fig. 6).

IsoNet achieved isotropic reconstruction of the synaptic tomogram with substantially higher contrast and better structural integrity (Figs. 4a, b, and Supplementary Videos 3-5). Synaptic vesicles that were smeared out along the Z-axis in the original tomograms now become spherical (Fig. 4c). The linker proteins between vesicles that are hardly seen in the

original tomograms now become visible in XZ and YZ planes (Figs. 4c). Even some horizontally oriented features can be resolved. For example, large patches of membrane on the top and the bottom of the synapse and the endoplasmic reticulum (ER) now appear smooth and continuous in the isotropic reconstruction (Figs. 4b and e). These structural integrity improvements facilitate the segmentation of the cellular structure, since the missing-wedge corrected structures can be directly displayed based on their density threshold in 3D. Particularly, placing the artificial spheres to represent synaptic vesicles, as in previous studies^{7,33}, is no longer needed (Fig. 4e). As the elongation effect of microtubules in the Z-axis being corrected, the protofilaments of microtubules have now become visible (Figs. 4d and f). Inside synapses, numerous small black dots can be observed in the cytoplasm but not in vesicular lumens. These dots represent small cytoplasmic proteins (orange arrows in Fig. 4c), indicating our reconstruction preserves delicate structural features.

As a prominent example, tomograms from IsoNet revealed various types of clathrin coats in hippocampal synapses. Clathrin-mediated endocytosis is a well-known presynaptic vesicle recycling mechanism and is a critical step in synaptic transmission^{34,35}. Clathrin proteins are also present in postsynaptic compartment for neurotransmitter receptor endocytosis, a process playing essential roles in synaptic plasticity³⁶. Those clathrin proteins are known to form cages that consist of pentagons and hexagons³⁷. We observed structures similar to clathrin cages of various sizes in the postsynaptic compartment in synaptic cryo tomogram. However, due to the missing-wedge effect, the geometry of these clathrin cages cannot be directly resolved *in situ* in typical cellular tomograms. We applied IsoNet software to one synaptic tomogram that contains many putative clathrin cages in the postsynaptic compartment (Figs. 5a and b, Supplementary Fig. 7). After the isotropic reconstruction, all the pentagons and hexagons, which made up the clathrin cages, are revealed (Figs. 5c and d). This contrasts with the original tomograms, where the polygons are smeared, especially in XZ and YZ planes.

The 25 clathrin cages can be categorized into five types based on their geometry

(Supplementary Video 6). The most abundant type is minicoat, which is the smallest cages the clathrin proteins can form³⁸. Intriguingly, the largest clathrin cage contains two heptagons, in addition to 8 hexagons and 14 pentagons (Fig. 5d, Supplementary Fig. 8), which has not been reported in previous single-particle analysis^{37,38}. This geometry of the cage deviates from the common belief that a closed polyhedral protein cage contains 12 polygons. This heterogeneous in the Platonic cages of the clathrin arises from the specific yet variable forms of clathrin triskelion interactions. Adapting those heptagons in neurons could likely be strategy to scale up the size of the clathrin coats that enables accommodating different sizes of vesicles. Intriguingly, we did not observe vesicles inside these clathrin cages, suggesting that clathrin protein molecules may spontaneously self-assemble into cages even when not involved in the endocytosis. It is important to note that the unexpected heptagon containing clathrin cage would be lost in averaging-based methods because it only has a single instance in the tomogram. Thus, these observations made in neurons demonstrated that IsoNet enables compensating for missing-wedge for structures that are highly heterogeneous, with limited copy numbers, and in the complex cellular environment.

4.4 Discussion

Here we have developed a deep learning-based package, IsoNet, to overcome the limitation of missing-wedge problem and low SNR plaguing all current cryoET methods. IsoNet embodies several measures that prevent the neural network from “inventing” molecule features. First, the neural network was initialized with random numbers, and all the information comes from original tomograms without prior knowledge. Second, we introduced the dropout factor of 0.5 in the neural network so that with 50% of randomly picked neurons remaining, the network can still reproduce the result. Third, to further prevent over-fitting, the extracted subtomograms for training can be divided into random halves, and the resolution estimation is based on the gold-standard 3D FSC.

To demonstrate its robustness, we have applied IsoNet to process three representative types of cryoET data—pleomorphic virus HIV, cellular organelle axoneme with PFR, and neuronal synapse—representing three levels of length and complexity. IsoNet significantly improved structural interpretability in all these cases, allowing us to resolve novel structures of lattice defects in immature HIV capsid, dynein subunits, and scissors-stack-network architecture of the paraflagellar rod in eukaryotic flagella, and heptagon containing clathrin cage inside a neuronal synapse. In the resulting tomograms, the *in situ* protein features appear isotropic and have high quality that sometimes matches that obtained through subtomogram averaging. For amorphous structures in the tomograms, such as membranes, IsoNet allows the network to learn the feature representation from many other similar structures in the tomogram and recover the missing information. Thus, IsoNet expands the utility of cryoET by overcoming its inherent missing-wedge problem, enabling 3D visualization of structures that are either complex as those in cells (Figs. 4 and 5) or are rare as those tomograms of patients tissues³⁹. Notably, IsoNet outperforms significantly program packages ICON⁸ and MBIR⁹ that were designed to deal with missing wedge problems (Supplementary Fig. 6).

Philosophically speaking, no information can emerge from vacuum/nothing. So where does IsoNet recover the missing information from? The questions touch upon the fundamentals of deep learning and can be thought of as relating to non-locality of information in space. That is, by learning from information scattered around in original tomograms with recurring shapes of molecules, IsoNet sophisticatedly eliminates distorted or missing information. The great advantage of the IsoNet approach is that similar features across different dimensions can be automatically discovered and “averaged” without human intervention. Such features could be related in translation and rotation manners in the three Cartesian dimensions, such as crystalline PFR subunits and axonemal microtubules and dyneins (Fig. 3); they could also be related through symmetries, such as those pentagons and hexagons of clathrin cages; ultimately, they could

also be related biologically, such as the facts that proteins are made up of only 20 amino acids and nucleic acid of four bases, and both are geometrically constrained as a linear molecule. IsoNet learns their relationships in the same tomogram or across multiple tomograms and reconstructs these features automatically. In essence, therefore, IsoNet and sub-tomogram averaging compensate the missing-wedge problem through the same principle.

Regardless of the details of information recovery, the substantial improvement in map interpretability afforded by IsoNet now allows visualization of structures for functional interpretation without the need of tedious and time-consuming sub-tomogram averaging, which typically involves *a priori* feature identification and manual particle picking. Visualizing such structures in cellular tomograms by IsoNet would also improve localization and subsequent sub-tomogram averaging of hundreds of thousand copies of like-structures, leading to *in situ* atomic resolution structures of cellular complexes in their native cellular environment.

4.5 Materials and Methods

4.5.1 Software implementation

We implemented IsoNet in Python using Linux as the native operating system. Typical hardware setup includes one node with four Nvidia GeForce 1080Ti GPU cards of 11 gigabytes memory, which is common in a cryoET research laboratory. The package can be run from command line and relies on Keras that acts as interface for Tensorflow¹⁵, and the package can be downloaded from Github (<https://github.com/Heng-Z/IsoNet>). A detailed document is provided, accompanied by the IsoNet software. Tutorial dataset and video can be found in: https://github.com/Heng-Z/IsoNet_tutorial.

This package is standard-alone and does not rely on other software such as IMOD, while some common Python modules must be installed prior to running IsoNet. Such Python modules are easy to install with the “pip” command. For example, the IsoNet uses Python module “mrcfile” for the read and write tomogram or subtomogram, and “numpy” for the image

processing such as rotation and Fourier transform. The U-net neural network is built by stacking multiple layers (Supplementary Fig. 3) provided in “tensorflow.keras.layers”. For example, three “Conv3D” layers are stacked together in each depth of the encoding path of the U-net.

The package can be launched through a single command entry, either “*isonet.py*” or “*isonet.py gui*”, for Linux command line operations or a graphical user interface (GUI) (Fig. 1b), respectively. Users can then access all the processing steps of the IsoNet procedure. The IsoNet procedure contains five steps, including three major steps: Extract, Refine, and Predict and two accessory steps: CTF deconvolve and mask generate. Each of these steps corresponds to one command of IsoNet in Linux terminal and will be described in the following sections.

4.5.2 Dataset preparation

To use IsoNet, users should prepare a folder containing all tomograms. Binning the tomograms to a pixel size of about 10Å is recommended. Typically, a folder containing 1 to 5 tomograms is used as input. These input tomograms can either be reconstructed by SIRT or WBP algorithm. The tilt axis is the Y-axis, and recommended tilt range is -60 to 60°, while other tilt ranges might also work. The tilt series can be collected with any tilt schemes, continuous, bidirectional, or dose-symmetric.

IsoNet uses Self-defining Text Archive and Retrieval (STAR) file format to store information of tomograms and program parameters been used during Isonet processing. Thus, it inter-operates seamlessly with such leading cryoEM software packages as Relion⁴⁰.

Tomogram STAR file, default named as *tomograms.star*, is required to run IsoNet and this file can be prepared with IsoNet GUI, with a text editor, or with the following command:

```
$ isonet.py prepare_star [tomogram_folder]
```

4.5.3 Deconvolve CTF

For the tilt series imaged without Volta phase plate (VPP), the sinusoidal CTF suppresses or

even inverts information at certain frequencies. To enhance the contrast of the tomograms and promote information retrieval, CTF deconvolution, similar to what is implemented in Warp software, is applied to tomograms in this step.

IsoNet uses Wiener-like filter¹⁶ for CTF deconvolution, with spectral signal-to-noise ratio (SSNR) set empirically:

$$SSNR = e^{-f \times 100F} \times 10^S \times H_f$$

Where f denotes the spatial frequency, H a high-pass filter, F the custom fall-off parameter, and S denotes the custom strength parameter. Because the SSNR of the Wiener-like filter is determined empirically, users can tune the SSNR fall-off or deconvolve strength parameters to enhance contrast of the tomograms. This step can be performed by “CTF deconvolve” function in GUI or with the following command:

```
$ isonet.py deconv [tomogram_star] --snrfalloff 1.0 --deconvstrength 1.0
```

4.5.4 Generate mask

Subtomograms for training would be better to contain rich information than empty areas with only ice, air, or carbon. In this optional mask generation step, IsoNet uses statistical methods to detect empty regions from which subtomograms will not be extracted. Two different types of masks can be applied: density mask that excludes areas with low cryoET density and standard deviation mask that excludes areas with low standard deviation.

The density mask will first suppress the noise with a Gaussian filter and then apply a sliding window maximum filter to the contrast-inverted tomogram. The areas with relatively smaller density values in the filtered tomogram will be deemed as empty space and excluded in the mask. The parameter “*density_percentage*” defines the percentage of the area kept in the tomogram by the density mask.

The standard deviation mask is achieved by calculating the standard deviation of voxels in a small cubic volume centered at each evaluating voxel. The voxels having relatively lower

standard deviation will be excluded. The parameter “*std_percentage*” defines the percentage of voxels kept by the standard deviation mask.

IsoNet uses the intersection of these two masks. The parameters for mask generation can be tuned to cover the region of interest but exclude empty areas. In addition to these two types of masks, IsoNet allows excluding top and bottom parts of tomograms which usually are empty areas by the “*z_crop*” parameter.

Usually, the default parameters will provide a good mask to exclude empty areas; users can also tune the parameters using GUI by “Generate mask” function or the following command, for example:

```
$ isonet.py make_mask [tomogram_star] --density_percentage 50
```

4.5.5 Extract subtomograms

In each tomogram, specified number of seeds are randomly generated within the whole tomogram or the region of interest defined by the mask. Then, cubic volumes centered at the generated seeds are boxed out and saved as subtomograms. The extracted subtomograms should be large enough to cover typical features in tomograms, such as a patch of membrane or vesicle. However, due to the GPU memory limitation, this size cannot be arbitrarily large. We usually extract 300 subtomograms of 96^3 voxels in total. After extraction, the contrast of those subtomograms is inverted. Then, tomograms are normalized by percentile to ensure 80% of the voxel values fit into the range between zero and one. The subtomograms can be randomly split into two halves and used for performing missing-wedge correction independently to eliminate overfitting and calculate gold-standard FSC.

The information of extracted subtomograms is stored in another STAR file, default named as *subtomo.star*. Subtomogram extraction can be performed by either IsoNet GUI or the following command:

```
$ isonet.py extract [tomogram_star]
```

4.5.6 Refine

This process iteratively trains neural networks to fill the missing-wedge information using the same tomograms whose missing-wedge artifacts were added to other directions. The denoising module can also be enabled in this step, making the network capable of reducing noise and recovering the missing-wedge. After refinement, the resulting subtomograms and neural network model in each iteration are saved. The network models with a suffix of “.h5” can be used for the prediction step.

Four steps, including *training dataset generation*, *adding noise*, *network training*, and *subtomograms prediction*, will be performed during each iteration. These steps will be described in the following sections. The missing-wedge restored subtomograms by *subtomograms prediction* in every iteration will be used for *training dataset generation* in the next iteration. Usually, 10-15 iterations in the refine step are sufficient to obtain a well-trained network for the missing-wedge correction, whereas more iterations can be performed for refinement with denoising.

The refine step can be performed from the GUI or with the following command, for example:

```
$ isonet.py refine [tomogram_star] --iterations 30 --gpuID '0,1,2,3'
```

Users can also continue training from the previous iterations using “*continue_from*” or from previously trained models using “*pretrained_model*” parameter.

Refine step 1: training dataset generation

To generate paired datasets for neural network training, IsoNet rotates the extracted subtomograms to different orientations. Twenty rotated copies can be obtained for each extracted subtomogram as follows (Supplementary Fig. 2). First, each subtomogram is a cube with six faces. Each face can be rotated with an out-of-plane angle to face toward the positive direction of the Z-axis. Second, each out-of-plane rotation can be followed by four in-plane

rotations, making 24 possible rotations. However, among the 24 rotations, four of them result in subtomograms with the same missing-wedge direction as the original subtomograms. Thus, these four rotations are excluded, resulting in 20 orientations for each subtomograms. This rotation process enlarges the original dataset by 20 times for training, making it possible to achieve a good performance of missing-wedge correction even with a small dataset, *e.g.*, a single tomogram.

After the rotation, the IsoNet program then applies missing-wedge filter to the rotated subtomograms. The missing-wedge filter volume has the same size as that of the subtomograms. In the missing-wedge filter volume, voxel value is zero inside the missing-wedge region and one in the rest of the volume. Then, the Fourier transforms of the rotated subtomograms are multiplied by the missing-wedge filter volume and then transformed back to the real space, generating missing-wedge filtered subtomograms.

To avoid the incomplete information along the edge of the subtomograms when applying missing-wedge filter, both rotated subtomograms and missing-wedge filtered rotated subtomograms are trimmed into smaller volumes (often 64^3 voxels), generating “target” and “input” for the network training, respectively. These generated data pairs are used to train neural network that maps the “input” to “target”.

Refine step 2: adding noise.

This optional step allows performing missing-wedge correction and denoising simultaneously using IsoNet. IsoNet uses a noisier-input strategy^{20,21} that learns to map “input” with additional noise to the “target”.

IsoNet simulates the noise pattern in reconstructed tomograms with the assumption that in every acquired projection, the noise is an additive Gaussian noise and independent among all images acquired in a tilt series. During the adding noise step, a set of 3D noise volumes are constructed by back-projecting a series of 2D Gaussian noise images to reflect the effect of the back-projection algorithm on noise formation.

The denoise level is defined as the ratio of the standard deviation between the subtomograms and the added noise. The noise volumes are scaled to match the denoise level before being added to the “input” subtomograms. Thus, the lower denoise value means less noise is added to individual subtomograms.

Because the added noise may further corrupt the 3D subtomograms, making the network hard to train, it is recommended to start the first several iterations of refinement without denoising. After the refinement results are nearly converged, the noise volume can then be added to the “input” subtomograms in the following iterations. A typical routine is to train ten iterations without denoising and then increase the denoise level by 0.05 for every five iterations. This step-wised noise addition can be performed automatically in the refine step of the IsoNet software.

Refine step 3: network training

Neural network used in IsoNet is based on U-net, which is well recognized in biomedical image restoration and segmentation¹⁹. The main building blocks of the U-net are 3D convolution layers with non-linear activation functions called Rectified Linear Units (Relu), which are applied per voxel. Those convolution layers have kernel sizes of 3x3x3. Three 3D convolution layers are stacked together to form a convolution block in our network, which can extract complicated features.

By stacking the convolutional blocks, the U-net is built based on encoder-decoder architecture (Supplementary Fig. 3). The encoder path is a set of convolution blocks and strided convolution layers that compress 3D volumes. Strided convolution layers reduce the spatial size of the input of this layer by 2x2x2, allowing the network to learn more abstract information. A convolution block followed by a strided convolution layer makes one encoder block in the contracting path. Total three encoder blocks form the entire encoding path. The number of convolution kernels for each convolution layer doubles after each encoder block. After the encoder path, the 3D volumes are processed with a convolution block and enter the decoder

path of the network. The decoder path is symmetrical to the encoder but uses transpose convolution layers, opposite to strided convolution layers, to enlarge the dimension of features.

Although the down-sampling of the 3D volumes captures the essence of the features, high-resolution information is lost by stride convolution operations. In particular, the skip connections that concatenate the feature layers of the same dimension in two paths are implemented to preserve the high-resolution information. Dropout strategy that randomly set 50% of neurons' activation to 0 in the convolution layers is used to prevent overfitting during the training.

This network uses the mean absolute error between the output of the network and the target subtomograms as loss function. The loss function is minimized by employing Adam optimizer⁴¹ with an initial learning rate of 0.0004. The neural network training is performed on GPU and consists of several epochs or cycles (typically ten epochs). Each epoch will traverse through the paired dataset. The data pairs are grouped into batches (which generally have a size of 8 or 16) to feed into each epoch. After the training, the trained neural networks are saved for the next iteration of the refine step.

Refine step 4: Subtomogram prediction

After each iteration of refinement, the network is applied to the original subtomograms, generating predicted subtomograms. Then IsoNet generates an inverse missing-wedge filter volume with value one inside the missing-wedge region but zero in the rest of the 3D volume. The predicted subtomograms are then transformed to Fourier space and multiplied with the inverse missing-wedge filter volume to extract the added information inside the missing-wedge regions. Then, these filtered volumes are added to the original subtomograms, generating the missing-wedge restored subtomograms for subsequent iterations of refinement.

4.5.7 Predict

After the refine step, the trained network is saved in a model file. It will be used to correct the

missing-wedge for the original tomogram or other similar tomograms. For most tomograms, the full-size 3D images can hardly fit into the memory of a regular GPU. Thus, the IsoNet program splits the entire tomogram into smaller 3D chunks to apply the network on them separately. Then output 3D chunks are montaged to produce the final output. To avoid the line artifact between adjacent chunks caused by the loss of information on the edges of subtomograms. We implemented a seamless reconstruction method called overlap-tile strategy, which predicts the overlapping chunks to avoid the edge effect. The “crop_size” parameter defines the size of the cubic chunks. This predicting step can be performed with IsoNet GUI or with the following command, for example:

```
$isonet.py predict [tomogram_star] [output_folder] --model [network_model]
```

4.5.8 Benchmarking with simulated data

We performed IsoNet reconstruction on simulated subtomograms using the public available atomic models: apoferritin model²² (PDB: 6Z6U) and ribosome model²³ (PDB: 5T2C). For both tests, density maps were simulated from the atomic models using “*molmap*” function in ChimeraX²⁴ and filtered to 8Å resolution (Fig. 1d, e). Those simulated maps with 2.67 Å/pix pixel size were then rotated in 10 random directions and imposed with missing-wedge filter in Fourier space, resulting in simulated subtomograms with missing-wedge artifacts (leftmost columns in both Fig. 1d, e).

For the simulated Apoferritin maps, we created a subtomogram STAR file with the “*isonet.py prepare_subtomo_star*” command. With this subtomogram STAR file as input, we performed an IsoNet refine step for ten iterations without denoising. For benchmarking with the simulated ribosome maps, we extracted eight smaller subtomograms (96³ voxels) from each ribosome map due to the larger dimension of the ribosome map. The subtomogram STAR file generated in the extract step was used for subsequent refine step. After ten iterations, trained network was obtained and was then used to produce missing-wedge corrected maps of

ribosome using “isonet.py predict” command.

4.5.9 Processing tomograms of HIV virus

For pleomorphic viruses, we downloaded an HIV dataset from public repository EMPIAR-10164²⁶. Three tilt series, TS_01, TS_43 and TS_45, were used for testing. The movie stacks were drift corrected with MotionCorr⁴² and reconstructed with IMOD¹¹ using WBP algorithm. The defocus value of each image was determined by CTFFIND4⁴³. Eight-time binned tomograms with 10.8Å pixel size were used for further processing. For the CTF deconvolution of the tomograms, SSNR fall-off and the deconvolve strength parameters were set to 0.7 and 1, respectively. Then, we created one mask for each tomogram using “isonet.py make_mask” command. Total 300 subtomograms with 96³ voxels were randomly extracted from the three tomograms and then split into random halves. For each half of the subtomograms, we performed refine step for 35 iterations independently, generating two trained neural networks. In the predict step of IsoNet, tomogram TS_01 was used to generate two missing-wedge corrected tomograms by the two independently trained networks. These two tomograms were then averaged to create a final map.

These two missing-wedge corrected tomograms enabled calculating gold-standard FSC. Instead of calculating a global FSC, we performed 3D FSC calculation for all the directions¹⁷ to measure the resolution anisotropy of the missing-wedge corrected tomogram. Because the 3D FSC calculation works for cubic volumes while the size of the tomogram is non-cubical, we cropped the generated tomograms into cubic subtomograms for the 3D FSC calculation. As for the HIV dataset, the 3D FSC was calculated for four 200³ volumes cropped from both missing-wedge corrected HIV tomograms. The resulting four 3D FSC were then averaged to produce the final 3D FSC, the orthogonal sections of which are shown in Figure 2b.

4.5.10 Processing tomograms of the Eukaryotic flagella

For cellular organelles, we chose the demembrated flagella of *Trypanosoma. Brucei*. The

datasets described here were obtained in our previous studies^{32,44}. Tilt series were recorded with SerialEM⁴⁵ by tilting the specimen stage from -60° to $+60^\circ$ with 2° increments. The cumulative electron dosage was limited to 100 to 110 $e^-/\text{\AA}^2$ per tilt series. The movie stacks were drift corrected with MotionCorr⁴² and reconstructed with IMOD using SIRT algorithm. The tomograms were binned by four, resulting in a pixel size of 10.21 $\text{\AA}/\text{pixel}$.

Three tomograms were chosen for missing-wedge correction. These tilt series were acquired with VPP so that we did not perform the CTF deconvolution. We generated one mask for each tomogram using “isonet.py make_mask” command. Then, we extracted a total of 360 cubic subtomograms with 128^3 voxels from three tomograms. Using these subtomograms, we trained a network by running the refine step for 35 iterations with default denoise levels, which were automatically changed from 0 to 0.2. Trained network produced in the refine step was then used to run the predict step of IsoNet to obtain a final missing-wedge corrected tomogram, which is shown in Figure 3.

4.5.11 Processing tomograms of hippocampal neurons

Tomograms of hippocampal neurons were obtained in our previous study⁷. The two tomograms used in this study were collected on a Titan Krios microscope equipped with K2 summit in counting mode. The energy filter (Gatan image filter) slit was set at 20 eV. The Titan Krios was operated at an acceleration voltage of 300 KV. Tilt series were acquired using SerialEM⁴⁵ with tilt scheme: from $+48^\circ$ to -60° and from $+50^\circ$ to $+66^\circ$ at an interval of 2° . The total accumulated dose was $\sim 150 e^-/\text{\AA}^2$. The pixel size of the tomograms is 4.35 $\text{\AA}/\text{pixel}$. Each recorded movie stack was drift-corrected and averaged to produce a corresponding micrograph using MotionCorr⁴². The tilt series were aligned using IMOD¹¹. One tilt series of the tomogram shown in Figure 4 was imaged with VPP, while the other shown in Figure 5 was acquired without VPP. When VPP was used, the defocus value was maintained at $-1 \mu\text{m}$; otherwise, it was kept at $-4 \mu\text{m}$.

For the tilt series recorded with VPP, the aligned tilt series were reconstructed using NovaCTF⁴⁶, generating a tomogram reconstructed with WBP. The tomogram was binned by four, and 300 subtomograms (96³ voxels) were extracted from that tomogram. Those subtomograms were then used 35 iterations of refinement. The trained network produced in refine step was used for missing-wedge correction for the entire tomogram (Fig. 4).

For the tilt series recorded without VPP, the defocus value of each image was determined by CTFFIND4⁴³, and CTF phase flipped tomogram was obtained by NovaCTF⁴⁶. This tomogram (Fig. 5) was reconstructed with SIRT-like filter, with CTF phase flipping performed on the individual tilt images. The tomogram was binned by four for missing-wedge correction with IsoNet. Then, 200 subtomograms (96³ voxels) were extracted from the tomogram in the extract step of IsoNet. A trained network was obtained with the refine step of IsoNet for 35 iterations. The trained network was then used for the predict step of IsoNet, producing a missing-wedge corrected tomogram (Fig. 5).

4.5.12 3D visualization

IMOD⁴⁷ was used to visualize the 2D tomographic slices. UCSF ChimeraX²⁴ was used to visualize the resulting IsoNet generated tomograms in their three dimensions. Segmentation of densities maps and surface rendering were performed by the volume tracer and color zone in UCSF ChimeraX.

4.6 Acknowledgments

We thank Chong-Li Tian, Yi-Nan Xia and Zhuo Li for testing the program, and Jiayan Zhang, Simon Imhof, Ivo Atanasov, Wong Hui and Kent Hill for the Eukaryotic flagellum data. This work was supported in part by grants from the National Key R&D Program of China (2017YFA0505300), the National Natural Science Foundation of China (31630030, 31761163006, and 31621002), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB32030200). Research in the Zhou group is supported in part by the

U.S. National Institutes of Health (GM071940). We acknowledge the use of resources at the Center for Integrative Imaging of Hefei National Laboratory for Physical Sciences at the Microscale, and those at the Electron Imaging Center for Nanomachines of UCLA supported by U.S. NIH (S10RR23057 and S10OD018111) and U.S. NSF (DMR-1548924 and DBI-133813).

4.7 Author contributions

Y.-T.L. conceptualized the method, wrote code, processed data, made illustrations and documentation, and wrote the paper; H.Z. wrote code, processed the HIV and neuronal synapse data, made documentation and assisted in illustrations, and wrote the paper; H. W. participated in coding, processed the axoneme with PFR data, made documentation and assisted in illustrations; C.-L.T. assisted in testing the method with the neuronal synapse data and in making illustration based on the test results; G.-Q. B. and Z.H.Z. oversaw the project, interpreted the results, and wrote the paper. All the authors edited and approved the manuscript.

4.8 Conflict of interest

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.-Q.B. (gqbi@ustc.edu.cn) or Z.H.Z. (Hong.Zhou@UCLA.edu).

4.9 Data Availability

All data reported in the main text or the supplementary materials are available upon request. The software has been deposited to GitHub page at <https://github.com/Heng-Z/IsoNet>.

4.10 Figures

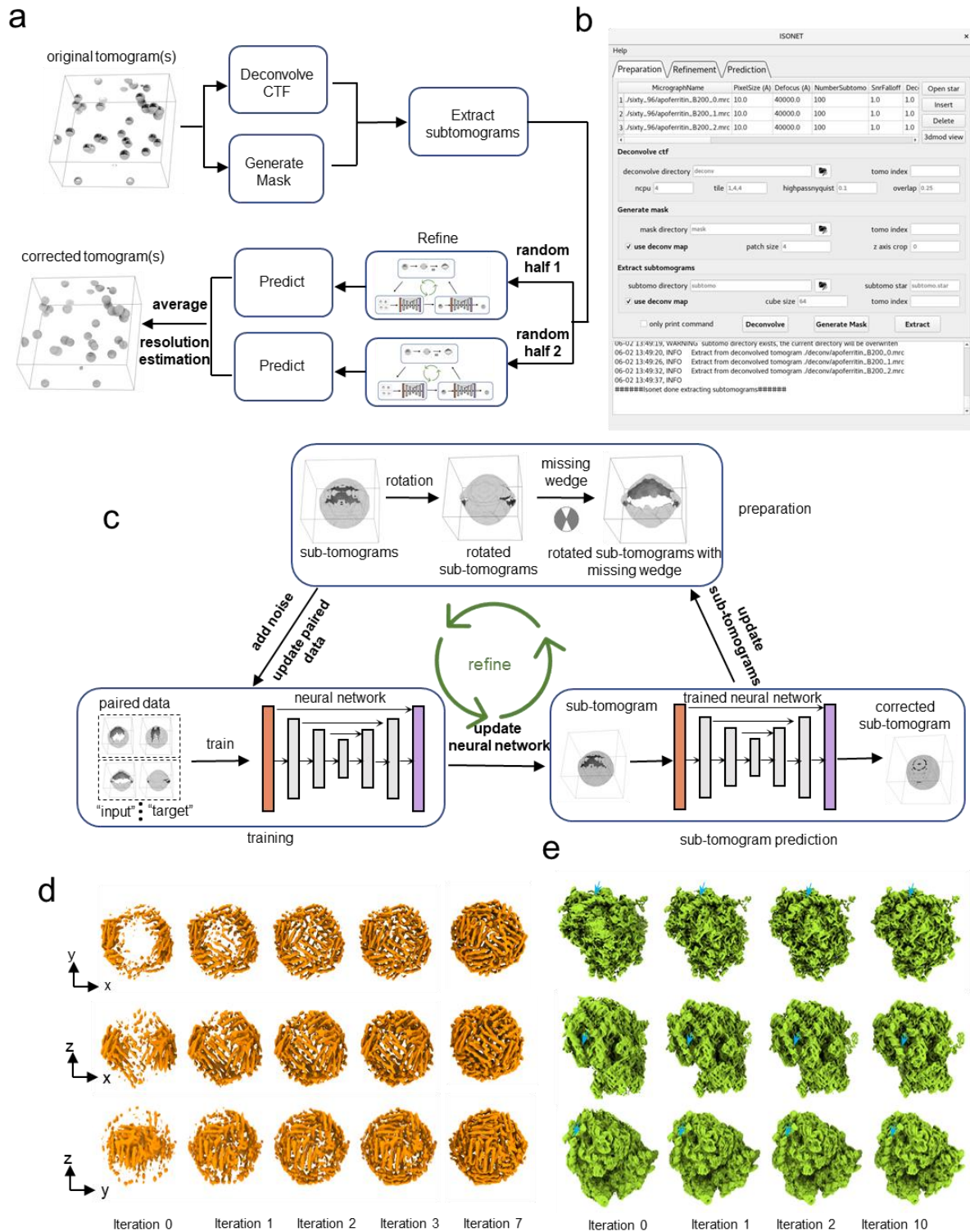


Figure 4- 1. Principle and workflow of IsoNet.

a, Workflow of the IsoNet software. **b**, GUI of IsoNet. **c**, Illustration of *Refine* step: First, subtomograms are rotated and then applied with additional missing-wedge artifacts to produce

paired data for training. Second, the paired data is used to train a neural network with U-net architecture. Third, the trained neural network is applied to the extracted subtomograms to produce missing-wedge corrected subtomograms. The recovered information in these subtomograms is added to the original subtomograms, producing new datasets for the next iteration. **d**, Validation of IsoNet with simulated sub-tomogram of apoferritin and ribosome. Surface views from three orthogonal directions of the reconstructions are shown after increasing iterations of IsoNet processing. Blue arrows indicate segments of RNA duplexes.

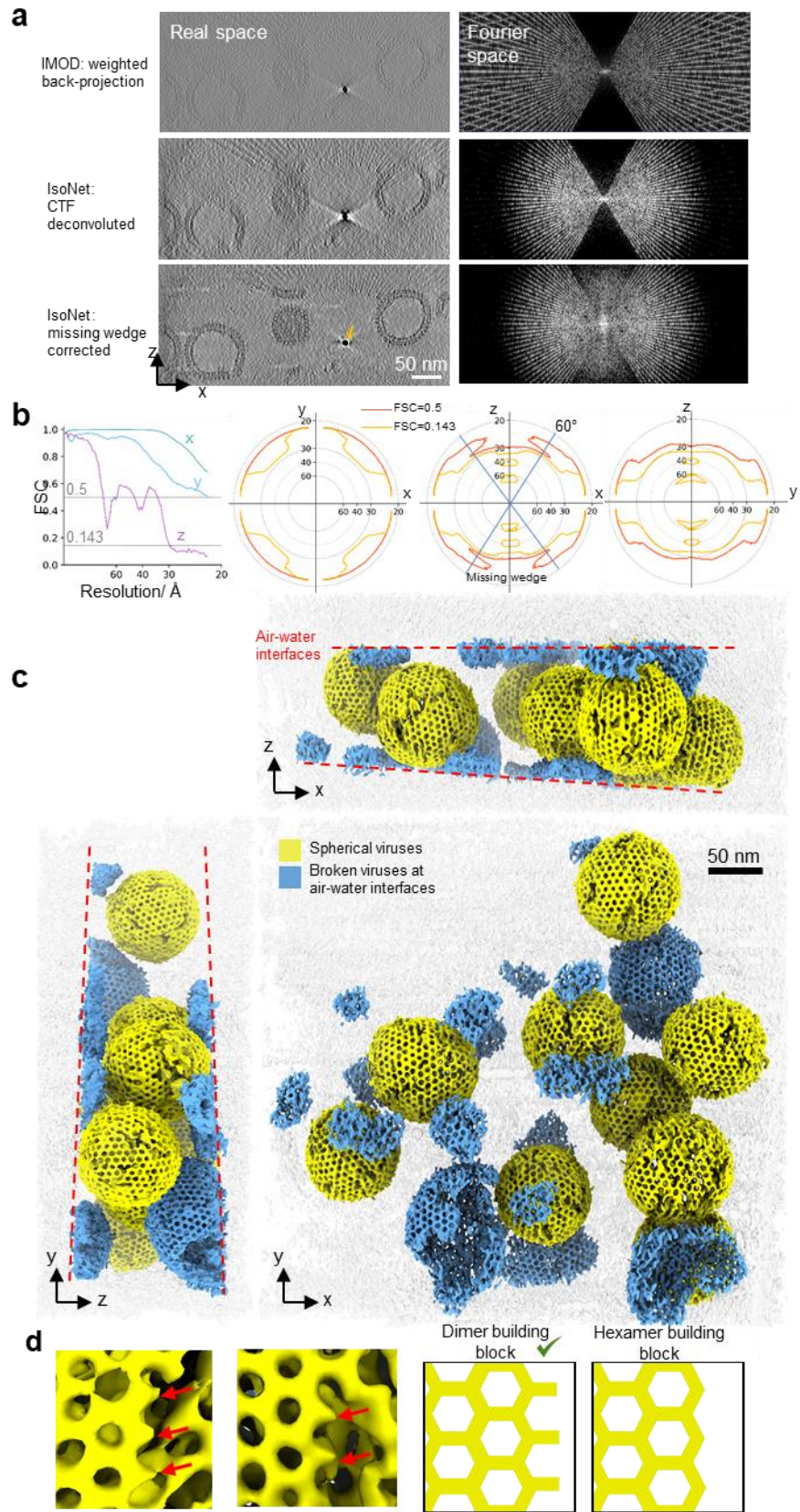


Figure 4- 2. IsoNet reveals lattice defects in immature HIV capsid.

a, XZ slice views of tomogram reconstructed with WBP (top), CTF deconvoluted WBP tomogram (middle), and missing-wedge corrected tomograms (bottom), with their Fourier transforms on the right. Orange arrow indicates a gold bead. **b**, 3D FSC of the two independent isotropic reconstructions, the left panel shows the FSC along the X, Y and Z directions. Three panels on the right show the 3D FSC at 0.5 and 0.143 cutoffs on XY, XZ, and YZ planes, respectively. **c**, 3D rendering of the missing-wedge corrected tomogram. Dashed lines show the air-water interfaces. **d**, Examples (left) and illustrations of the lattice edges of HIV capsids. Red arrows point out the density protrusions on the edges of hexagonal lattices.

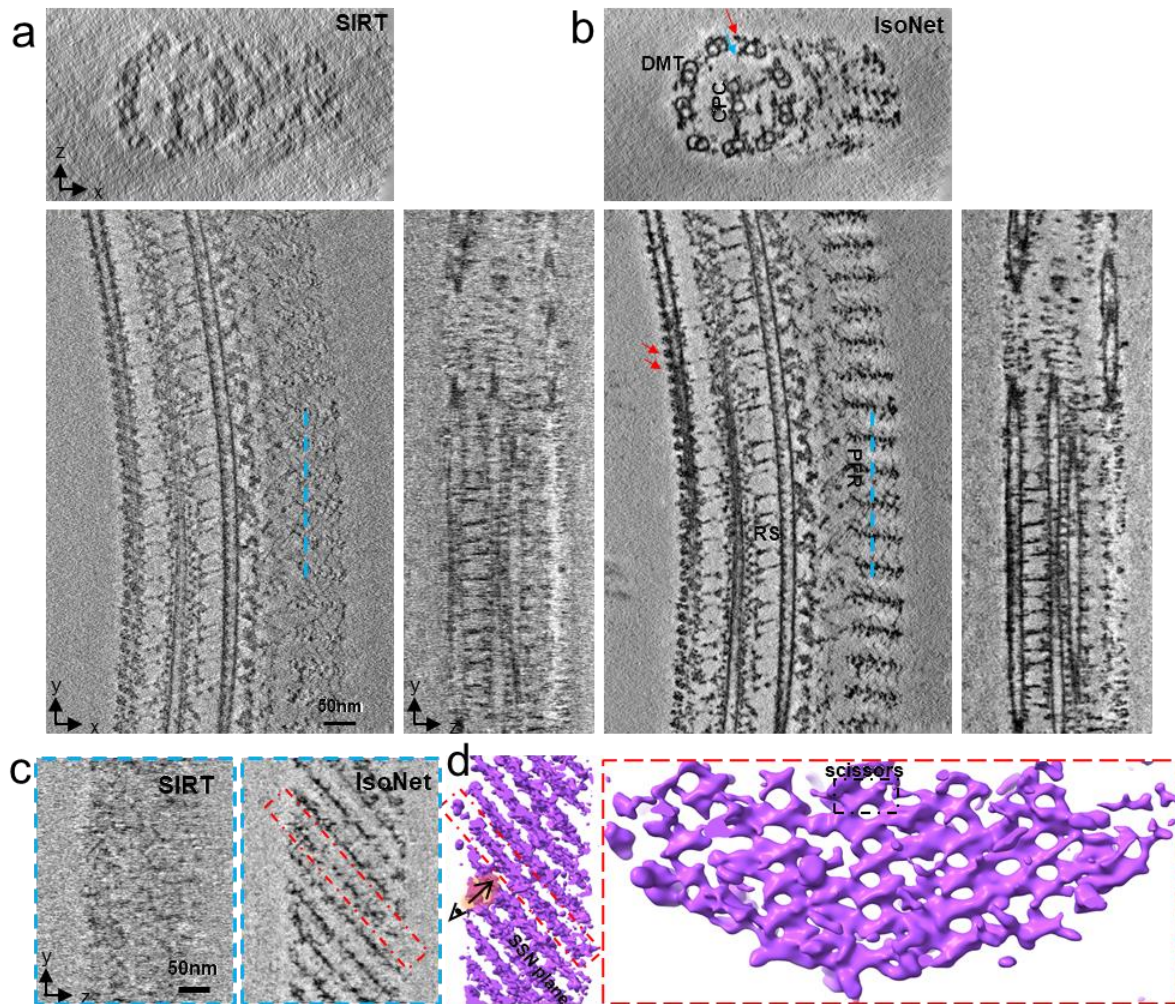


Figure 4- 3. Architecture of the PFR revealed after missing-wedge correction.

a-b, Orthogonal slices of a tomogram of flagellar for SIRT reconstruction **(a)** and IsoNet reconstruction **(b)**. DMT: Doublet microtubule; CPC: central pair complex; RS: radial spoke; Red arrows: outer arm dyneins; Blue arrow: inner arm dynein. **c**, YZ slices show the cross-sections corresponding to the cyan lines in **(a)** and **(b)**. **d**, 3D rendering of PFR in the IsoNet generated tomogram. Left panel is the 3D view of PFR in the direction corresponding to **(c)**. Right panel shows the *en face* view of a PFR scissors-stack-network (SSN) plane.

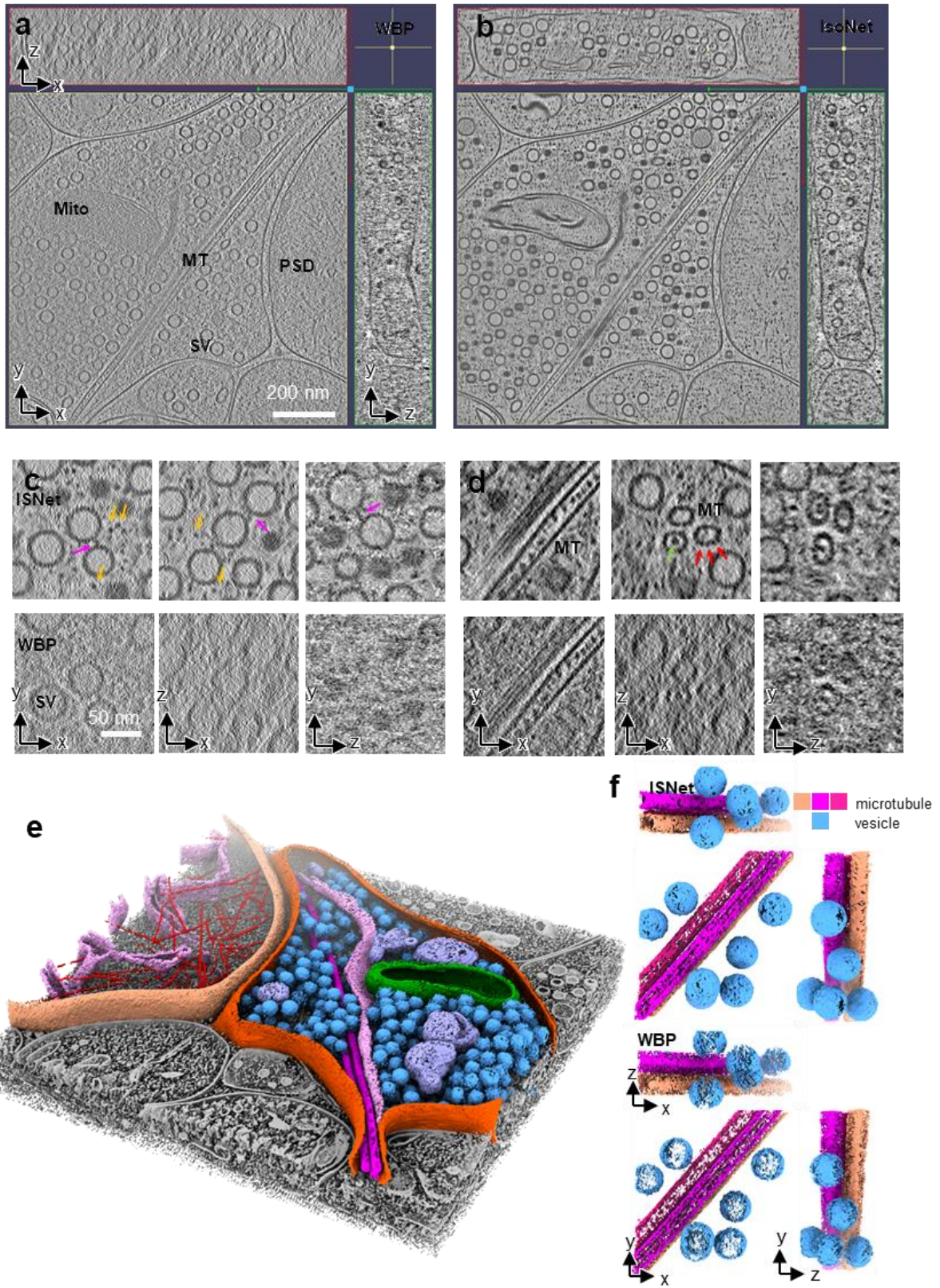


Figure 4- 4. IsoNet recovers missing information in the tomograms of neuronal synapses.

a-b, Orthogonal slices of a synaptic tomogram reconstructed with WBP (**a**) and IsoNet (**b**). SV: synaptic vesicle; Mito: mitochondria; MT: microtubule; PSD: postsynaptic density. **c-d**, Zoomed-in orthogonal slices of WBP reconstruction and IsoNet produced reconstruction. Magenta arrows: vesicle linker; Orange arrows: small cellular proteins; Green arrow: microtubule luminal particles; Red arrows: microtubule subunits. **e**, 3D rendering of the tomogram shown in (**b**). **f**, 3D rendering of a slab of tomogram with WBP reconstructions and Isotropic reconstructions, showing microtubules and vesicles.

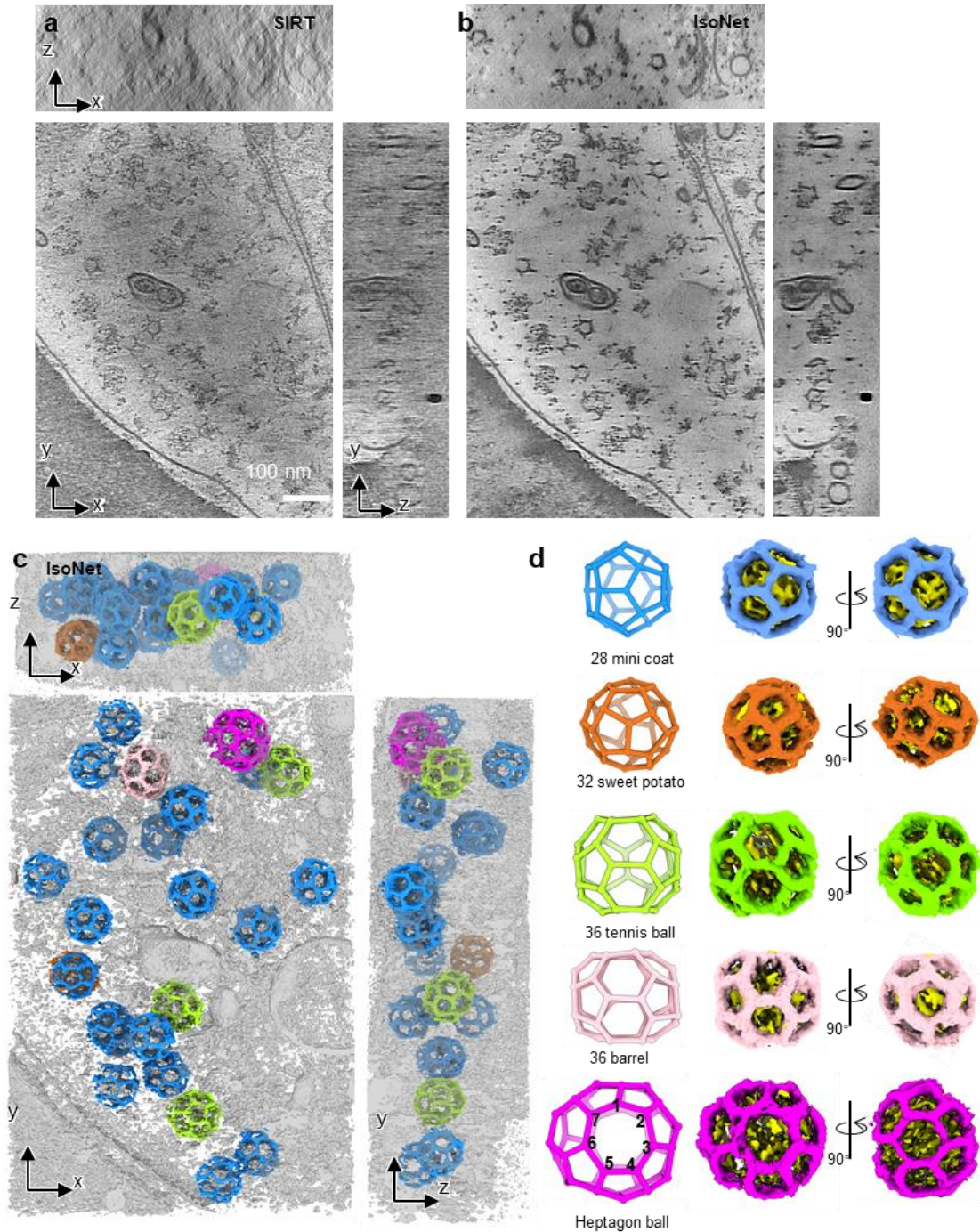
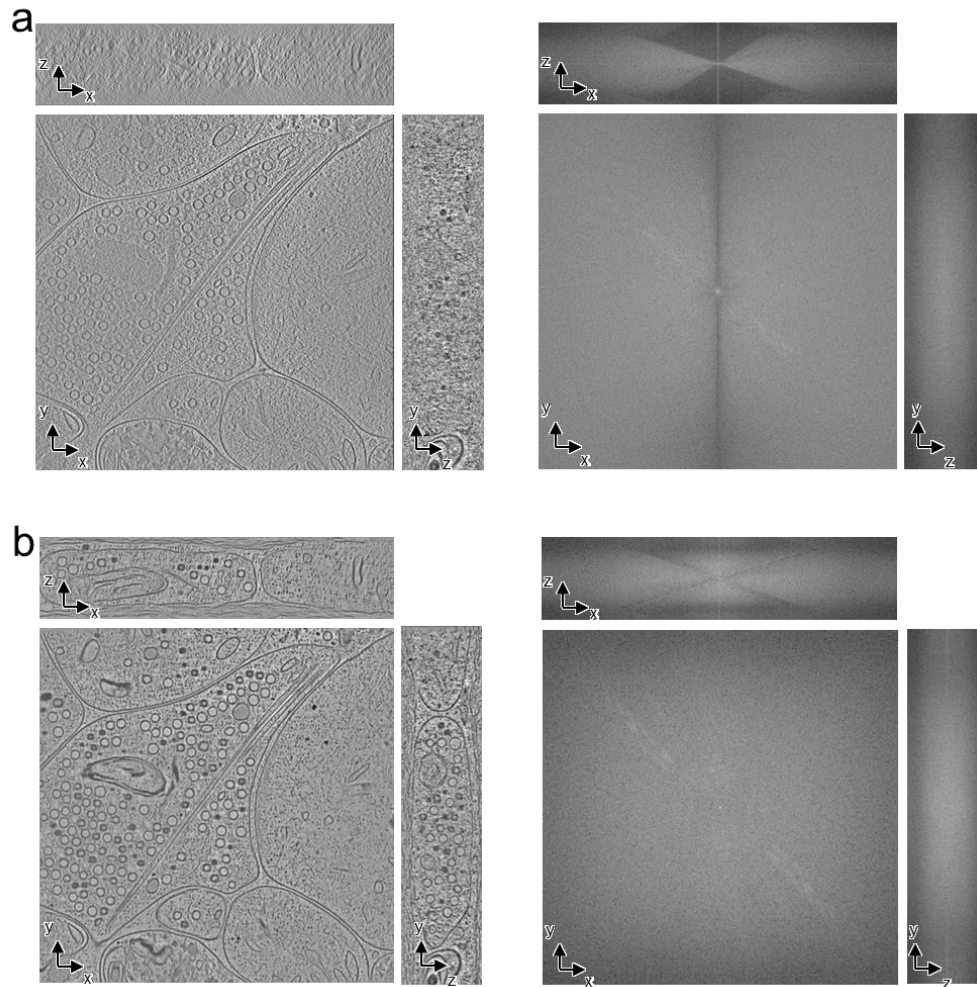


Figure 4- 5. IsoNet reveals various types clathrin coats in a synapse.

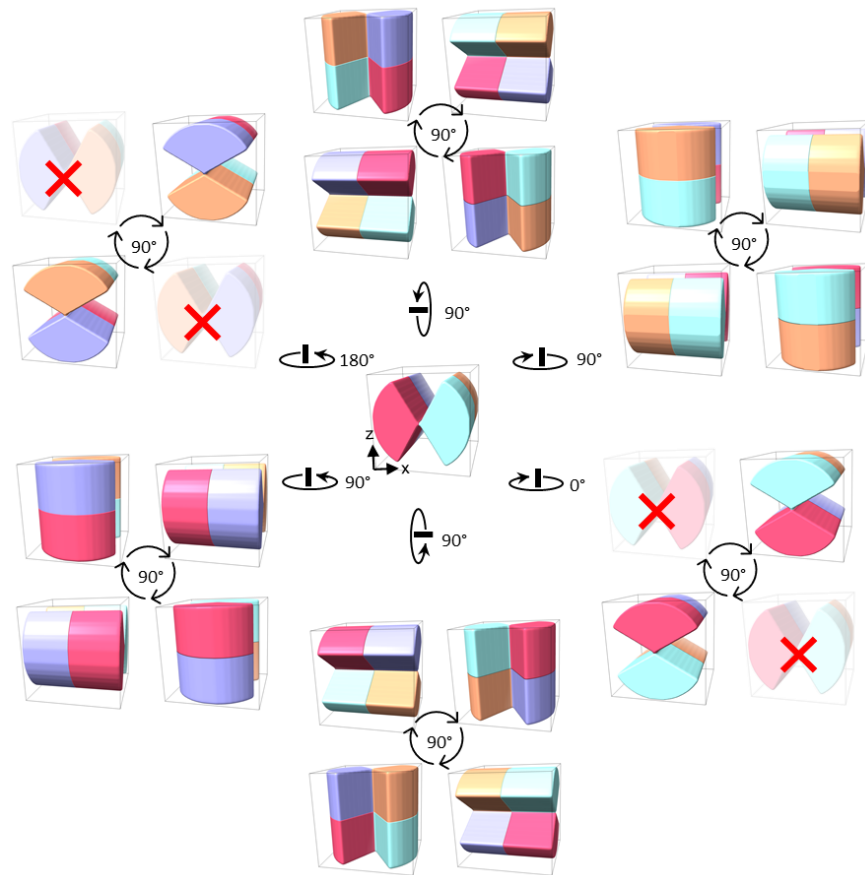
a-b, Orthogonal slice views of another synaptic tomogram reconstructed with SIRT algorithm **(a)** and IsoNet **(b)**. **c,** 3D rendering of the tomogram shown in **(b)**. **d,** 3D views of the five types of clathrin cages in **(c)**.

4.11 Supplemental Figures



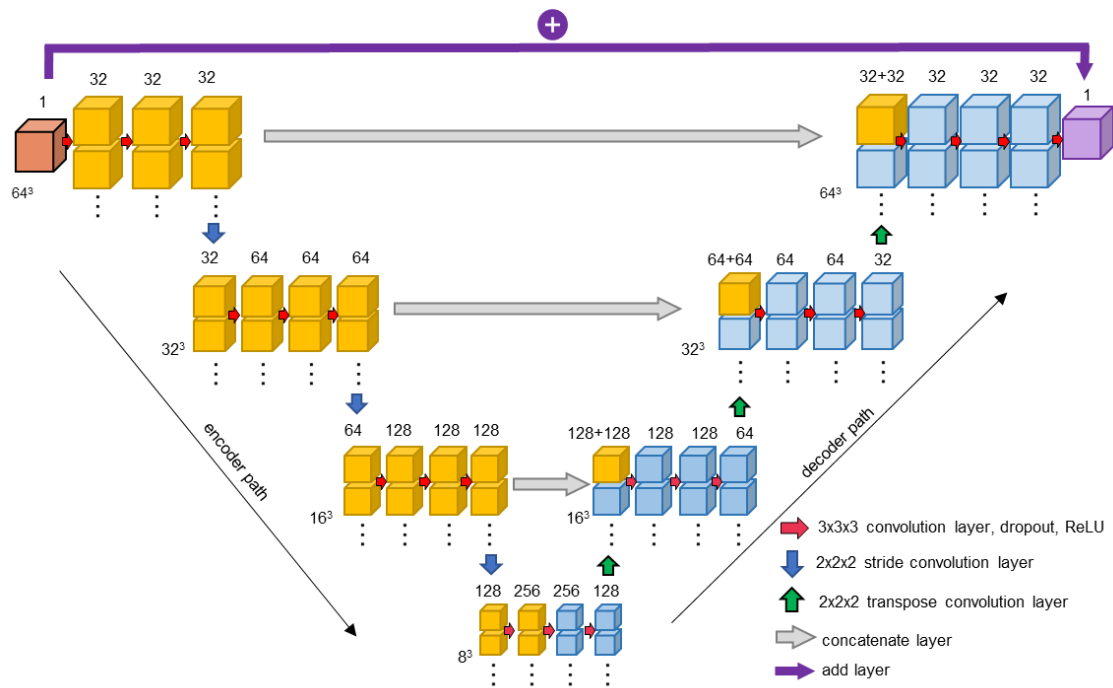
Supplementary Figure 4- 1.

a, Orthogonal views of the WBP reconstructed tomograms and their corresponding Fourier transforms. **b**, Orthogonal views of the IsoNet reconstructed tomograms and their corresponding Fourier transforms.



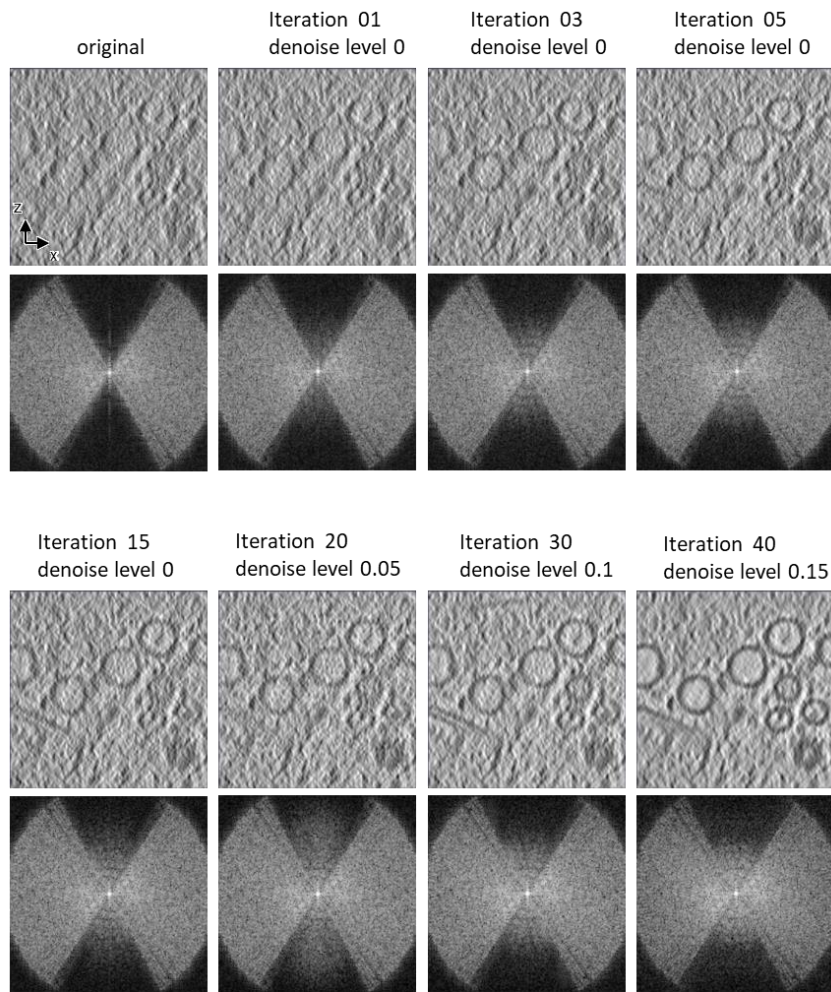
Supplementary Figure 4- 2. Rotation schemes.

Twenty rotated copies are obtained for each extracted subtomograms demonstrated in the center. First, each subtomogram has six faces. Each face can be rotated with an out-of-plane angle to face toward six positive directions of the Y-axis. Second, each out-of-plane rotation is followed by four in-plane rotations, making a total of 24 possible rotations. However, among the 24 rotations copies, four have the same Z-axis missing-wedge as the original subtomograms. Thus, these four rotations are excluded (red cross).



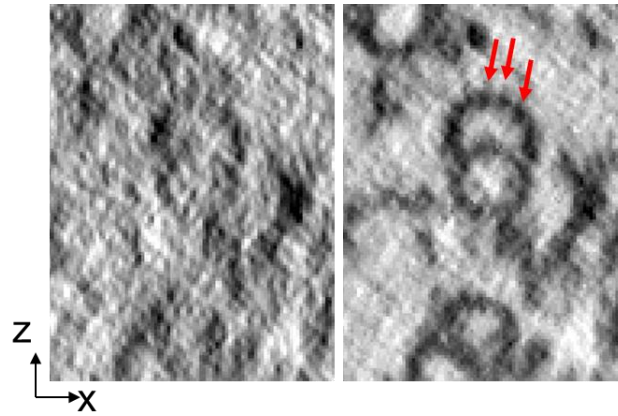
Supplementary Figure 4- 3. The architecture of neural network based on U-net.

The values at the bottom left of boxes show sizes of 3D feature maps or subtomograms, while the values on top of the boxes are their numbers.



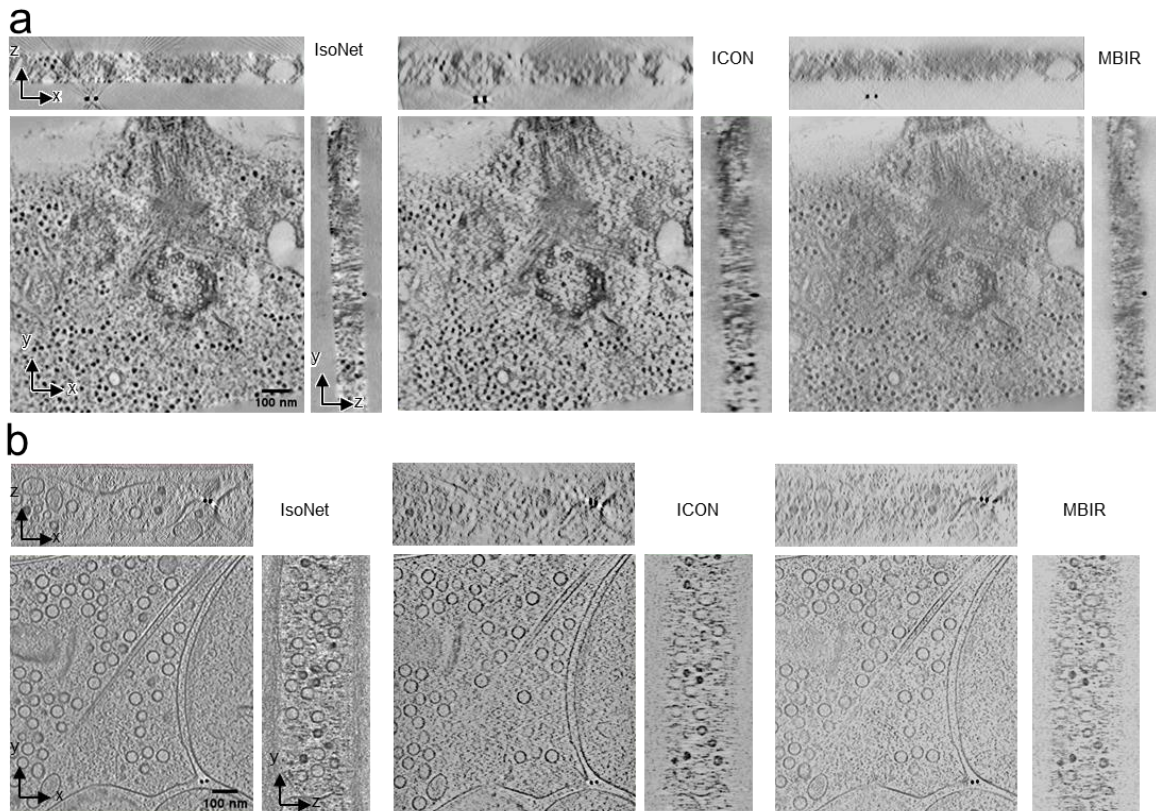
Supplementary Figure 4- 4. Iteratively filling the missing-wedge region.

XZ slice views of subtomograms and corresponding power spectrum at different iterations in refine step.



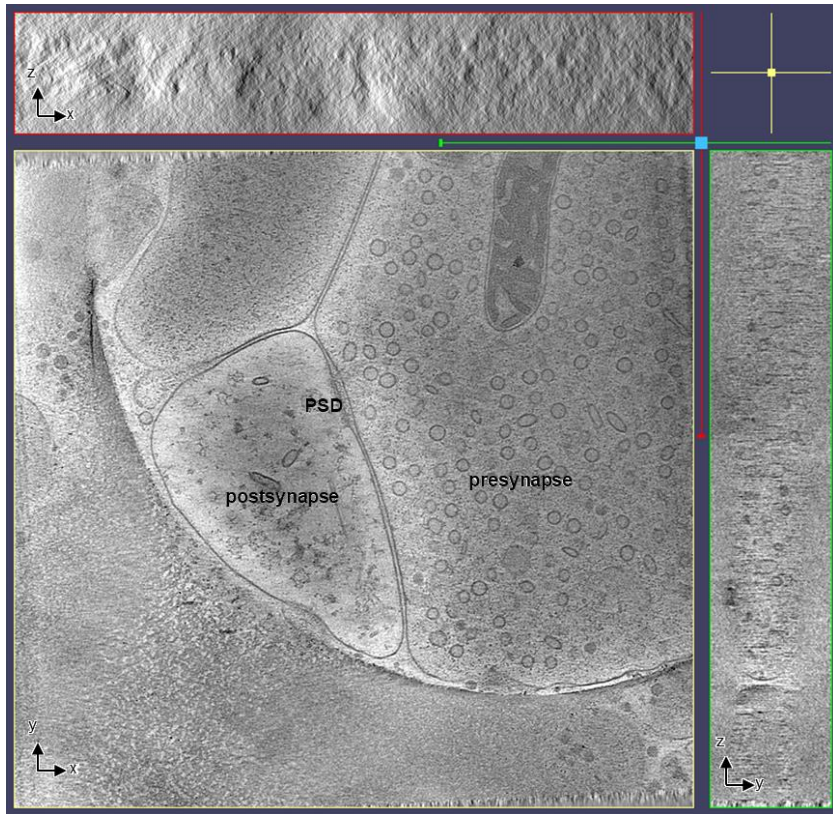
Supplementary Figure 4- 5. XZ slices of microtubule doublets in axoneme.

Left: XZ slices of the SIRT reconstructions. Right: the slices of the corresponding tomogram generated by IsoNet. Red arrows indicate microtubule subunits.



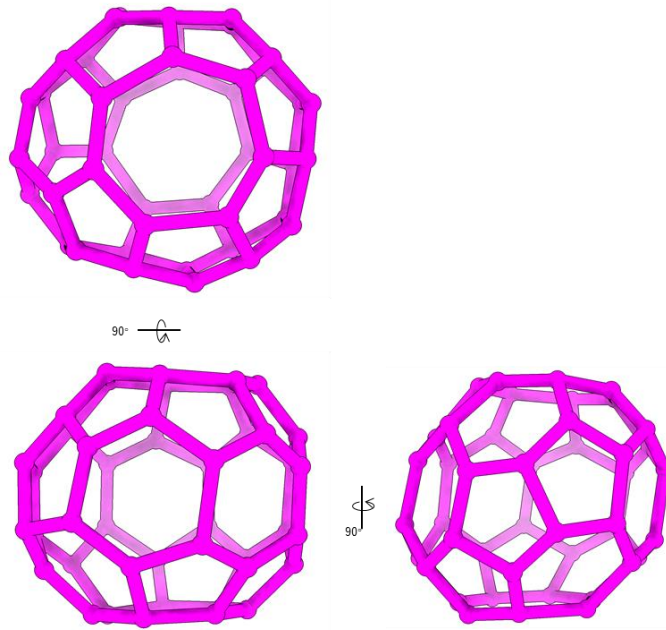
Supplementary Figure 4- 6. Comparison the IsoNet reconstruction with ICON and MBIR reconstructions.

a, Orthogonal slices of IsoNet, ICON and MBIR reconstructions of tilt series of thin sections provided in IMOD tutorial. **b**, Orthogonal slices of reconstructions of cryoET tomogram of neuronal synapses in cultured cells.



Supplementary Figure 4- 7. Orthogonal views of the tomogram containing clathrin cages reconstructed with the SIRT algorithm in IMOD.

PSD: postsynaptic density.



Supplementary Figure 4- 8. 3D views for the shape of the heptagon containing clathrin cage.

4.12 References

- 1 Beck, M. *et al.* Visual proteomics of the human pathogen *Leptospira interrogans*. *Nat Methods* **6**, 817-823, doi:10.1038/nmeth.1390 (2009).
- 2 Lucic, V., Forster, F. & Baumeister, W. Structural studies by electron tomography: from cells to molecules. *Annu Rev Biochem* **74**, 833-865, doi:10.1146/annurev.biochem.73.011303.074112 (2005).
- 3 Bloodgood, R. A. Sensory reception is an attribute of both primary cilia and motile cilia. *J Cell Sci* **123**, 505-509, doi:10.1242/jcs.066308 (2010).
- 4 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).
- 5 Liu, Y. T., Tao, C. L., Lau, P. M., Zhou, Z. H. & Bi, G. Q. Postsynaptic protein organization revealed by electron microscopy. *Curr Opin Struct Biol* **54**, 152-160, doi:10.1016/j.sbi.2019.02.012 (2019).
- 6 Liu, Y. T. *et al.* Mesophasic organization of GABAA receptors in hippocampal inhibitory synapses. *Nat Neurosci* **23**, 1589-1596, doi:10.1038/s41593-020-00729-w (2020).
- 7 Tao, C. L. *et al.* Differentiation and Characterization of Excitatory and Inhibitory Synapses by Cryo-electron Tomography and Correlative Microscopy. *J Neurosci* **38**, 1493-1510, doi:10.1523/JNEUROSCI.1548-17.2017 (2018).
- 8 Deng, Y. *et al.* ICON: 3D reconstruction with 'missing-information' restoration in biological electron tomography. *J Struct Biol* **195**, 100-112, doi:10.1016/j.jsb.2016.04.004 (2016).
- 9 Yan, R., Venkatakrishnan, S. V., Liu, J., Bouman, C. A. & Jiang, W. MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *J Struct Biol* **206**, 183-192, doi:10.1016/j.jsb.2019.03.002 (2019).

- 10 Zhai, X. *et al.* LoTToR: An Algorithm for Missing-Wedge Correction of the Low-Tilt Tomographic 3D Reconstruction of a Single-Molecule Structure. *Sci Rep* **10**, 10489, doi:10.1038/s41598-020-66793-1 (2020).
- 11 Mastronarde, D. N. Dual-axis tomography: an approach with alignment methods that preserve resolution. *J Struct Biol* **120**, 343-352, doi:10.1006/jsbi.1997.3919 (1997).
- 12 Guesdon, A., Blestel, S., Kervrann, C. & Chretien, D. Single versus dual-axis cryo-electron tomography of microtubules assembled in vitro: limits and perspectives. *J Struct Biol* **181**, 169-178, doi:10.1016/j.jsb.2012.11.004 (2013).
- 13 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).
- 14 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185, doi:10.1038/s41592-020-01049-4 (2021).
- 15 Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv*, arXiv:1603.04467 (2016).
- 16 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).
- 17 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).
- 18 Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).
- 19 Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sc* **9351**, 234-241, doi:10.1007/978-3-319-24574-4_28 (2015).

- 20 Moran, N., Schmidt, D., Zhong, Y. & Coady, P. Noisier2Noise: Learning to Denoise From Unpaired Noisy Data. *CVPR*, 12064-12072 (2020).
- 21 Xu, J. *et al.* Noisy-As-Clean: Learning Unsupervised Denoising from the Corrupted Image. . *arXiv* (2019).
- 22 Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157-161, doi:10.1038/s41586-020-2833-4 (2020).
- 23 Zhang, X. *et al.* Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. *Nat Commun* **7**, 13223, doi:10.1038/ncomms13223 (2016).
- 24 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).
- 25 Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat Methods* **13**, 387-388, doi:10.1038/nmeth.3806 (2016).
- 26 Schur, F. K. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508, doi:10.1126/science.aaf9620 (2016).
- 27 Noble, A. J. *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7**, doi:10.7554/eLife.34257 (2018).
- 28 Tan, A., Pak, A. J., Morado, D. R., Voth, G. A. & Briggs, J. A. G. Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2020054118 (2021).

- 29 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, doi:10.1073/pnas.1103634108 (2011).
- 30 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, doi:10.1371/journal.pone.0025700 (2012).
- 31 Hoog, 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, doi:10.1016/j.jsb.2012.01.009 (2012).
- 32 Zhang, J. *et al.* Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell Discov* **7**, 51, doi:10.1038/s41421-021-00281-2 (2021).
- 33 Fernandez-Busnadiego, R. *et al.* Quantitative analysis of the native presynaptic cytomatrix by cryoelectron tomography. *J Cell Biol* **188**, 145-156, doi:10.1083/jcb.200908082 (2010).
- 34 Mettlen, M., Chen, P. H., Srinivasan, S., Danuser, G. & Schmid, S. L. Regulation of Clathrin-Mediated Endocytosis. *Annu Rev Biochem* **87**, 871-896, doi:10.1146/annurev-biochem-062917-012644 (2018).
- 35 Rodal, A. A. & Littleton, J. T. Synaptic endocytosis: illuminating the role of clathrin assembly. *Curr Biol* **18**, R259-261, doi:10.1016/j.cub.2008.02.014 (2008).
- 36 Man, H. Y. *et al.* Regulation of AMPA receptor-mediated synaptic transmission by clathrin-dependent receptor internalization. *Neuron* **25**, 649-662, doi:10.1016/s0896-6273(00)81067-3 (2000).
- 37 Paraan, M. *et al.* The structures of natively assembled clathrin-coated vesicles. *Sci Adv* **6**, eaba8397, doi:10.1126/sciadv.aba8397 (2020).

- 38 Morris, K. L. *et al.* Cryo-EM of multiple cage architectures reveals a universal mode of clathrin self-assembly. *Nat Struct Mol Biol* **26**, 890-898, doi:10.1038/s41594-019-0292-0 (2019).
- 39 Wang, R. *et al.* Electron cryotomography reveals ultrastructure alterations in platelets from patients with ovarian cancer. *Proc Natl Acad Sci U S A* **112**, 14266-14271, doi:10.1073/pnas.1518628112 (2015).
- 40 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).
- 41 Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* (2014).
- 42 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
- 43 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
- 44 Imhof, S. *et al.* Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*. *Elife* **8**, doi:10.7554/eLife.52058 (2019).
- 45 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).
- 46 Turonova, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J Struct Biol* **199**, 187-195, doi:10.1016/j.jsb.2017.07.007 (2017).

- 47 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76, doi:10.1006/jsbi.1996.0013 (1996).

Chapter 5: A streamlined cryoET software pipeline with automatic particle picking on flexible lattices

Hui Wang^{1,2,3,*}, Shiqing Liao^{2,3}, Xinye Yu³, and Z. Hong Zhou^{1,2,3,*}

¹Department of Bioengineering, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

²California NanoSystems Institute, UCLA, Los Angeles, CA 90095, USA

³Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095, USA

***Corresponding author:**

Z. Hong Zhou (Hong.Zhou@UCLA.edu, 1-310-694-7527)

ORCID:

Hui Wang: 0000-0002-9922-7170

Shiqing Liao: 0000-0002-9723-8968

Xinye Yu: 0000-0002-1764-1141

Z. Hong Zhou: 0000-0002-8373-4717

Keywords: cryoET and STA workflow; automatic particle picking; protein in lattice configuration; deep learning

5.1 Abstract

Cryogenic electron tomography (cryoET) is capable of determining *in situ* biological structures of molecular complexes at sub-nanometer to near atomic resolution by subtomogram averaging (STA) of tens of thousands of copies of the complex. While abundant complexes (referred to as “particles” herein) often exist in polymeric arrays that manifest as flexible and imperfect lattices in one to three dimensions, precisely locating and seamlessly averaging such particles across many tomograms present major challenges. Here, we have developed TomoNet, a software package with a modern graphical user interface (GUI) to carry out the entire pipeline of cryoET and STA to achieve high resolution. TomoNet features automatic particle picking and 3D classification functions and integrates commonly used cryoET packages to streamline high-resolution STA for structures that constitute flexible lattices. Automatic particle picking is accomplished in two complementary ways, one based on a geometric template matching, and the other employing deep learning. TomoNet’s hierarchical file organization and visual display of intermediate results facilitate efficient data management required for large cryoET projects; and its interfacing utilities enable inter-operability with other leading data processing platforms. Applications of TomoNet to three types of cryoET data demonstrate its capability of efficient and accurate particle picking on flexible and imperfect lattices to obtain high-resolution 3D biological structures: virus-like particles formed by human immuno-deficiency virus Gag proteins, bacterial outermost cell wall made of surface layer proteins within lamellae generated by focused-ion-beam milling, and membranes decorated with nuclear egress protein complexes. These results demonstrate TomoNet’s potential for broad applications to various cryoET projects targeting high-resolution *in situ* structures.

5.2 Impact Statement

Cryogenic electron tomography (cryoET) has become a powerful approach to visualize organization and high-resolution structures of biological complexes as they exist in their native

environment. Subtomogram averaging (STA) of hundreds of thousands of copies of such complexes is necessary to obtain near atomic resolution structures. While abundant biological complexes often cluster in arrays manifest as one to three-dimensional lattices, flexibility and imperfection of such lattices pose challenges for efficient and accurate particle picking. To overcome these challenges and to meet the growing demand for efficient data processing and management in the cryoET and STA workflow, we have developed TomoNet, a user-friendly software package with a modern graphical user interface that allow users to execute the entire data processing pipeline seamlessly with integration of commonly used software packages. TomoNet addresses the particle picking challenge by two solutions, one is based on geometric template matching, and the other employs artificial intelligence. Application of TomoNet to three representative datasets demonstrate its capability for high-resolution structure determination of biological complexes on flexible and imperfect lattices.

5.3 Introduction

Single-particle cryogenic electron microscopy (cryoEM) is employed to elucidate atomic-level structures of purified biological proteins and homogeneous complexes. This methodology adheres to a standardized and well-established workflow, supported by advanced software packages such as Relion¹ and cryoSparc². In parallel, cryogenic electron tomography (cryoET), coupled with subtomogram averaging (STA), expands the investigative scope to encompass heterogeneous macromolecules in their native context³⁻¹⁰. To enhance the resolution of subunits within *in situ* macromolecules, subtomograms (*i.e.*, particles) are extracted from each tomogram and then subjected to 3D alignment and averaging, thereby improving signal-to-noise ratio. Notably, STA has achieved resolutions up to sub-3 Å for *in situ* structures of large cellular complexes such as ribosomes, approaching the capabilities of single-particle cryoEM methodologies¹¹⁻¹⁴.

The workflow for cryoET and STA typically involves five key components across specific

software packages. In cryoET preprocessing, dose fractionated frames are collected from an electron microscope, motion correction is performed on those frames, and motion corrected images are organized and stacked into individual tilt series. In tomogram reconstruction, three-dimensional reconstructions are generated from generated tilt series. In particle picking, particles of interest are identified and extracted from tomograms. Complexity varies based on the diverse and intricate nature of *in situ* cellular samples and their unique configurations. Many packages include its own particle picking methods, such as geometry oversampling in Dynamo¹⁵, machine learning in crYOLO¹⁶, and template matching in emClarity¹⁷. In 3D refinement and classification, particles are iteratively classified and refined to improve resolution. Structural variations are identified by software like Relion^{13,18}, emClarity¹⁷, EMAN2⁴ and Warp¹⁹. Finally, activities in post-processing include map sharpening, Fourier shell correlation (FSC) calculation, placing back averaged maps into the original tomogram, etc. Users often need to navigate between several specialized software packages for optimal results, which poses a barrier to users and often demands a certain level of computational proficiency.

The method for particle picking varies on a case-by-case basis, dictated by the characteristics of *in situ* cellular samples. In the early works of STA, manual particle picking was employed, particularly when aiming for resolutions between 20-50 Å with a maximum of several hundred particles²⁰⁻²². However, for biological samples exhibiting periodic structures, oversampling can be leveraged to significantly reduce the labor associated with acquiring enough particles for improved resolution. For instance, HIV virus-like particles (VLPs) adopt a hexagonal Gag protein lattice in its sphere-like configuration²³. With an increasing demand for automation to enhance efficiency with minimal manual intervention, template matching has emerged as a popular method for automatic particle picking, relying on a user-provided reference map^{17,24}. Simultaneously, deep neural network shows promising results for cryoET automatic particle picking given its capacity to analyze three-dimensional feature maps and

autonomously identify prominent features within specific samples²⁵⁻²⁸. This deep learning approach typically operates in a template-free fashion and sometimes even obviates the need for human annotation²⁹.

The expanding array of specialized software tools designed for specific tasks post a critical need for seamless software integration within the cryoET workflow. Transitioning between various software packages can be a cumbersome process, often requiring a significant level of computational proficiency. Remarkably, recent initiatives have made notable progress in tackling this integration challenge. For example, TomoBEAR³⁰ offers an integrated solution while ScipionTomo³¹ and nextPYP³² provides a comprehensive web-based platform for managing various tasks in the cryoET pipeline. Notably, none of these packages takes specific advantage of the fact that many abundant complexes exist in arrays of some sort, albeit with imperfections, variability, or flexibility.

In this context, we have developed TomoNet, a software package designed for streamlining the entire cryoET data processing workflow, with a modern GUI (Fig. 1, Suppl Fig. 1). Our methodology employs a geometric template matching approach, rooted in the concept of "Auto Expansion," which serves as a general particle picking solution for protein complexes organized in flexible, variable, or imperfect arrays. TomoNet is also powered by a deep learning-based solution to automate particle picking, which only needs 1-3 tomograms with known particle locations as ground truth for model training. Importantly, while TomoNet is particularly powerful for locating and averaging particles arranged on flexible or imperfect lattices, it can be applied to a broader range of particle types, offering a more generalizable trained model. These methods significantly diminish the need for manual inputs, and their outcomes can be seamlessly imported into Relion for subsequent high-resolution 3D classification and refinement. We demonstrate capabilities of TomoNet by applying it to three datasets with distinct protein lattice types, highlighting its accuracy and efficiency in identifying particles across diverse scenarios.

5.4 Results

5.4.1 Overall design of TomoNet

TomoNet is a Python-based software package that integrates commonly used cryoET packages to streamline the cryoET and STA pipeline, with a particular emphasis on automating particle picking of lattice-configured structures and cryoET project management. As shown in the main menu and the entire pipeline (Fig. 1), after data collection from electron microscopy, TomoNet can perform motion correction with integration of MotionCorr2³³; stacked tilt series generation and tomogram reconstruction with integration of IMOD³⁴; CTF estimation with integration of CTFFIND4³⁵; manual particle picking with IMOD; particle picking with build-in geometry-based template match algorithms and integration of PEET³⁶; automatic particle picking with build-in deep learning-based algorithms; particle cleaning and subtomograms placing back with build-in algorithms. The design allows for on-the-fly processing of tomogram reconstruction during data collection which facilitates quick quality check. TomoNet generates particle picking results in STAR format³⁷ which can be incorporated in Relion for high-resolution 3D refinement, it also can read Relion results in STAR format for particle cleaning and subtomograms place back (Fig. 1).

5.4.2 “Auto Expansion” particle picking

"Auto Expansion" module is designed to pick particles on flexible lattices with minimal manual inputs, its basic concept is elucidated in Figure 2. Such particles exist in array configurations of some sort, manifest as flexible, partial, and imperfect lattices of one, two and three dimensions (1-3D). Examples abound: microtubule doublets, ubiquitous in most cells, consist of 96 nm axonemal 1D translational repeat units²¹; The external sheath hoop of archaeal cells consists of subunits repeat every 2.8 nm along the circumferential direction³⁸ (1D rotational lattice); HIV VLPs³⁹ and surface layer lattice of prokaryotic cells^{40,41} are made of hexameric subunits (2D lattice); The paraflagellar rod (PFR) of protozoan species is organized into para-crystalline arrays in its distal zone²⁰ (3D lattice). Within the input tomograms, each of these distinct lattice

densities is called a patch, for instance, there are two distinct patches illustrated in [Figure 2](#).

To start “Auto Expansion”, users need to prepare only a few “seed” particles that sparsely distribute over all patches observed. A typical number of such “seed” particles per tomogram ranges from 20 to 200, which depends on the number and size of patches. Then, “Auto Expansion” expands the particle set by adding new unpicked particles for each tomogram in an iterative manner; each iteration includes the following three steps ([Fig. 2](#)). First, potential particles that are adjacent to each “seed” particle are selected as “candidate” particles. Second, those “candidate” particles undergo alignments to user-provided reference, and evaluation based cross-correlation coefficient, such that “wrong” particles with low cross-correlations can be excluded. Third, qualified “candidate” particles are added into the final particle set, and they become “seed” particles for the next iteration. “Auto Expansion” stops until either no “candidate” particles detected, or it reaches the user-defined maximum iteration number. The final particle set finally could encompass all available particles within the selected patches, expanded from only a few “seed” particles through above iterative iterations.

“Auto Expansion” is a template matching-based method, since cross-correlation coefficient is used in the evaluation process. However, compared with conventional template matching, “Auto Expansion” incorporates the prior knowledge of lattice configuration, to guide the search for unpicked particles iteratively by following user-defined paths, as detailed in TomoNet’s user manual. Thus, “Auto Expansion” can significantly reduce computational complexity by searching in the regions of interest only, with restricted angular and translational search range defined by users; as a result, it reduces the number of incorrectly picked particles. Notably, “Auto Expansion” works for flexible, imperfect, or variable lattices in 1D, 2D and 3D, and has no intrinsic size limit of subunits.

5.4.3 Automatic particle picking by deep learning

TomoNet provides another particle picking module named “AI AutoPicking”, which is designed

for automatic particle picking using deep learning. “AI AutoPicking” uses supervised learning and requires an input training dataset consisting of just 1-3 tomograms paired with their corresponding particles coordinate files, and it can output predicted particles for tomograms in the entire dataset. There are three main steps in “AI AutoPicking”: training data preparation, neural network training, and particles coordinate prediction as detailed in Method section (Fig. 3).

Compared to the template matching-based “Auto Expansion”, “AI AutoPicking” has several advantages. First, “AI AutoPicking” does not use prior knowledge of particle distribution pattern, which means it works for both particles on flexible lattices, and ones distributed in scattered pattern like cellular ribosomes. Second, it utilizes GPU for fast convolution operations which enables particles prediction in just several minutes. Third, “AI AutoPicking” does not require the “seed” particles used in “Auto Expansion”, which further reduce human efforts around 5 minutes per tomograms, especially when processing a large tomography dataset could including up to hundreds of tomograms. However, as the final output, “AI AutoPicking” normally picks less particles than “Auto Expansion” because of the miss of certain particles on the flexible lattices. To further explore those missing particles, TomoNet allows the users to use the “AI AutoPicking” predicting particles as “seed” particles at the start of “Auto Expansion”.

5.4.4 3D Classification using TomoNet

In addition to the above two commentary modules for particle picking, TomoNet allows users to eliminate “bad” particles based on user-defined geometric constraints, which could serve as 3D classification during high-resolution particle refinements. Lattice variation present in cryoET data could have multiple causes, including biological and experimental causes. For example, particles may be incomplete near the lattice edge as a result of paused biology assembly process⁴². Experimentally, lattice tends to become flattened near the air-water interface of the sample. These variabilities cause difficulties for 3D classification in the process of high-

resolution STA and make it hard to exclude the “bad” particles, which normally show unexpected coordinates and orientations assignment as subunits of lattices.

Removing of those “bad” particles is necessary to achieve better final resolutions⁴³. To accomplish this, for each particle, TomoNet counts its neighboring particles, and calculates the averaged tilt angle to their neighbors which represents local surface curvature of a lattice. For those particles with too few neighbors or having large tilt angles to the neighbors, TomoNet considers them as “bad” particles since they possibly deviate from the lattice configuration. This step can be integrated into high-resolution refinement in Relion to be an alternative 3D classification method based on analyzing spatial relationships between particles.

5.4.5 Application to in situ viral protein arrays: the matrix protein lattice in HIV VLPs

To validate TomoNet as an integrated high-resolution STA pipeline and an efficient particle picking tool, we processed four tomograms from the HIV-1 Gag dataset from EMPIAR-10164³⁹ and resolved the Gag hexamer structure at 3.2 Å resolution. We performed tilt series assembly, CTF estimation, tomographic reconstruction using TomoNet after downloading the motion corrected images. Within these tomograms, the VLP hexagonal lattice and its building blocks were observed from different views and some of these observed VLPs exhibit sphere-like geometry (Fig. 4a).

As detailed in the Method section, we applied a combination of “Auto Expansion” and “AI AutoPicking” on the generated four tomograms, as the outcome, particles were readily picked on all the observed lattice patches (Fig. 4b, c). Then, those picked particles were imported to Relion package to perform high-resolution refinements and get the final reconstruction of Gag hexamer structure (Suppl Fig. 2).

Using the subtomogram place back function in TomoNet, we generated 3D visualizations of the *in situ* architecture of the VLPs lattices (Fig. 4d, Suppl Fig. 3), which

supports that TomoNet is capable of picking particle with arbitrary lattice geometry. Such 3D visualization also enabled us to identify the lattice defects on each VLPs that are consistent with previous studies⁴⁴, which could potentially facilitating the understanding of lattices assembly mechanism⁴⁵.

5.4.6 Application to focused ion beam (FIB)-milled cellular sample: the surface layer lattice of prokaryotic cell

We validated TomoNet's particle picking capability by processing one tomogram of FIB-milled *Caulobacter crescentus* cells from EMD-23622⁴⁶. The surface layer (S-layer) covers the cell body as a component of cell wall; its lattice geometry is normally defined by the shape of cells (Fig. 5a). The *C. crescentus* cell's shape is pleomorphic with variable sizes, which raises difficulty for efficient particle picking on its S-layer lattice. Besides the uncharacterized lattice shape, the low contrast shown in this tomogram also hindered locating subunits on the S-layer lattice (Fig. 5a).

TomoNet can overcome the above challenges by utilizing the hexagonal configuration of S-layer lattices. With a minimal manual input, "Auto Expansion" easily picked over a thousand hexamer S-layer subunits. The intermediate subtomogram average result shows the hexagonal distribution of S-layer inner domains clearly, which is hard to recognize in the tomogram because of low contrast (Fig. 5b). Visualization of S-layer lattices also showed that the picked particle arranged in the expected hexagonal arrangement, again confirms the reliability and applicability of TomoNet and as a particle picking tool (Fig. 5c).

5.4.7 Application to in vitro assembled arrays: nuclear egress complex (NEC) lattice

We also validated TomoNet as an integrated high-resolution STA pipeline and an efficient particle picking tool by processing sample containing NEC lattices budded vehicles. Nuclear egress is a pivotal step in herpes virus replication driven by NEC, and responsible for

translocating nascent virus particles from nucleus to cytoplasm. In our reported dataset⁴⁷, NEC heterodimers budded into large vesicles with diameters spanning from 100 nm to 500 nm, and formed beehive-like lattices in the inner surface of these vesicles (Fig. 6a, b). Because of their large sizes, noticeable compressions caused in the process of sample freezing were observed, making the vesicles and NEC lattices reshaped from sphere to flattened disk shapes (Fig. 6a, b). This conformational change is due to the ice thickness limitation in cryoET that restricts the sample thickness to approximately 250 nm, thus, poses challenges for particle picking.

TomoNet is capable of picking the hexamer subunits using “Auto Expansion” following the NEC lattice topology. The intermediate averaged result generated in TomoNet already showed the six heterodimers in one hexamer subunit (Fig. 6c). With those picked particles, we achieved the reconstruction of NEC hexamer subunit at 5.4 Å resolution, with no preferred orientation bias (Fig. 6d) and all the helices were well resolved (Fig. 6e). Visualization of an NEC lattice generated by placing back averaged maps shows that the large vesicle is compressed, and the compression generated by sample freezing stretched the lattice and makes it appear flat and split at the air-water interface, while the middle part of the lattice appears to be more curved.

5.5 Discussion

We have demonstrated the remarkable efficacy of TomoNet's particle picking on three distinct datasets featuring particles with varying lattice configurations. TomoNet stands out as the first software to exhaustively trace lattices following its inherent topology. This unique approach ensures that the particle picking results faithfully reflect *in situ* or *in vitro* lattices shape, providing valuable insights into how these lattices are formed by their constituent subunits. In the case of HIV VLPs, employing our subtomogram place back function enabled us to directly visualize the VLPs lattices. VLPs lattice defects are potentially caused by the absence of pentamer subunits. Similarly, for NEC dataset, TomoNet facilitated a more direct observation of lattice conformation

changes resulting from the sample freezing process. Since vesicles in this dataset were too large to be compressed from a sphere into a disk-like shape, the lattice regions near air-water interface became stretched and subsequently divided into smaller fragments. Moreover, TomoNet has demonstrated its exceptional performance, even when dealing with datasets characterized by extremely low contrast. For instance, in the cellular S-layer tomogram of a lamella, S-layer subunits were nearly imperceptible to human observations. Thus, "Auto Expansion" excelled in particle picking without requiring denoising or contrast-enhancement algorithms.

Additionally, "AI AutoPicking", the deep learning-based module, demonstrated great performance on automatic particle picking. It showcased potential in handling a wide range of particle types, extending beyond those arranged in lattice-like arrangements. Some implemented deep learning-based particle picking method uses 2D slices for the network training such as in crYOLO¹⁶, while TomoNet directly use 3D volumes for that purpose. This definitely enriched the training dataset with an additional dimension, but the missing wedge artifact induced in 3D tomogram may cause misleading effects to neural network learning.

Previously, achieving high-resolution cryoET and STA was predominantly hindered by several factors. First, the cryoET data collection process is inherently more intricate due to the involvement of stage tilting, which impedes the speed of data collection. Second, computational algorithms for cryoET and STA were not as mature, which constrained their full potential. Third, efficient particle picking methods were essential but not always readily available. However, the landscape of cryoET has evolved significantly, largely addressing these limitations. With advancements in high-throughput tilt series collection, acquiring large tomogram datasets has become more feasible and cost-effective⁴⁸⁻⁵⁰, which is a fundamental requirement for achieving high-resolution STA. Developments in 3D CTF correction⁵¹ and per-particle refinement algorithms^{4,13,17,52} have alleviated the bottleneck of cryoET 3D refinement, making sub-3Å STA achievable. In parallel, template matching and neural networks now play significant roles in

particle location and orientation determination for high-resolution STA^{15-17,25-29}. These techniques have significantly accelerated the particle picking step, reducing the time required from weeks to just a few days or even hours, and thus streamlining the entire workflow. Despite these advancements, completing the cryoET and STA pipeline often entails navigating between multiple specialized software packages to optimize results. This process is not always convenient and often demands a certain level of computational proficiency. To address these challenges, TomoNet offers an intuitive solution for managing large tomogram datasets and intermediate results, eliminating the need for users to engage in time-consuming transitions between various cryoET packages or tools. Additionally, TomoNet provides a user-friendly GUI, particularly beneficial for inexperienced users or those with limited computational backgrounds, as demonstrated in previous studies^{13,19,42}.

Each module within TomoNet is intentionally designed to be highly independent, ensuring flexibility for seamless integration of future advancements in specific tasks. This adaptable framework positions TomoNet as an ideal platform for other developers to incorporate their innovations. At present, TomoNet is primarily tailored for integration with the Relion-related pipeline. However, it possesses the capability to accommodate specific demands and can be extended to integrate alternative pipelines, including emClarity¹⁷, EMAN2⁴, M⁵², and potentially others in the future. In summary, TomoNet significantly simplifies the overall process for users in managing and monitoring every step of the complete cryoET and STA pipeline. Its user-friendly GUI design notably reduces the entry barrier for newcomers to this field. The particle picking modules of TomoNet furnish a general solution for particles organized in lattice-like arrangements, ensuring both accuracy and efficiency, thereby facilitating the high-resolution STA pipeline.

5.6 Methods

Software implementation

TomoNet is an open-source software package developed using Python. It follows a highly modularized architecture with each module responsible for specific tasks in a typical cryoET and STA data processing pipeline. Built-in tools in TomoNet mainly cover the upper stream of the cryoET and STA pipeline including procedures of motion correction, tilt series generation, tomogram reconstruction, CTF estimation and particle picking, while leave the high-resolution 3D refinement to established software package like Relion (Fig. 1). The design of a modern GUI enhances user-friendliness and helps with tracking the processing progress (Suppl Fig. 1). With the table view, users can have a census overview of the entire dataset, this allows the direct and intuitive management for each tomogram (Suppl Fig. 1).

Motion Correction, tomogram reconstruction and CTF estimation

Motion correction, tomogram reconstruction, and CTF estimation are organized into separate modules in TomoNet, with the integration of corresponding external software: MotionCorr2, IMOD and CTFFIND4, respectively. Since we do not rewrite the code in the TomoNet package, users have to install the external software before using the corresponding module.

The “Motion Correction” module is used to correct beam-induced sample motion. First, it requires an input folder path that contains all the dose fractionated frames. Then, user can specify the MotionCorr2 parameters in the GUI. Finally, after clicking the run button, TomoNet will perform motion correction on all the input images and save the result images in a separated folder. This feature also enables on-the-fly motion correction during data collection.

The “3D Reconstruction” module comprises two tabs: “TS Generation” and “Reconstruction”. Using the “TS Generation” tab, users can readily generate tilt series from the above motion corrected images. It also provides options for data cleaning, “min # tilts” helps to set up the minimum number of tilt images in a tomogram, and “rm dup” helps to remove duplicate images at the same tilt angle by excluding the images with older time stamps. The result of “TS generation” are the stacked tilt series files and their corresponding tilt information files. The “Reconstruction” tab automatically reads all the generated tilt series and lists them in a

table view with essential information such as tilt image number and alignment errors for simplifying the assessment of the reconstruction results and direct tomogram reconstruction management including restart, continue and delete.

The "CTF Estimation" module is used for the tilt series defocus estimation, and it supports parallel processing through multiple CPUs. Similar to the "Reconstruction" module, it presents outcomes in a table view with visualization features like displaying defocus at 0 degree and plotting the defocus distribution across all tilts.

Prepare "seed" particles

To perform STA, users need to identify target subunits, i.e., particles, within the outcome tomograms, in some cases, those particles arrange in flexible lattice configuration. TomoNet contains the "Manual Picking" module for the purpose of "seed" particle set in "Auto Expansion" particle picking. Normally, there are both top/bottom and side views that can be seen in the tomograms (Fig. 4a). We implemented the IMOD *stalknit* for the manual particle picking by defining the Y-axis of each particle which is typically the rotating axis. Thus, picking takes 2 points for each side view particle and 1 point for each top/bottom view particles (Fig. 4a). In the example of HIV dataset, we manually picked 5-10 particles as the "seed" particles for each VLPs, and it only takes several minutes for the entire tomogram picking. This module also uses the table view for easy management of the manual picking.

"Auto Expansion"

We decompose the "Auto Expansion" into three steps: "Generate tomogram.star", "Generate Picking Parameter" and "Run Particle Expansion". "Generate tomogram.star" is used to generate a STAR format file that defines what are the tomograms will be used in the "Auto Expansion". "Generate picking parameter" is used to set up the searching parameters that are required in the particle set expansion including: angular search ranges and steps, translational search ranges and steps, transition list, alignment box size, subunit repeating distance, reference and mask map, cross correlation threshold, etc. The transition list is a vector array,

the vectors represent the 3D translational shifts in pixel from the “seed” particle center to its associated “candidate” particles centers, as measured in the provided reference map. Click the button, it will generate a JSON format file containing all the defined picking parameters. “Run Particle Expansion” takes the above two files as the input to perform the iterative particle set expansion.

During the “Auto Expansion” running, three folders will be generated for each tomogram. “*TomoName*” is the working directory for the current round expansion, “*TomoName_cache*” stores the intermediate results for each iteration, and “*TomoName_final*” stores the final particle outcomes. Typically, this iteration number should be greater than one. However, the design of the “Auto Expansion” module allows for some special usage cases: Firstly, in the scenario with no known reference map, user can generate a rough reference map using the initial “seed” particles by setting the iteration number as 1 and select “refine reference” as “Yes”. Secondly, in the scenario when users wish to modify the picking setting such as a different cross-correlation threshold, or the exclusion of specific rounds due to an increased number of false positives, they can set the iteration number as 0. This prompts the program to skip the “candidate” searching steps, gathering all intermediate results stored in the “*TomoName_cache*” folders and generate the new “*TomoName_final*” result.

“AI AutoPicking”

“AI AutoPicking” comprise 3 main steps, “Prepare Training Dataset”, “Train Neural Network” and “Predict Particles coordinates”. TomoNet uses supervised learning, it requires the user to provide ground truth, i.e., tomogram with the associated particle coordinates information, for the model training. In this study, the ground truth data were prepared by the “Auto Expansion”.

First, the input tomograms will be converted to subtomograms as the network training dataset for two reasons. One is the whole tomogram size is typically around 1000x1000x1000 which is not efficient to be loaded in the GPU memory, the size of extracted subtomograms can be under 100x100x100. The other is it helps with increasing the training sample size to avoid

over fitting during the training. For the model output, we embedded the particles coordinate information into a same size 3D segmentation map where the pixels involved with particles were set to 1 otherwise 0 (Fig. 3a).

Next, the above extracted subtomograms with their associated segmentation maps can be used to train a neural network to learn how to segment the density map into region with or without particles. The network architecture we employ in the “AI AutoPicking” module is based on U-net as it is well-suited for capturing generalized features of 3D objects (Fig. 3b). Our neural network setup is derived from the U-net used in IsoNet⁴², but with the following key modification. The model output is changed from density map to segmentation map, such that the learning task here becomes pixel-wise classification. Thus, we use cross entropy loss function instead of minimum squared error (MSE). Equipped with one RTX 3080Ti graphic card, the training process can be completed swiftly within 1-2 hours using the default parameters.

Finally, users can use the trained model for particle prediction for the entire tomogram dataset (Fig. 3c). For each input tomogram, TomoNet crops it into subtomograms near the same size as the one in the training dataset. After applying the trained model on those subtomograms, the network output corresponding segmentation maps which will be assembled together to get a full-size segmentation map. Then, we employed the hierarchical clustering algorithm to extract the particles information from this segmentation map. The final resulting particles set will be saved individually for each tomogram.

Other Utilities

This “Other Utilities” module consists of two separate tabs: “Recenter | Rotate | Assemble to .star file” and “3D Place Back”. Once the particle picking process is completed, the “Other Utilities” module is used. The first subfunction allows the users to assemble and convert the particle picking results into STAR format file following the Relion4 convention, reset particles center to its symmetric center, and align the rotation axis to Z-axis which is required in Relion. The second subfunction takes a user-provided STAR format file that contains particles

information as input and generates a ChimeraX⁵³ session file for 3D subtomogram place back. On the one hand. This function allows users to validate the accuracy of particle picking before importing the STAR format particle file into Relion. On the other hand, it also enables direct observation of subunit distribution and configuration after the high-resolution 3D refinement in Relion, for an overall *in situ* lattice observation (Suppl Fig. 3).

Overall, the modularized structure offers users the flexibility to engage with TomoNet at any stage. Users can initiate their workflow from the raw stacked frames, or directly import 3D density maps for particle picking management. These features contribute to efficiency improvements and establish a seamless framework for effective project management with GUI.

Processing tomograms of HIV VLP dataset

We downloaded the HIV VLP dataset from the Electron Microscopy Public Image Archive (EMPIAR) with the accession code EMPIAR-10164³⁹. Four tilt series, TS_01, TS_43, TS_45 and TS_54, were used in this study. The micrographs were loaded into the TomoNet pipeline to perform tilt series generation, CTF estimation, and tomogram reconstruction using the WBP algorithm. Four-time binned tomograms with 5.4 Å pixel size were used for further particle picking. First, TS_01 was used for manual particle picking (“seed” particles) and generate the initial reference map, and then all the four tomograms were used for one run of “Auto Expansion” to get more particles and refine the reference. Then with the refined reference, 3 VLPs were selected in TS_01 for a new run of “Auto Expansion” (Fig. 4b), and the result was used for neural network training in “AI AutoPicking”. After the particle prediction with the trained model, we got result as shown in Figure 4c which is already acceptable. To exploit even more particles, the predicted particles results were again used for the final run of “Auto Expansion”. 17,625 particles were exported to Relion to perform high resolution refinement.

Following the same procedure carried out in the Relion4 tutorial, we achieved the Gag hexamer structure at 3.2 Å, with 13,558 particles from four tomograms finally after the geometric classification. Resolution was calculated in Relion and on 3DFSC Processing Server⁵⁴, the

global resolution reported is based on the “gold standard” refinement procedures and the 0.143 Fourier shell correlation (FSC) criterion (Suppl Fig. 2).

Processing tomograms of *C. Crescentus* surface layer

We downloaded the FIB-milled *C. crescentus* data of one reconstructed tomogram from Electron Microscopy Data Bank (EMDB) with the accession code EMD-23622⁴⁶. We directly use the tomogram for manual picking on two of the cells. Around 30 “seed” particles were manually picked and then performed template free STA using PEET to generate the initial reference map. Then “Auto Expansion” takes the “seed” particles and initial reference map for 5 iterations to get more particles such as to refine the reference map. With the improved reference map, another run of “Auto Expansion” applied to the “seed” particles again for 15 iterations to explore all possible particles on the outer surface of the cells, and finally get around 1,500 S-layer hexamer particles. TomoNet demonstrated its performance of particle picking for FIB-milled tomogram even with limited contrast. The S-layer have been observed as the cell envelope component in a wide range of bacteria and archaea with either hexagonal, square or other lattice arrangements, as well as distinct cellular shapes^{41,56}. Thus, TomoNet is well-suited to facilitate resolving the cell envelope structures and has broad applications.

Processing tomograms of NEC budding *in vitro*

CryoET grid preparation and data collection were previously described⁴⁷. The motion correction, tomogram reconstruction and CTF estimation were performed using TomoNet. Around 50-150 “seed” particles were manually picked for each tomogram. “Auto Expansion” were applied on a total of 35 tomograms and yield the ~48K particles before Relion refinement. Following one round of 3D auto-refine job under bin4 pixel size and several rounds of 3D auto-refine jobs under bin2 pixel size and one round of 3D auto-refine under bin1 pixel size, with the TomoNet geometric classifications in between, we resolved the NEC hexamer structure at 5.4 Å with 35,039 particles.

3D visualization

IMOD³⁴ was used to visualize the 2D tomographic and segmentation map slices. UCSF ChimeraX⁵³ was used to visualize the STA results and the lattices generated by 3D subtomogram place back. The atomic models were fitted into the density map using the “fit in map” tool in ChimeraX.

5.7 Availability

TomoNet code is available on Github website at <https://github.com/logicvay2010/TomoNet>.

For the HIV VLPs dataset, the raw data was deposited with accession code EMPIAR-10164, the Gag atomic model is downloaded from PDB with accession code 5L93³⁹. For the *C. Crescentus* S-layer dataset, the reconstructed tomogram was deposited with accession code EMD-23622, and the subunit for 3D place back was generated using atomic model with PDB accession code 6P5T⁵⁷. The STA map of NEC hexamer was deposited with accession code EMD-40224.

5.8 Acknowledgement

We thank Elizabeth Draganova and Ekaterina Heldwein for the NEC dataset. This project is supported by grants from the US National Institutes of Health (GM071940 to Z.H.Z.) and the National Science Foundation (DMR-1548924 to Z.H.Z.).

5.9 Authorship contributions

HW and ZHZ initialized and ZHZ supervised research; HW wrote the code and developed the software GUI with help from SL; HW, SL and XY tested the software on different datasets; HW, and ZHZ wrote the manuscript; XY assisted the manuscript writing; all authors reviewed and approved the paper.

5.10 Competing Interests Statement

The authors declare that there is no conflict of interest.

5.11 Figures

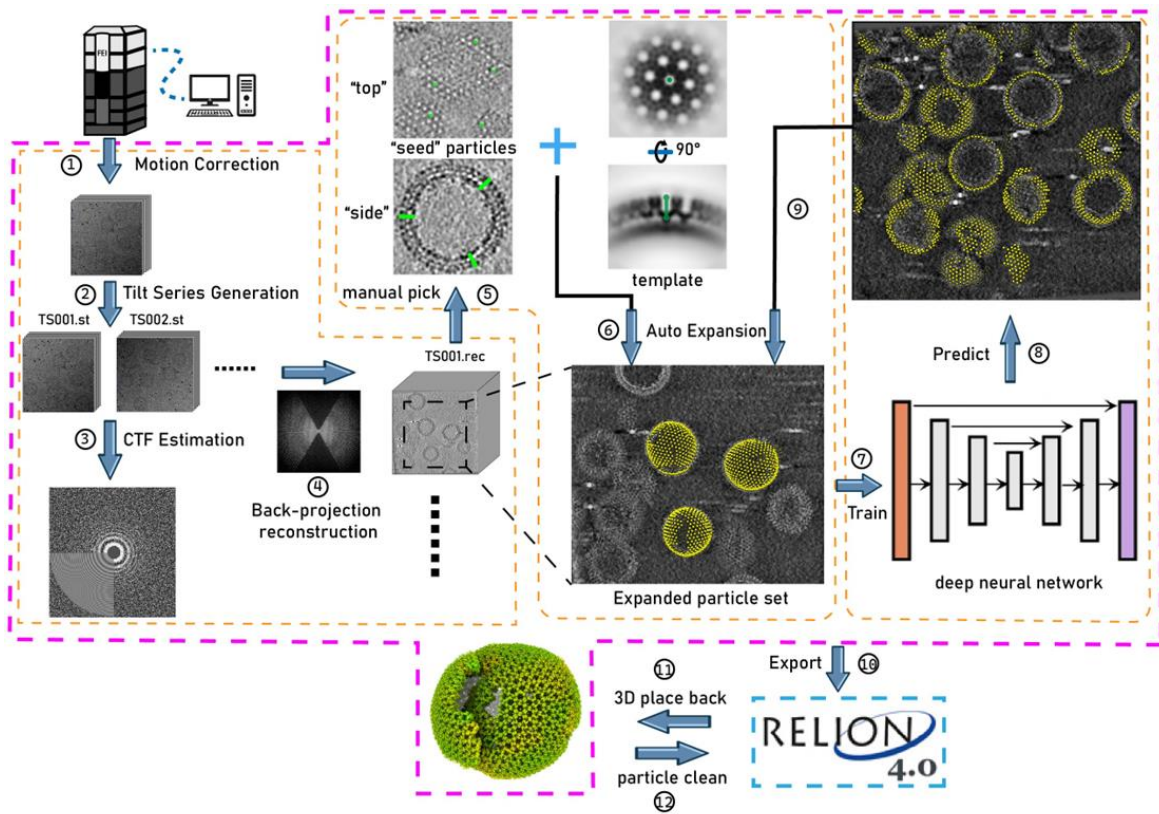


Figure 5- 1. Illustration of TomoNet’s comprehensive pipeline for cryoET and STA.

Enclosed within the pink border are the sequential functions that constitute the pipeline. This framework can be subdivided into three principal segments that precede the generation of the files needed for Relion, delineated by the orange border. These segments include tomogram preparation on the left, particle set expansion in the center, and deep learning-based automatic particle picking on the right.

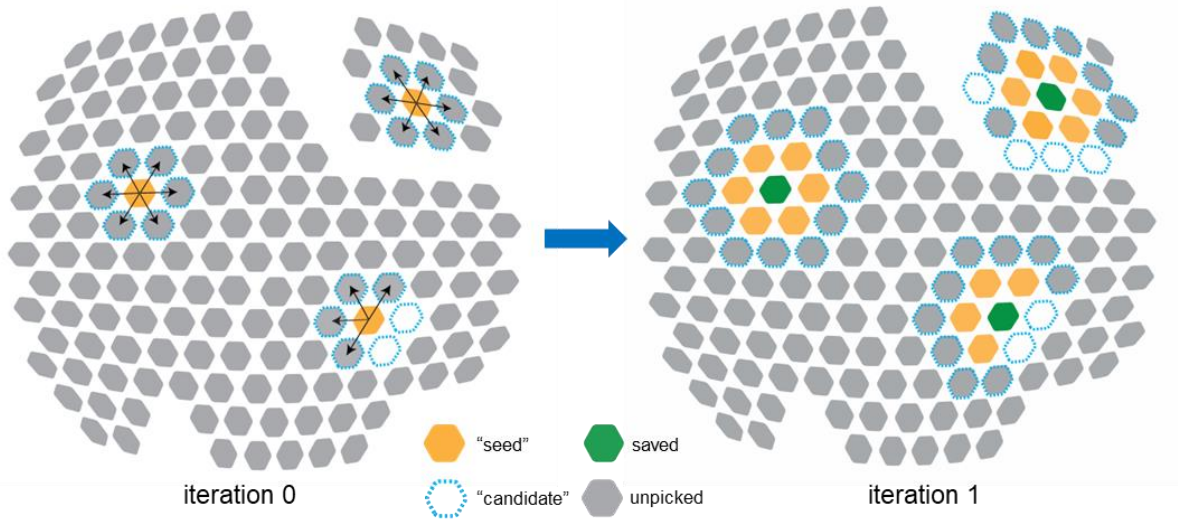


Figure 5- 2. Illustration of the first two iterations of "Auto Expansion" particle picking.

There are two patches of a hexagonal lattice with individual particles represented by solid hexagons. At iteration 0, 18 "candidate" particles were selected from the neighbors of 3 "seed" particles. 14 good particles remained and will serve as "seed" particles in iteration 1, and 3 "seed" particles were saved in the final particle set. At iteration 1, 35 "candidate" particles were selected from the neighbors of 14 "seed" particles. 29 good particles remained and will serve as "seed" particles in iteration 2, and 14 "seed" particles were saved in the final particle set. "Auto Expansion" is an iterative process and will stop when no "candidate" can be detected.

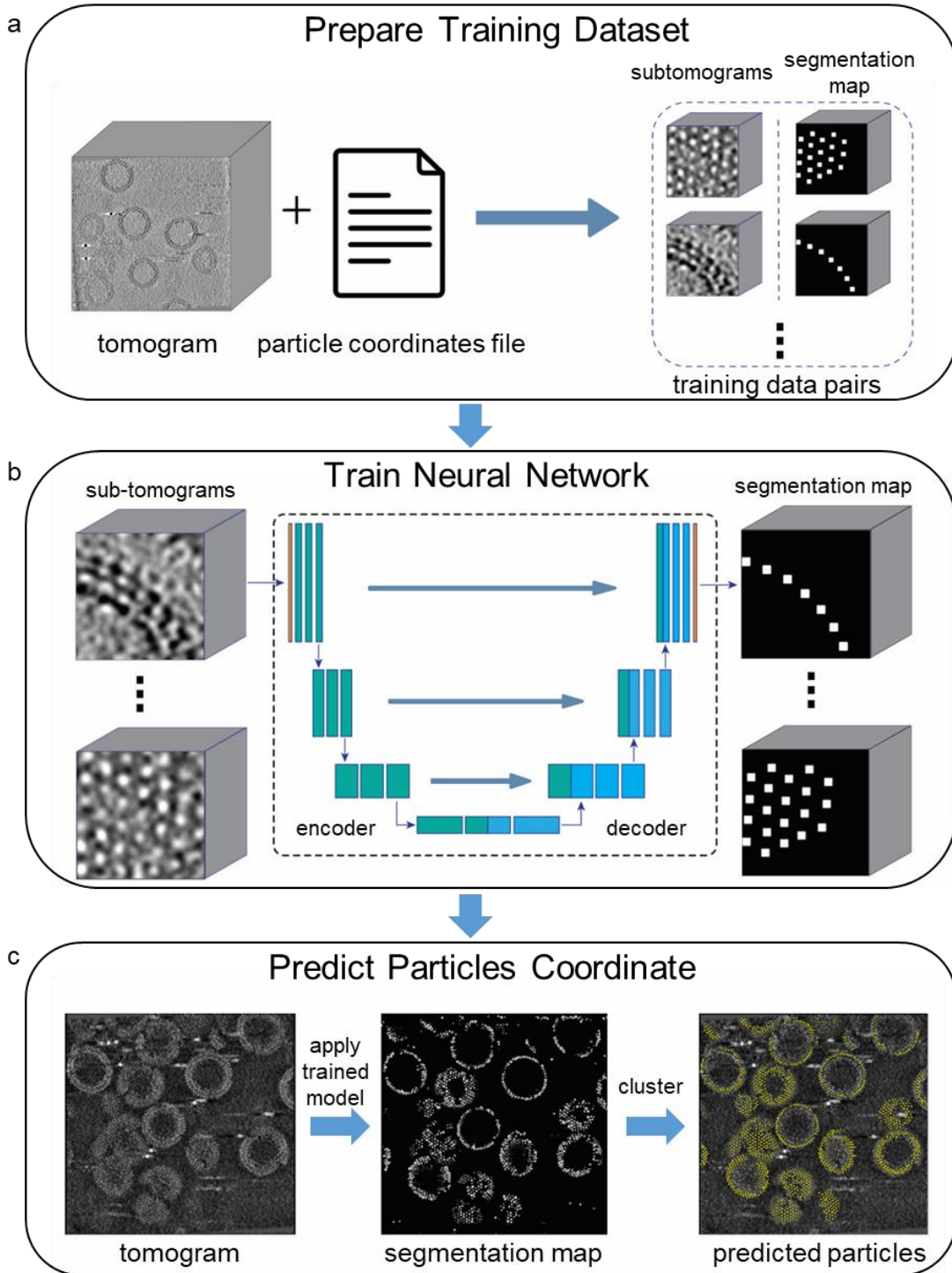


Figure 5- 3. Illustration of "AI AutoPicking" process consisting of three steps.

The HIV VLPs data were used for this illustration, and the particles referred to Gag hexamers. a, The first step is training dataset preparation. Using the user-provided tomograms with

associated particle coordinate files, subtomograms containing particle densities were extracted. For each of them, TomoNet generated a segmentation map based on the coordinates of particles, where the pixels near a particle's center are shown as white and the others as black. b, The second step is neural network training. The generated subtomograms and segmentation maps were used as the input and output to train the convolutional neural network in learning how to segment out particle densities. c, The third step is particle coordinate prediction. Firstly, TomoNet applied the trained neural network model to unseen tomograms and generated associated predicted segmentation maps. Then, the particle coordinate information was obtained from the segmentation maps using clustering algorithms.

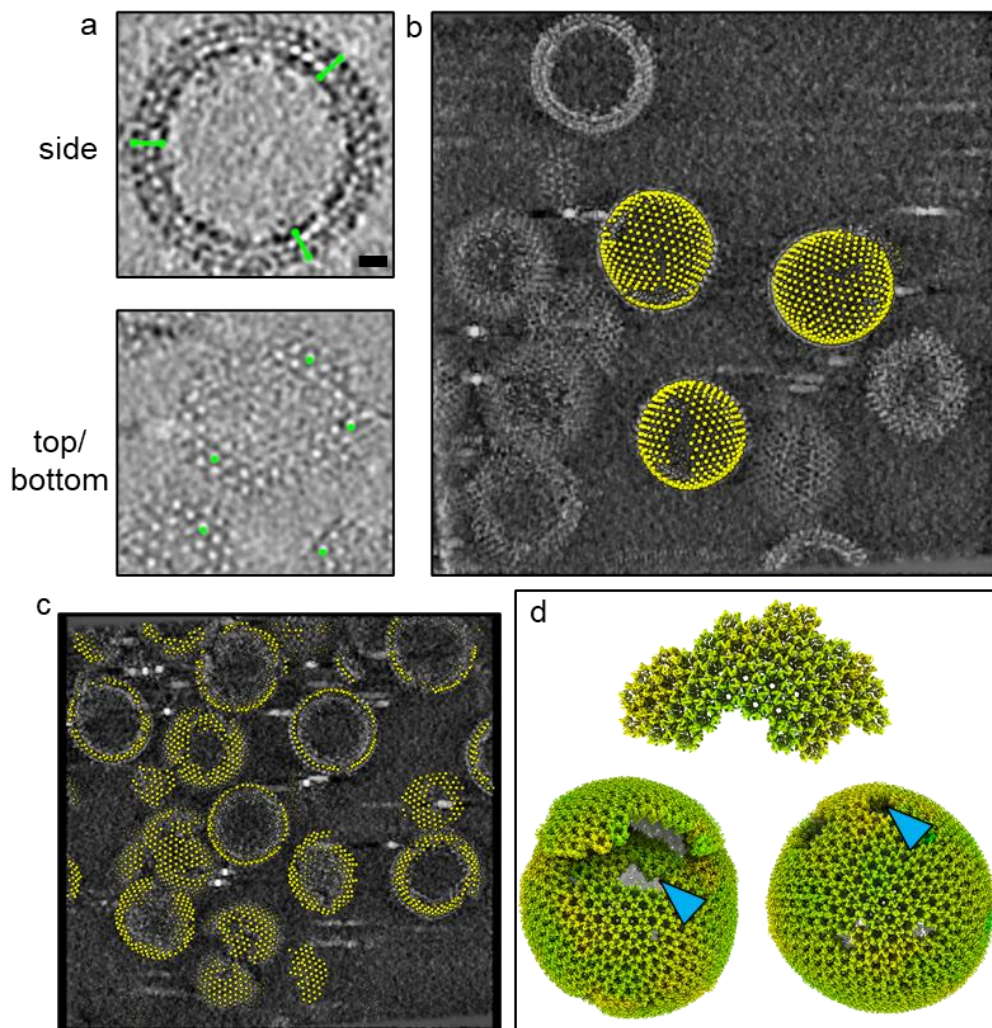


Figure 5- 4. TomoNet application on HIV VLPs.

a, Illustration of picked “seed” particles in different views. Green segments represent the direction of particles’ Y-axis. Scale bar is 20 nm. b, the “Auto Expansion” result on three VLPs within tomogram TS_01; yellow dots represent the center of the hexamer subunits. c, “AI AutoPicking” particle prediction results of tomogram TS_45 show it can pick particles on all lattices of different sizes and shapes. d, Visualization of three different variations of the HIV Gag lattices generated by placing back averaged structures, two exhibiting a spherical shape, and one presented as a fragment. The defects in the lattice are indicated by blue arrowheads.

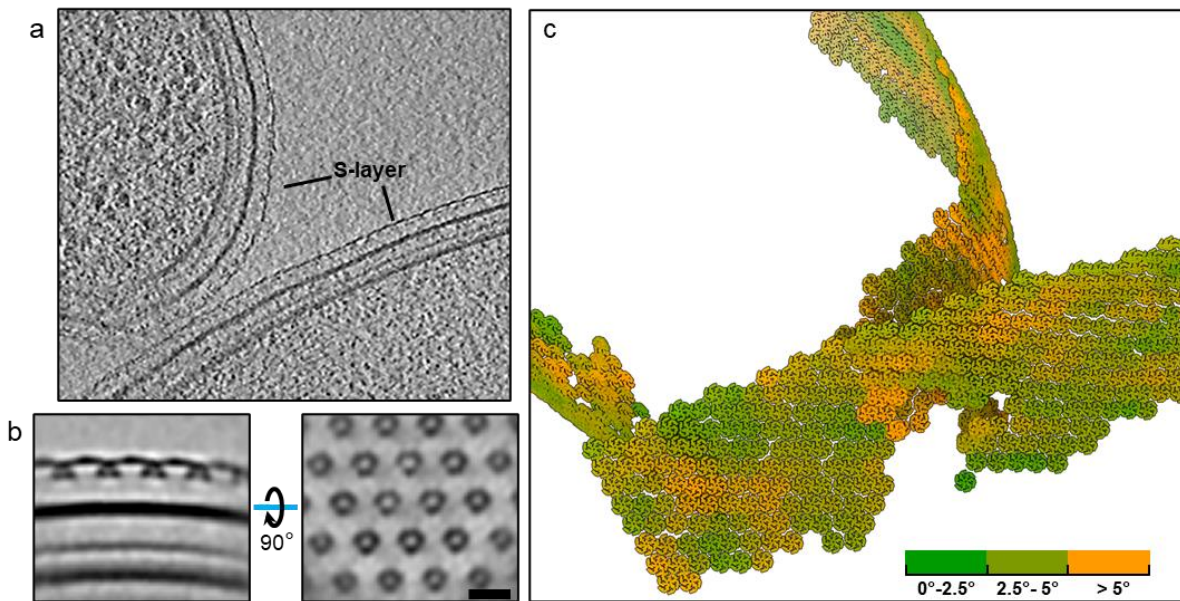


Figure 5- 5. TomoNet application on S-layer.

a, A tomographic slice view shows two *C. crescentus* cells in a FIB-milled lamella. b, Orthogonal slice views of the averaged density map generated in TomoNet, showing the hexagonal distribution of S-layer inner domains. Scale bar is 20 nm. c, Visualization of S-layer lattices generated by placing back hexamer subunit maps simulated from PDB: 6P5T. Coloring is based on surface curvatures at the point of each subunit.

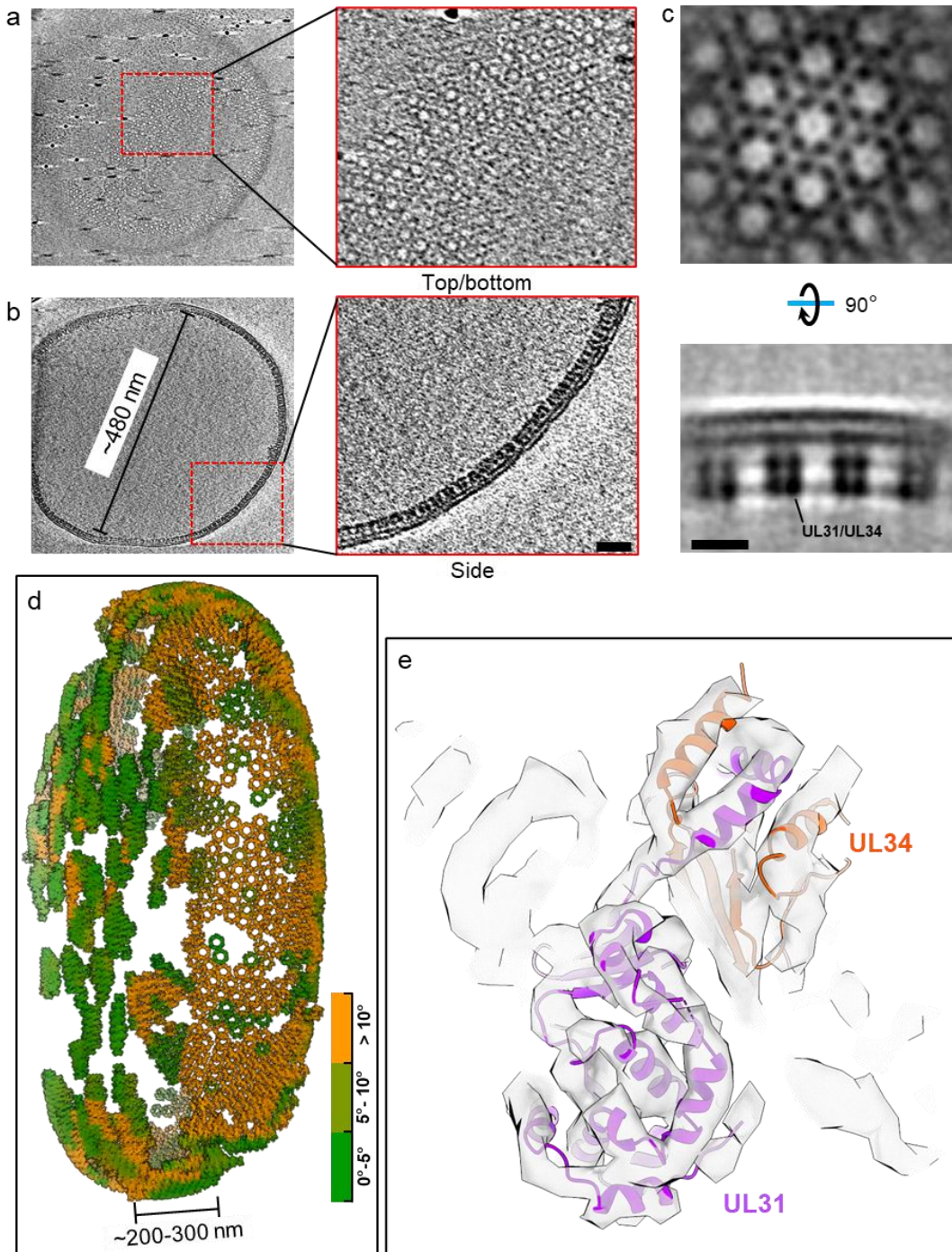
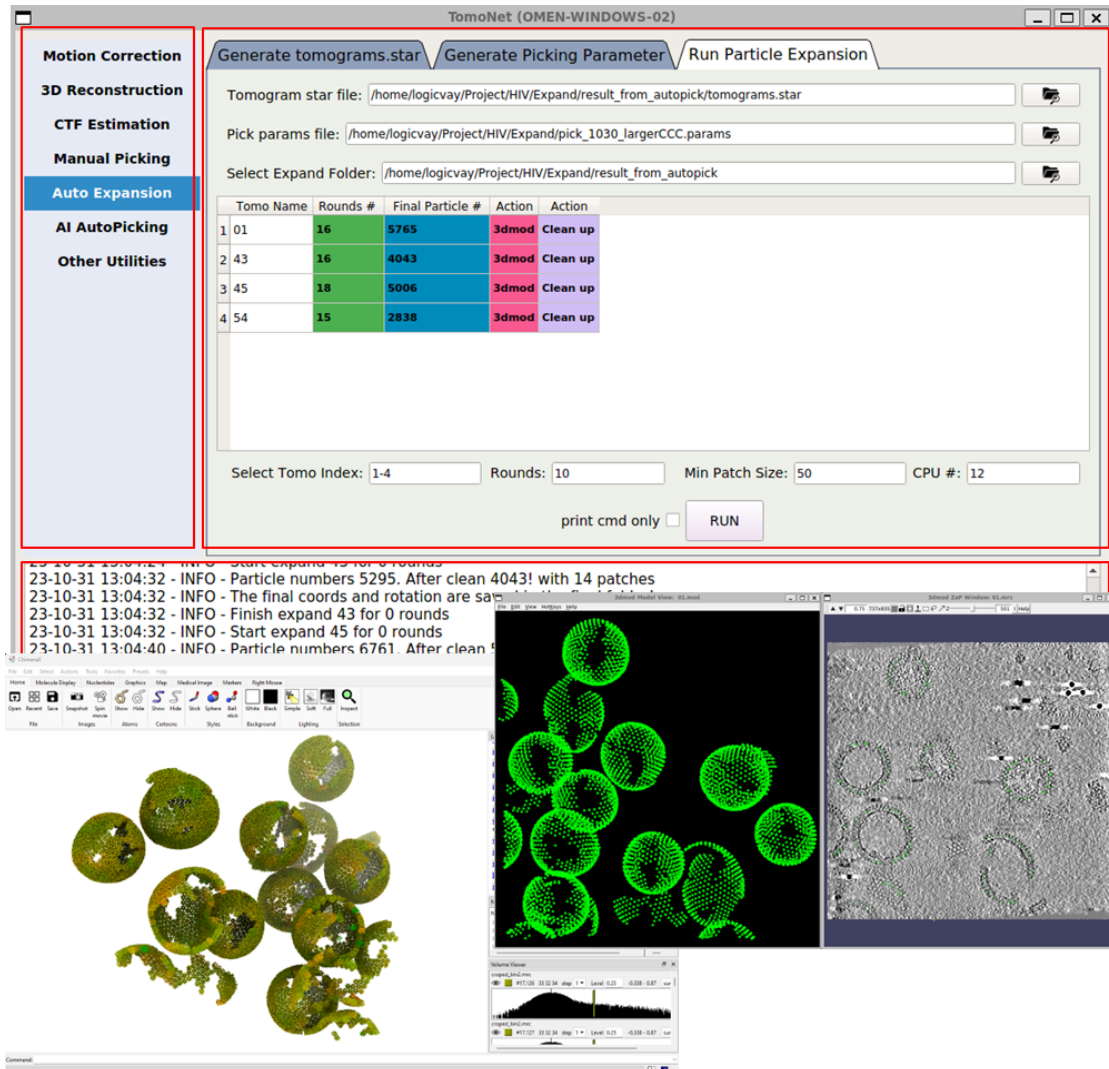


Figure 5- 6. TomoNet application on NEC lattice.

a, b, Tomographic slice views show a large NEC lattice; the insets show different views of NEC hexamer subunits. Scale bar is 20 nm. c, Orthogonal slice views of an averaged density map generated in TomoNet show that NEC hexamer subunits consist of UL31/UL34 heterodimers.

Scale bar is 10 nm. d, Visualization of an NEC lattice generated by placing back averaged maps shows that the large vesicle is compressed into a disk-like shape. The compression generated by sample freezing stretched the lattice and makes it appear flat and split at the top surface. Coloring is based on surface curvatures at the point of each subunit. e, Atomic model of the UL31/UL34 heterodimers fits into the final averaged map, with all helices well resolved.

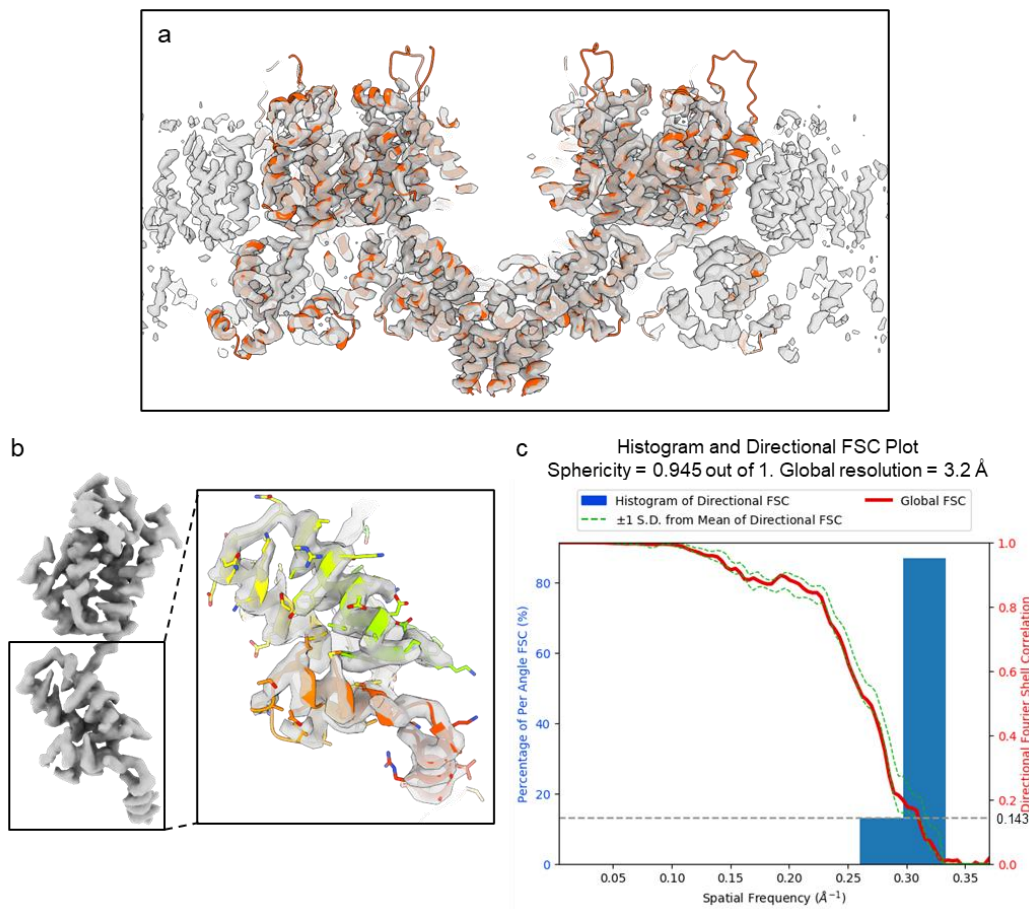
5.12 Supplemental Figures



Supplementary Figure 5- 1. A screenshot of the TomoNet GUI.

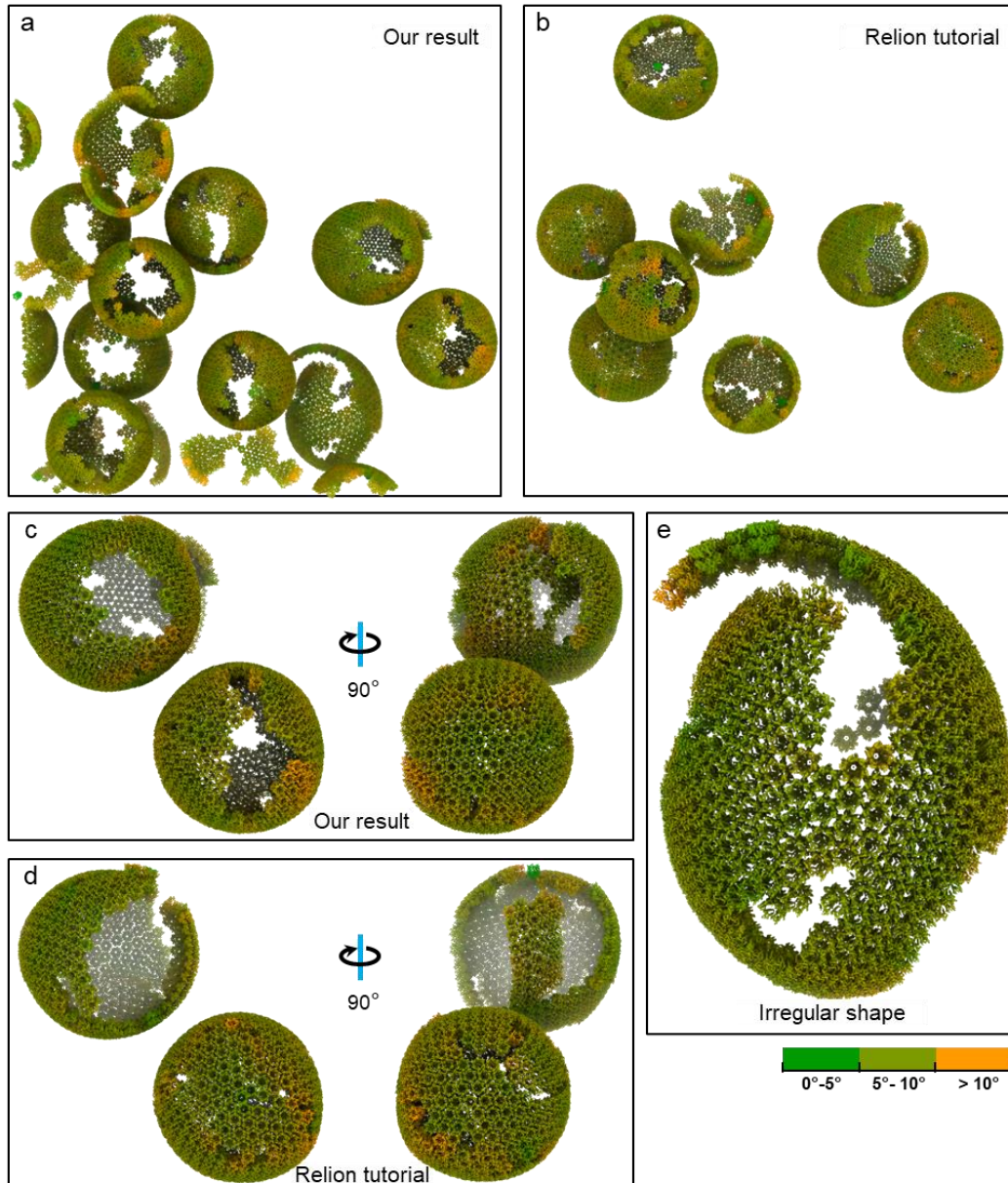
The TomoNet GUI contains three main areas: the menu bar (top left); the input and operate

area (top right); the log window (bottom). Results generated by “3D Subtomogram Place Back” function can be visualized in ChimeraX at the bottom left. An intermediate result of picked particles viewed with IMOD can be seen on the bottom right.



Supplementary Figure 5- 2. Final map resolution of the HIV Gag hexamer.

a, The final reconstruction of Gag hexamer (grey) fitted with atomic model (PDB: 5I93). **b**, One segmented Gag monomer structure, inset shows a closer view of carboxy-terminal domain overlay with the atomic model (PDB 5I93), colored in rainbow (blue to red) from N-terminal to C-terminal. **c**, Directional Fourier shell correlation (FSC) curves for the STA of Gag hexamer structure, with a global resolution at 3.2 Å.



Supplementary Figure 5- 3. Comparative visualization of the lattices obtained from TomoNet and Relion tutorial.

a, b, Visualization of a comparison of particles used in TomoNet and Relion tutorial within tomogram TS_01. TomoNet can pick particles not only on a sphere-like lattice but also on others with random shapes. **c, d**, A comparison of particle picking results on two sphere-like shape VLPs from TomoNet and Relion tutorial. **e**, A zoom-in view of an irregularly shaped lattice. Coloring is based on surface curvatures at the point of each subunit.

5.13 References

- 1 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169-4185, doi:10.1042/bcj20210708 (2021).
- 2 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
- 3 Wan, W. & Briggs, J. A. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol* **579**, 329-367, doi:10.1016/bs.mie.2016.04.014 (2016).
- 4 Chen, M. *et al.* A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat Methods* **16**, 1161-1168, doi:10.1038/s41592-019-0591-8 (2019).
- 5 Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr Opin Struct Biol* **58**, 249-258, doi:10.1016/j.sbi.2019.05.021 (2019).
- 6 Castano-Diez, D. & Zanetti, G. In situ structure determination by subtomogram averaging. *Curr Opin Struct Biol* **58**, 68-75, doi:10.1016/j.sbi.2019.05.011 (2019).
- 7 Hong, Y., Song, Y., Zhang, Z. & Li, S. Cryo-Electron Tomography: The Resolution Revolution and a Surge of In Situ Virological Discoveries. *Annu Rev Biophys* **52**, 339-360, doi:10.1146/annurev-biophys-092022-100958 (2023).
- 8 Huang, Y., Zhang, Y. & Ni, T. Towards in situ high-resolution imaging of viruses and macromolecular complexes using cryo-electron tomography. *J Struct Biol* **215**, 108000, doi:10.1016/j.jsb.2023.108000 (2023).

- 9 Sibert, B. S. *et al.* Workflow for High-resolution Sub-volume Averaging from Heterogenous Viral and Virus-like Assemblies. *Microsc Microanal* **29**, 943-944, doi:10.1093/micmic/ozad067.470 (2023).
- 10 Kopylov, M., Bobe, D., Johnston, J. D. & Paraan, R. M. Modern Tools for In-situ Tomography. *Microsc Microanal* **29**, 954-955, doi:10.1093/micmic/ozad067.476 (2023).
- 11 Ni, T. *et al.* High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity. *Nat Protoc* **17**, 421-444, doi:10.1038/s41596-021-00648-5 (2022).
- 12 Xue, L. *et al.* Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205-211, doi:10.1038/s41586-022-05255-2 (2022).
- 13 Zivanov, J. *et al.* A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *Elife* **11**, doi:10.7554/eLife.83724 (2022).
- 14 Obr, M. & Schur, F. K. M. in *Advances in Virus Research* Vol. 105 (ed Félix A. Rey) 117-159 (Academic Press, 2019).
- 15 Castano-Diez, D., Kudryashev, M., Arbeit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J Struct Biol* **178**, 139-151, doi:10.1016/j.jsb.2011.12.017 (2012).
- 16 Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218, doi:10.1038/s42003-019-0437-z (2019).
- 17 Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* **15**, 955-961, doi:10.1038/s41592-018-0167-z (2018).

- 18 Bharat, T. A. M. & Scheres, S. H. W. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols* **11**, 2054-2065, doi:10.1038/nprot.2016.124 (2016).
- 19 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).
- 20 Zhang, J. *et al.* Structure of the trypanosome paraflagellar rod and insights into non-planar motility of eukaryotic cells. *Cell Discov* **7**, 51, doi:10.1038/s41421-021-00281-2 (2021).
- 21 Imhof, S. *et al.* Cryo electron tomography with volta phase plate reveals novel structural foundations of the 96-nm axonemal repeat in the pathogen *Trypanosoma brucei*. *eLife* **8**, e52058, doi:10.7554/eLife.52058 (2019).
- 22 Si, Z. *et al.* Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLOS Pathogens* **14**, e1007452, doi:10.1371/journal.ppat.1007452 (2018).
- 23 Scaramuzza, S. & Castaño-Díez, D. Step-by-step guide to efficient subtomogram averaging of virus-like particles with Dynamo. *PLOS Biology* **19**, e3001318, doi:10.1371/journal.pbio.3001318 (2021).
- 24 Böhm, J. *et al.* Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences* **97**, 14245-14250, doi:doi:10.1073/pnas.230282097 (2000).
- 25 de Teresa-Trueba, I. *et al.* Convolutional networks for supervised mining of molecular patterns within cellular context. *Nature Methods* **20**, 284-294, doi:10.1038/s41592-022-01746-2 (2023).

- 26 Moebel, E. *et al.* Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods* **18**, 1386-1394, doi:10.1038/s41592-021-01275-4 (2021).
- 27 Wu, S., Liu, G. & Yang, G. in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. 1-5.
- 28 Hao, Y. *et al.* VP-Detector: A 3D multi-scale dense convolutional neural network for macromolecule localization and classification in cryo-electron tomograms. *Computer Methods and Programs in Biomedicine* **221**, 106871, doi:<https://doi.org/10.1016/j.cmpb.2022.106871> (2022).
- 29 Rice, G. *et al.* TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nature Methods* **20**, 871-880, doi:10.1038/s41592-023-01878-z (2023).
- 30 Balyschew, N. *et al.* Streamlined structure determination by cryo-electron tomography and subtomogram averaging using TomoBEAR. *Nature Communications* **14**, 6543, doi:10.1038/s41467-023-42085-w (2023).
- 31 Jimenez de la Morena, J. *et al.* ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation. *J Struct Biol* **214**, 107872, doi:10.1016/j.jsb.2022.107872 (2022).
- 32 Liu, H.-F. *et al.* nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods*, doi:10.1038/s41592-023-02045-0 (2023).

- 33 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
- 34 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76, doi:10.1006/jsbi.1996.0013 (1996).
- 35 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
- 36 Heumann, J. M., Hoenger, A. & Mastronarde, D. N. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *Journal of Structural Biology* **175**, 288-299, doi:<https://doi.org/10.1016/j.jsb.2011.05.011> (2011).
- 37 Hall, S. R. The STAR file: a new format for electronic data transfer and archiving. *Journal of Chemical Information and Computer Sciences* **31**, 326-333, doi:10.1021/ci00002a020 (1991).
- 38 Wang, H. *et al.* Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein. *Nature Communications* **14**, 6720, doi:10.1038/s41467-023-42368-2 (2023).
- 39 Schur, F. K. M. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508, doi:doi:10.1126/science.aaf9620 (2016).
- 40 von Kügelgen, A., Alva, V. & Bharat, T. A. M. Complete atomic structure of a native archaeal cell surface. *Cell Reports* **37**, 110052, doi:<https://doi.org/10.1016/j.celrep.2021.110052> (2021).

- 41 Pum, D., Breitwieser, A. & Sleytr, U. B. Patterns in Nature—S-Layer Lattices of Bacterial and Archaeal Cells. *Crystals* **11**, 869 (2021).
- 42 Liu, Y.-T. *et al.* Isotropic reconstruction for electron tomography with deep learning. *Nature Communications* **13**, 6482, doi:10.1038/s41467-022-33957-8 (2022).
- 43 Tan, A., Pak, A. J., Morado, D. R., Voth, G. A. & Briggs, J. A. G. Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proceedings of the National Academy of Sciences* **118**, e2020054118, doi:doi:10.1073/pnas.2020054118 (2021).
- 44 Guo, S., Saha, I., Saffarian, S. & Johnson, M. E. Structure of the HIV immature lattice allows for essential lattice remodeling within budded virions. *eLife* **12**, e84881, doi:10.7554/eLife.84881 (2023).
- 45 Talledge, N. *et al.* HIV-2 Immature Particle Morphology Provides Insights into Gag Lattice Stability and Virus Maturation. *Journal of Molecular Biology* **435**, 168143, doi:<https://doi.org/10.1016/j.jmb.2023.168143> (2023).
- 46 Lasker, K. *et al.* The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nature Communications* **13**, 5643, doi:10.1038/s41467-022-33221-z (2022).
- 47 Draganova, E. B. *et al.* The universal suppressor mutation in the HSV-1 nuclear egress complex restores membrane budding defects by stabilizing the oligomeric lattice. *bioRxiv*, 2023.2006.2022.546118, doi:10.1101/2023.06.22.546118 (2023).
- 48 Bouvette, J. *et al.* Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. *Nat Commun* **12**, 1957, doi:10.1038/s41467-021-22251-8 (2021).

- 49 Eisenstein, F. *et al.* Parallel cryo electron tomography on in situ lamellae. *Nature Methods* **20**, 131-138, doi:10.1038/s41592-022-01690-1 (2023).
- 50 Khavnekar, S. *et al.* Multishot tomography for high-resolution in situ subtomogram averaging. *J Struct Biol* **215**, 107911, doi:10.1016/j.jsb.2022.107911 (2023).
- 51 Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *Journal of Structural Biology* **199**, 187-195, doi:<https://doi.org/10.1016/j.jsb.2017.07.007> (2017).
- 52 Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods* **18**, 186-193, doi:10.1038/s41592-020-01054-7 (2021).
- 53 Meng, E. C. *et al.* UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein Sci*, e4792, doi:10.1002/pro.4792 (2023).
- 54 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).
- 55 Mendonça, L. *et al.* CryoET structures of immature HIV Gag reveal six-helix bundle. *Communications Biology* **4**, 481, doi:10.1038/s42003-021-01999-1 (2021).
- 56 Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol Rev* **38**, 823-864, doi:10.1111/1574-6976.12063 (2014).
- 57 Herrmann, J. *et al.* A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proceedings of the National Academy of Sciences* **117**, 388-394, doi:doi:10.1073/pnas.1909798116 (2020).

Chapter 6: Hierarchical organization and assembly of the archaeal cell sheath from an amyloid-like protein

Hui Wang^{1,2,3}, Jiayan Zhang^{2,3}, Daniel Toso^{1,2,3}, Shiqing Liao^{2,3}, Farzaneh Sedighian³, Robert Gunsalus^{3,4} and Z. Hong Zhou^{1,2,3,*}

¹Department of Bioengineering, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

²California NanoSystems Institute, UCLA, Los Angeles, CA 90095, USA

³Department of Microbiology, Immunology, and Molecular Genetics, UCLA, Los Angeles, CA 90095, USA

⁴The UCLA-DOE Institute, UCLA, CA 90095, USA

***Address for Correspondence:** Z. Hong Zhou (Hong.Zhou@UCLA.edu, 1-310-694-7527)

6.1 Abstract

Certain archaeal cells possess external proteinaceous sheath, whose structure and organization are both unknown. By cellular cryogenic electron tomography (cryoET), here we have determined sheath organization of the prototypical archaeon, *Methanospirillum hungatei*. Fitting of AlphaFold-predicted model of SH monomer into the 7.9 Å-resolution structure reveals that the sheath cylinder consists of axially stacked β -hoops, each of which is comprised of two to six 400 nm-diameter rings of β -strand arches (β -rings). With both similarities to and differences from amyloid cross- β fibril architecture, each β -ring contains two giant β -sheets composed of ~450 SH monomers that entirely encircle the outer circumference of the cell. Tomograms of immature cells suggest models of sheath biogenesis: oligomerization of SH monomers into β -ring precursors after their membrane-proximal cytoplasmic synthesis, followed by translocation through the unplugged end of a dividing cell, and insertion of nascent β -hoops into the immature sheath cylinder at the junction of two daughter cells.

6.2 Introduction

Essential to life emergence, the barriers separating the interior of a cell from the exterior differ significantly among the three domains of life. While all cells possess lipid bilayer that enclose their cytoplasm, some types specialize by assembling additional exterior layers that confer shape and size. For example, plant cells possess a cellulose based outer layer and most bacteria have peptidoglycan layer(s). Some bacteria and most archaea have an S-layer. A distinct group of microorganisms possess an additional outer layer, termed the sheath, with the best studied member of this group being the methanogenic archaeal species, *Methanospirillum hungatei*. First isolated from a municipal sewage sludge digester, it operates in combination with fermentative and syntrophic bacteria to recycle complex organic matter into methane, water, and carbon dioxide¹. It provides the driving force for carbon decomposition in anaerobic food chains and thus performs an essential role in the carbon cycling on earth², *M. hungatei*, together with other specialized methanogens, possess an unusual cylindrical cell morphology^{3,4}. Its cylindrical shape is determined by an outermost proteinaceous sheath layer that ensures a uniform cell dimension of 0.4-0.5 μm in diameter^{3,5}, and encapsulates from 1 to over 70 cells within, forming a slightly wavy filament up to 500 μm in length^{1,6}. Within the sheath tube, individual cells are enclosed by a proteinaceous surface layer (S-layer) outside their lipid membrane and further separated by plug-like structures. The polar tufts of flagella and pili at the two ends of the sheath tube allow for cell motility and taxis^{1,4,5}. As a hydrogenotrophic methanogen, *M. hungatei* is both a consumer and producer of gases, resulting in differential internal pressures relative to the outside environment. Its cylindrical shape enhances gas exchange, given that a representative *M. hungatei* cell that is ~ 7 μm in length⁷ and 0.4 μm in diameter, its sheath can double the surface area compared to that of a sphere shape with same volume. Since the first observations of these cells nearly half a century ago^{1,5}, negative-stain transmission electron microscopy (TEM) has defined the basic cell architecture of *M. hungatei*,

both for the whole cell and cell fraction^{5,8}, as well as the organization of the plug, S-layer lattices and sheath⁹⁻¹². The sheath is an extremely stable structure and maintains cylindrical shape of the cell under common denaturants¹³. Early attempts to isolate the sheath protein (SH) under strongly reducing conditions first yielded flattened crystalline arrays⁸ and subsequently multiple hoops^{14,15}, suggesting that the subunits of the cylindrical sheath layer do not follow a helical arrangement. Recent results from a combination of conventional TEM, mass spectrometry, and sequence analysis establish that SH is the 40.6 kDa protein WP_011449234.1 previously annotated from the *M. hungatei* genome and suggested to share amyloid properties¹⁶.

However, molecular descriptions of the sheath are lacking. Indeed, *M. hungatei* sheath's diverse, and seemingly unrelated characteristics (single protein composition, large cylindrical shape, resistance to high pressure¹⁷, hydrogen-methane gas exchange etc.) raise more questions than answers. What is the structure of the sheath protein? How does it assemble and maintain the characteristic cylindrical shape of the cell to permit gas exchange and direct cell growth? How are the newly synthesized SH translocated through multiple barriers to assemble into well-organized polymers outside the cell? Here, by cellular cryogenic electron tomography (cryoET) with subtomogram averaging and modeling with SH monomer from AlphaFold¹⁸, we establish the molecular organization and biogenesis of the cylindrical sheath layer of *M. hungatei*. About 450 SH monomers circumferentially polymerized around the cylindrical cell. They form a ring structure (termed as β -ring), which consists of two enormous β -sheets with characteristics similar to the so-called " β -arch kernel" coined for amyloids¹⁹. β -hoops, each comprising 2 to 6 β -rings, stack axially into a cylindrical sheath extending to hundreds of micrometers long. In addition, visualization of SH structures with an immature cell allows us to propose a route for nascent β -hoop synthesis, assembly, and insertion into a growing sheath, answering the question of how proteins are transported across multiple barriers to assemble into the largest known assembly from a single protein—the external sheath of archaeal cells.

6.3 Results

6.3.1 Organization of the *M. hungatei* sheath by cellular cryoET

To examine *M. hungatei* cells and resolve its sheath structure at higher resolution and without dehydration and fixation artifacts, we captured *in situ* cell images by cellular cryoET (Fig. 1). Lower magnification tomograms reveal the overall cylindrical shape of the sheath layer with ~400 nm diameter, as well as organelle distribution within the cell (Figs. 1a, b). Cross-section views of higher magnification tomograms (Fig. 1c) show that the sheath is composed of a series of hoops, which we refer to as β -hoops, to reflect their composition of β -sheet-rich SH (see below). Thousands of β -hoops stack axially to form a tube-like structure that encases cells, whose topology and “crystalline” arrangement have been examined at low resolution by cryoET³ and negative-stain TEM^{14,15}, respectively.

Our high-resolution tomograms now show that the sheath actually is a non-periodic stack of β -hoops with variable numbers of SH rings (β -rings), thus deviating from the previously thought crystalline arrangement (Fig. 1c). The cross section of each β -hoop has a hand-like appearance (Fig. 1c insets) with average “hand” length at around 9 nm. However, the number of “fingers” of the “hand” varies from 2 to 6, which results in the width variation of β -hoops (Fig. 1c and Supplementary Fig. 1a), thus the previous TEM-based interpretation of sheath being crystalline¹⁴ is actually incorrect, explaining the failure of the early efforts, including our own, to resolve the sheath structure based on crystallinity. Each of these “fingers” is the cross-section view of a β -ring. The predominant β -hoop type is the one containing 4 β -rings, while β -hoops containing 3 or 5 β -rings occur less frequently, and those with 2 or 6 β -rings rarely (Fig. 1d and Supplementary Fig. 1a; termed 2-, and 6- β -ring hoops, respectively). As indicated by the arrangement of different membered β -hoops, no long-range periodicity is observed along the axial direction. In addition, this pattern at the opposite side of the cell is not always maintained, indicating that β -rings in adjacent hoops can switch memberships (Supplementary Fig. 2).

To further improve resolution and investigate the building block of the sheath layer, we classified subtomograms into different membered β -hoops. By averaging β -hoops with the same number of β -rings, we obtained subtomogram averages and resolved the 4- β -ring hoop at 7.9 Å global resolution (anisotropically varying from 7 to 20 Å at different directions) (Supplementary Figs. 1b, c and Supplementary Movie 1). Placing this subtomogram average map back into the original tomogram allowed us to generate a large region of the sheath, containing 3 full β -hoops and 2 half β -hoops (hoops colored differently in Fig. 1e and Supplementary Movie 2), which furthers structural understanding of the sheath within its *in situ* environment. The structural feature appears to repeat every 28 Å circumferentially along the β -ring, consistent with previous observation¹⁴ (Fig. 1f, repeating units are demarcated by dashed line of the same color). 3D segmentation was conducted to visualize this repeating unit (Fig. 1g), which matches the AlphaFold-predicted model of an SH monomer (Fig. 1h). This SH monomer model can be divided into three domains: a cap domain containing six α helices and one β sheet, an amyloid-like domain containing two six-stranded large β sheets, named β -sheet 1 and β -sheet 2, and a linker projecting from the joint between the cap and amyloid-like domain (Figs. 1g, h). A portion of a β -ring was modeled by fitting the AlphaFold-predicted monomer model into the cryoET density map, where the interior parallel β -sheets circumferentially extend, and the cap domains repeat every 28 Å (Fig. 1i and Supplementary Movie 2). This SH-SH spacing is identical to that revealed in the TEM image of a sheath component (Figure 6 in Southam et al.¹²), suggesting the previous isolated sheath component are a β -ring segment. The subtomogram average also reveals that β -rings are assembled into a β -hoop through both β -sheet to β -sheet and linker to β -sheet interactions, and that neighboring β -hoops are connected through cap-cap interactions (Fig. 1j). Within a β -hoop, β -sheets interact with neighbors closely while caps do not, consequently the interior of the β -hoop, formed by the bottoms of the β -sheets, is narrower than the exterior of the β -hoop, formed by the caps; in other words, the axial cross-section of the β -hoop exhibits an annulus sector shape (top panel of Fig. 1j).

6.3.2 SH contains a motif similar to the β -arch kernel found in amyloid fibrils

The remarkable size of the β -sheets in the *M. hungatei* sheath (Fig. 1i) prompted us to compare the β -ring structures with those in various amyloid cross- β fibrils¹⁹⁻²³. β -sheets 1 and 2 are separated by 10 Å (Figs. 2a, b), consistent with previous X-ray diffraction and FTIR spectrum observations of archaeal sheaths^{14,16,24}, which also suggested that the sheath has a high content of structural elements characteristic of amyloid cross- β structure^{20,23,25,26}. Recurring in amyloid fibrils is the β -arch motif, in which two β -strands of a continuous polypeptide chain interact via their amino-acid side chains, forming an arch^{27,28} that dictates fibril formation, as observed in α -synuclein fibrils^{19,21} (Figs. 2c, d). Strands within each β -sheet of the *M. hungatei* SH monomer are arranged in an anti-parallel fashion (Fig. 2b), and β -strands across β -sheets 1 and 2 are connected through β -arches (Fig. 2b). Thus, each SH has 6 β -arches with different side chain compositions unlike the identical β -arches within a pathological amyloid fibril. In addition, the two β -strands within a β -arch in SH are 15° tilted from each other instead of being parallel as in disease-related β amyloids (Fig. 2c). The SH structure allows side chain interactions between different β -arches instead of forming the typical steric zipper structure in pathological amyloid fibrils. For example, α -synuclein¹⁹ and the Alphafold-predicted structure of bacterial curli major subunit csgA protein²⁹. These unusual β -arch properties indicate that SH can acquire stability through inter-sheet side chain interactions possibly through hydrogen bonds and salt bridges (Fig. 2d and Supplementary Fig. 3), and by forming hydrophobic cores with aromatic amino acids (Fig. 2d).

The cylindrical shape and consistent dimension of the sheath tube raise two significant questions: how could the β -arches in SH form β -rings in archaeal cell sheath instead of β -solenoids in amyloid fibrils, and what determines the radius of the β -ring? Interestingly, at the interface between two adjacent SHs in a β -ring, the two amino acids connecting β -strands S11 to S12 are not part of the β -sheet (Fig. 2e), creating a disruption to β -sheet 2. As a result,

instead of participating in backbone interactions as in a typical β -sheet, residues K363 and D43 from adjacent SH subunits could potentially interact through a salt bridge between their side chains (Fig. 2e inset). Also, cap-cap interactions between adjacent SHs likely cause steric hindrance at the opposite side of this salt bridge between K363 and D43 (Figs. 1i and 2e). These two structural features can lead to the shorter interior rim distance than the exterior rim distance, causing a slight curvature around 0.8° . The long-range consequence of this curvature is that ~450 SH subunits assemble into a self-limiting circular β -ring (Fig. 1e and Supplementary Movie 2), in contrast, the α -synuclein forms a non-terminating, continuous amyloid β fibril.

Thus, one β -ring contains a “ring-amyloid” structure comprising ~2700 β -strands for each of its two giant β -sheets, a remarkable β -sheet size only second to that seen in various amyloid cross- β fibrils. Axially, each micrometer of *M. hungatei* sheath structure contains ~89 β -hoops, corresponding to ~356 β -rings or ~160,200 SH subunits. Accordingly, the sheath of a 500 μm multi-cellular filament contains about 44,500 β -hoops, 178,000 β -rings, or 80,100,000 SH subunits, representing the largest known protein polymer in a living cell.

6.3.3 SH interactions within a β -hoop and between adjacent β -hoops

Docking AlphaFold-predicted SH monomer model into the 4- β -ring hoop subtomogram average map reveals how the β -hoops are held together to form a cylindrical sheath layer (Fig. 3a). In the cap domain, a 23 Å-long helix in the AlphaFold-predicted model matches well with a helix resolved in the density map (Fig. 3b); this, together with the matched β -sheets 1 and 2 (Fig. 3a inset), indicates that the atomic model was docked without ambiguity in orientation. Densities were observed between β -sheets 1 and 2, corresponding to a cluster of aromatic residues between the two β -sheets in the AlphaFold model (circled in Fig. 3a inset and Fig. 3c). These extensive connections between β -sheets 1 and 2, together with potential salt-bridges between charge pairs (aspartic acid-lysine and glutamic acid-lysine) (Supplementary Fig. 3), stabilize the amyloid-like β -ring structure. The docking also reveals possible sites of glycosylation⁴, where

amino acids Thr86, Ser144 and Asn157 in the cap domain are exposed to the exterior of the cell (Fig. 3b).

The β -ring arrangement in a β -hoop suggests the base loops (Fig. 3c) in the β -sheet domain may serve multiple functions. First, the two extending loops at the left and right bottom of SH monomer could form extra connections and establish the binding point when β -rings start assembling into a β -hoop (Fig. 3c and Supplementary Figs. 4a-b). Second, the asymmetric arrangement leaves a free charged loop on both the left and right side of a β -hoop (Supplementary Figs. 4a-b). Coincidentally, extra densities attached to the sheath interior at the same place are only observed near the plug region (blue arrowheads in Fig. 1c insets), suggesting that the exposed base loops could serve as binding sites for proteins that regulate plug formation and/or cell division.

The aforementioned linker connections between adjacent β -rings were analyzed by examining the spatial relationship between two adjacent SH monomers to localize the previous speculated disulfide bond^{11,15,30} (Figs. 3d, e). To illustrate the shape, organization and inter-SH interactions, we built a 4- β -ring hoop segment (Fig. 3f, Supplementary Figs. 4a, 5 and Supplementary Movie 3) by placing the above β -ring segment models into the tomogram and following the β -ring stacking arrangement within a β -hoop visible in the TEM image of flattened sheath layer²⁴. The model of this hoop segment reveals that each SH monomer contains two cysteine residues⁷ on opposite sides, one on β -sheet 1 and the other on the linker. These cysteine residues from neighboring SH subunits both within a β -hoop (Fig. 3d) and between two β -hoops (Fig. 3e) are only 8.3 Å and 13 Å apart, respectively. Although the distances between β -rings inside a β -hoop (i.e., intra-hoop β -rings) is longer than that of a typical disulfide bond, Cys257 is at the tip of the linker loop, whose flexibility should allow formation of a disulfide bond between the two cysteine residues. When two β -ring models are aligned by this putative disulfide bond, the complementary electrostatic potential distribution at their interfaces further supports the above β -hoop assembly pattern (Supplementary Fig. 5b). Indeed, there is clear

density in the linker region between intra-hoop β -rings (arrowhead in Fig. 3g) but such density is missing between inter-hoop β -rings. Thus, we propose that disulfide bonds connect SH monomers between intra-hoop β -rings, providing strong interactions to impart structural rigidity for the β -hoop.

Beyond rigidity, the existence of different β -hoop types raises another vital question: how the sheath establishes axial linearity leading to the cylindrical shape of *M. hungatei* cells (Figs. 1c inset and 3a)? The variation of numbers of β -ring per β -hoop results in different sizes of the axial cross-sectional annulus sectors among these β -hoops, e.g., a 5- β -ring hoop has a longer annulus sector arc than a 3- β -ring or 4- β -ring hoop (Fig. 3g). These different sizes would seem to create varied contact points between different β -hoop types, which could affect the molecular bonds between β -hoops. Therefore, we compared the subtomogram averages of 4- β -ring, 5- β -ring, and 3- β -ring hoop to investigate the local subunit conformational changes and inter- β -hoop interfaces (Fig. 3g).

Adjacent β -rings contact each other not only at the base of the β -sheet domain, but also in the middle through the linker, probably involving disulfide bonds (Figs. 3d, g). As shown in Figure 3g, the only contact point between neighboring β -hoops is at the cap domains of their outermost β -rings and remains the same; but the angle formed between them (*i.e.*, the angle formed by red and pink dashed lines in Fig. 3g) is larger for β -hoops containing more β -rings.

Within each β -hoop, the first β -ring (the leftmost of the 4- β -ring hoop in Fig. 3g) is perpendicular to the axis of the cylinder, and the subsequent β -rings are increasingly tilted to the right while the cap domains within these β -rings are increasingly less tilted as compared to their β -sheets domains (see the relationship between the yellow and red dashed line in Fig. 3g middle panel), suggesting a flexible connection between the cap and the β -sheets domain. Such flexibility would compensate for differences in the angles between adjacent β -hoops to maintain consistent molecular contact between contacting SH subunits, also allow all β -hoops to align linearly along the axial direction of the cylindrical cell.

6.3.4 *In situ* structure of an immature cell

By imaging actively dividing *M. hungatei* cells, we sought to understand how newly translated SH proteins are transported across the membrane and protein layers to assemble into such well-organized polymers outside the cells. The *M. hungatei* replicates through dividing a cell into two daughter cells such that the sheath layer encapsulating multiple cells elongates concomitantly with the cell growth. From one of our tomograms ([Supplementary Movie 4](#)), we observed an immature cell measured around 480 nm long ([Fig. 4a](#)), much shorter than that of the mature cell ([Fig. 1a](#)). Its nascent plug component is tilted and not fully sealed near the sheath layer ([Fig. 4a](#)), unlike that in the mature cell^{3,5,12}. Further, between the S-layer and the cell membrane is a pocket that is absent in the mature cell ([Fig. 4a](#)); the space within the pocket harbors densities which, as detailed below, we interpret as nascent SH polymers assembled immediately after SH synthesis and membrane translocation. These features are consistent with an immature cell.

To better observe the organization of this immature cell, a 3D rendering of the tomogram was generated ([Supplementary Movie 5](#)). Several filamentous densities with similar thickness around 10 nm were observed both within the pocket, adjoining the cell membrane and the S-layer ([Figs. 4b, c](#)), and in the interstitial space between the plugs of the two adjacent cells ([Fig. 4d](#)). Within the pocket, the filamentous densities have several contact points with the cell membrane, and the S-layer appears to be discontinuous near the pocket and there is a gap between the newly forming plug layers and the sheath ([Figs. 4b, c](#)); in the interstitial space, the filamentous densities have a curvature that is similar to curvature of the β -hoops, though these densities are not as aligned as the β -hoops in the sheath ([Fig. 4e](#)). Additionally, close to these densities are unfilled lens-like openings between adjacent sheath layers, where β -hoops are partially separated ([Fig. 4f](#)); and such an opening has never been observed in normal mature cells before. Thus, we propose that those filamentous densities are the precursors of β -rings in

different stages: the pocket ones, given their proximity to the membrane, are likely newly synthesized SH polymers; and the interstitial ones, given their location next to the opening in the sheath layer, are partially aligned β -rings in the process of assembling into β -hoops that could be subsequently inserted into the opening, driving sheath elongation and cell growth. In the absence of direct evidence establishing the identity of these densities, it is also possible that these filamentous densities are other extracellular polymers, such as flagella, with diameter around 10nm.

6.4 DISCUSSION

By providing a barrier against entropy flow as governed by the second law of thermodynamics, the compartmentalization of biological molecules and functions is synonymous with life emergence on earth. Membrane targeting through ribosome-bound Sec61/SecYEG complex enables delivery of trans-membrane proteins to serve as conduits across the membrane, typical of the situation for eukaryotic cells³¹. In bacteria, some of these trans-membrane proteins act as cell-wall synthesis enzymes, and others as lipid transporters for the assembly of outer membrane of Gram-negative bacterial cells. Highly ordered protein polymers typically are assembled at the location of protein synthesis within the confines of the cytoplasm and regulated by various protein or non-protein factors when a dynamical process between assembly and disassembly is necessary. A prominent example of a highly ordered protein assembly is the protein capsid of non-enveloped viruses, which typically manifest mostly as spherical assemblies with icosahedral symmetry³²⁻³⁷ and occasionally as filamentous assemblies with helical symmetry³⁸. Such complexes assemble within their host cytoplasmic site immediately after protein translation and subsequently translocate across cell membrane as fully assembled virions either by exocytosis of cells without cell walls or by lysis of cells from within cell walls (as in bacteria and eukaryotic plant cells). Distinctive from eukaryotic and bacterial cells, the existence of external, highly ordered proteinaceous layers in archaeal cells

poses additional challenges to ensure ordered assembly and timely delivery of such structures from inside of the cell to the outside.

Our study shows that the outmost layer of *M. hungatei* cell is composed of a single small protein of just 40 kDa with an amyloid-like cross- β structure, thus forming a functional amyloid polymer. Other microbial functional amyloids are known, including the extracellular fibrillar structures described in both Gram negative and positive bacteria with roles in biofilm formation and adhesion, for example the curli, Bap and Esp orthologs³⁹, P1 adhesions, Harpins and modulins (see review⁴⁰). Many structural and functional aspects of their biological roles remain to be fully elucidated. How does the *M. hungatei* SH amyloid-like molecule assemble into a biologically functional cylinder? The presence of additional loops at the base of its β -arch plus the “linker” loop, both absent in pathogenic amyloids, appear to direct the assembly of the β -ring and β -hoop. The subsequent axial stacking of many multi-ring β -hoops, each of which contains thousands of amyloid-like β -arches yields in essence, a gigantic supramolecular structure composed of ~13,456,800 β -strands for a typical mature *M. hungatei* cell of ~7 μ m in length. Axially, each micrometer of cylinder contains ~89 β -hoops, corresponding to ~356 β -rings or ~160,200 SH subunits. Thus, a 500 μ m multi-celled filament contains about 44,500 β -hoops, 178,000 β -ring. While both evolutionarily successful, the use of a single small protein to encase and protect the cell in *M. hungatei* is a genetically simpler solution than bacterial peptidoglycan-containing cell envelope, which requires numerous genes, complex synthesis-assembly pathway and associated regulatory system to accomplish.

Though much remains to be clarified and the temporary opening in the sheath needs to be verified, the visualization of the “growing” sheath in dividing cells provides some clues of archaeal sheath assembly. One provisional model is a four-stage model of sheath biogenesis as illustrated in [Figure 5](#). SH monomers are synthesized by polyribosomes near the inner surface of the cell membrane and SH signal peptide ([Fig. 2a](#)) at the N-terminus directs secretion of these monomers through a Sec61/SecYEG translocon⁴¹ or a secretion system encoded by one

of the four sets of type II/IV secretion genes within the *M. hungatei* genome⁷. Cleavage of the signal peptide triggers SH folding into a cross- β rich structure, which oligomerizes into short β -ring segments through β -sheet augmentation (Fig. 2e). These short β -ring segments could pass through S-layer through the pocket region, then elongate near the incomplete nascent plug. Upon arriving at the interstitial space between two daughter cells, the newly assembled short β -ring segments stack to form pre- β -hoops and inserts into a temporary opening in the sheath layer between two cells, elongating the sheath to provide more space for cell growth (Fig. 5a, b). These steps are based on the observations in tomograms of immature cells (Fig. 4, Supplementary Movies 4 and 6). Notably, our model of sheath synthesis, assembly and translocation is consistent with the distinctive property of many archaeal viruses—they fashion a highly ordered proteinaceous capsid (likely derived from archaeal cells⁴²) external to a bilayer membrane without transmembrane viral proteins. Nonetheless, due to the intrinsic lack of temporal progression information in the frozen hydrated sample used in cryoET, this proposed sheath biogenesis should only be considered as a putative model to be tested by orthogonal methods such as time-resolved fluorescence light microscopy.

As contemplated by quantum physicist Erwin Schrodinger in his 1944 book entitled “What is Life?”, crystalline arrangement of simple molecules provides the means to overcome increasing entropy in order for life to emerge from chaos. We now know that perfect crystalline arrangement, including three-dimensional forms as in table salt and in one-dimensional form as in amyloid β -fibrils, is not a solution to the many functional and dynamic processes of living organisms. Rather, some minor deviation from perfection might just be what is necessary to drive these processes. Archaea are considered as extremophiles that are related to Earth's primitive atmosphere and typically exist in extreme habitats. Found in sewage sludge and as a model organism of the Archaea, *M. hungatei* has been used for examining syntrophic symbiosis which drives anaerobic carbon recycling in nature. Our results might offer insight into how *M. hungatei* polymerizes the small SH into the sheath with small deviations from crystallinity to

accomplish functional tasks. *M. hungatei* sheath has the extraordinary capability of resisting ~300 atmospheric pressure¹⁷, yet should allow exchange of methane, hydrogen and water to support basic intracellular functions. Notably, the trunk of buoyancy-control gas vesicles within bacterial and archaeal cells is also a thin-walled cylinder with distinctive hydrophilic exterior and hydrophobic interior assembled from a single β -hairpin-containing protein^{43,44}, thus differs from the archaeal sheath which fashions alternating hydrophobic/hydrophilic patches (Supplementary Fig. 6). In archaeal sheath, the intra-hoop SH interactions are extensive, explaining its capability to resist high pressure. The interior of each β -hoop has no holes and thus is well sealed by its amyloid-like β -sheet domains (Fig. 1e and j), therefore the wedge-shaped cleft between β -hoops, which varies in shape and measures up to ~30° in angle and 40 Å in width (Fig. 1j), appears to be the only outlet for waste gas. Many questions still remain concerning the details of biogenesis, biophysical properties and regulatory mechanisms of sheath. Nonetheless, from the grandeur perspective, SH's deviation from a "perfect" β -sheet (Fig. 2e) and crystalline packing contributes to the cylindrical shape of sheath, thus its multiple functional roles: determining the cell size and cylindrical shape, serving as an environmental barrier against harmful agents, creating a novel periplasmic space bounded by the SH and S-layer, and providing a rigid cell platform for taxis-directed flagellar movement. Thus, while Mendel and Morgan's discoveries of genetics provided the answer for variation missing in Darwin's theory to drive evolution of life, the current work illustrates an example on how a small 40 kDa protein introduces variation to perfect periodicity/crystallinity to impart functions of life processes.

6.5 Materials and Methods

6.5.1 *M. hungatei* cell growth and sample preparation

M. hungatei strain JF1 (ATCC 27890) was cultured anaerobically in 28 ml anaerobic tubes with a vessel headspace pressurized to 10 psi with an 80:20 (vol/vol) mixture of H₂:CO₂⁴. Each 1000 ml of medium contained the following: 0.54 g NaCl, 0.12 g MgSO₄·7H₂O, 5.0 g NH₄Cl, 1.8 g

KH₂PO₄, 2.9 g K₂HPO₄, 0.06 g CaCl₂·2H₂O, 2.72 g Na Acetate ·3H₂O, 10 mL of 100x trace metal solution and 1 mL of 1000x vitamin solution. Following sterilization, the medium was supplemented with 10 ml filter-sterilized solution of reducing reagent (2.5% Na₂S· 9H₂O, 2.5% Cysteine HCl) and 20 ml of a 1 M NaHCO₃ solution. Following inoculation, anaerobic tubes containing 10 ml of medium were incubated at 37 °C horizontally on a rotary drum shaker (60 RPM, New Brunswick, Inc). Cells were serially transferred at least three times, with transfers made at mid-exponential phase, to achieve 10+ cell doublings prior to harvest. To capture dividing cells, cells were harvested early in the exponential growth phase by centrifuging 1 mL of cell suspension at low speed (5,000x g) and resuspending in 100 µl of medium over a three-day period, with 24 hours between each harvest.

6.5.2 CryoET and tomogram reconstruction

M. hungatei samples were mixed with 5 nm-diameter fiducial gold beads. 3 µl of the mixture was applied onto Quantifoil (3:1) holey carbon grids that were freshly glow-discharged for 30 s at -40 mA. With an FEI Mark IV Vitrobot cryo-sample plunger, excess sample on the grid was blotted away with filter paper at a blot force of -4 and blot time of 5 s. The sample was vitrified immediately by being plunged into liquid nitrogen-cooled liquid ethane. Plunge-freezing conditions and cell concentration on the grids were optimized with an FEI T20 transmission electron microscope equipped with an Eagle 2K HS CCD camera. Grids with vitrified cells were stored in a liquid nitrogen dewar until use.

With either FEI Batch Tomography (for lower Mag. Data collection) or *SerialEM*⁴⁵ (for higher Mag. Data collection), tilt series were collected in a Titan Krios instrument equipped with a Gatan imaging filter (GIF) and a post-GIF K2 direct electron detector in electron-counting mode at California NanoSystems Institute (CNSI); the data collection parameters are listed in [Table 1](#). Frames in each movie of the raw tilt series were aligned, drift-corrected, and averaged with *Motioncor2*⁴⁶. The tilt series micrographs were aligned and reconstructed into 3D

tomograms using the *IMOD* software package⁴⁷.

6.5.3 Subtomogram averaging

Subtomogram averaging was performed with *PEET*^{48,49} and *Relion4*⁵⁰. *In situ* details of the sheath layer and its assembly pattern along the longitudinal axis were visible in our tomograms. However, because of the preferred orientation of the *M. hungatei* cells (Supplementary Fig. 1b) and the missing wedge effect, the SH monomer repeat unit along the outer circumference was not distinguishable. Learned from both the previous study that described the diffraction patterns of the isolated β -hoops¹⁴, and the dimension of AlphaFold-predicted model, the SH monomer repeat unit along the outer circumference was presumed to be 3 nm for the later particle picking process.

Initial particles, *i.e.*, sub-tomograms, of the 3- β -ring, 4- β -ring and 5- β -ring hoops were manually picked, one per β -hoop, at the z-coordinate where they were most visible. The picked particles were averaged with *PEET*; 68, 465 and 85 particles were used to generate the initial reference maps for the 3- β -ring, 4- β -ring and 5- β -ring hoop, respectively. The geometry shown in the initial reference map, combined with the proposed 3 nm repeat unit of SH monomers, guided us to sample more particles circumferentially along the β -hoops to expand the particle set. *PEET* was used to align the final expanded particle set, resulting in one *PEET* motive list file per tomogram (*i.e.*, *MOTL.csv file), containing translational and angular alignment information.

Then, we first use *createAlignedModel* to generate new *PEET* motive list files to re-center all particles, and later use *MOTL2Relion* to convert the above angular alignment results stored in the motive list file into *Relion*'s convention. Such that the coordinates and orientations of the particles were formatted and imported into *Relion4* for further refinement. Following one round of 3D refine under 4-binned pixel size and two rounds of 3D refine under 2-binned or original pixel size, along with the removal of particles within 2.5 nm from each other to prevent

particle duplication, the resulting resolutions for the 3- β -ring, 4- β -ring and 5- β -ring hoop were 21.8 Å, 7.9 Å and 14.0 Å (Supplementary Figs. 1c-e), respectively. Resolution was calculated on 3DFSC Processing Server⁵¹ and the global resolution reported above is based on the “gold standard” refinement procedures and the 0.143 Fourier shell correlation (FSC) criterion.

6.5.4 Modeling and 3D visualization

The AlphaFold models for the SH monomer (Fig. 1h) and csgA monomer (Fig. 2d) were generated with AlphaFold2 Google Colab¹⁸. Tomograms displayed in Figure 1a, b and Figure 4a were missing-wedge corrected by *IsoNet*⁶². Visualization of tomograms and averaged electron density maps were done with *IMOD* and *UCSF ChimeraX*⁶³, respectively.

6.6 Author contributions

RG and ZHZ initialized and supervised research; FS, JZ, HW and RG prepared samples; DT, JZ and HW carried out cryoET imaging and structure determination; SL helped data processing and figure preparation; ZHZ, RG, HW, JZ and SL interpreted the data and wrote the manuscript; all authors reviewed and approved the paper.

6.7 Acknowledgements

We thank Zhu Si for initial structural determination efforts and Titania Nguyen for editorial assistance. This project is supported by grants from the US National Institutes of Health (NIH) (GM071940 to Z.H.Z.), US National Science Foundation (1515843 and 1911781 to R.P.G.) and U.S. Department of Energy (DOE) Office of Science (BER) (contract DE-FC-02-02ER63421 to R.P.G.). We acknowledge use of resources in the Electron Imaging Center for Nanomachines supported by UCLA and grants from the NIH (1S10OD018111 to Z.H.Z.) and the National Science Foundation (DBI-1338135 and DMR-1548924 to Z.H.Z.).

6.8 Data availability

The subtomogram average structure data generated during the current study have been

deposited in the Electron Microscopy Data Bank (EMDB) repository, with the accession codes EMD-29442 [<https://www.ebi.ac.uk/emdb/EMD-29442>] (4- β -ring hoop), EMD-29443 [<https://www.ebi.ac.uk/emdb/EMD-29443>] (3- β -ring hoop), and EMD-29448 [<https://www.ebi.ac.uk/emdb/EMD-29448>] (5- β -ring hoop). The previously published structure of α -synuclein shown in [Figure 2d](#) is available in the Protein Data Bank (PDB) repository under accession code 7LC9 [<https://www.rcsb.org/structure/7LC9>]. The major curlin subunit predicted atomic model shown in [Figure 2d](#) is available in AlphaFoldDB with accession code AF-P28307-F1 [<https://alphafold.ebi.ac.uk/entry/P28307>]. The atomic model of *Bacillus megaterium* gas vesicle segment shown in [Supplementary Figure 6](#) is available in PDB with accession code 7R1C [<https://www.rcsb.org/structure/7R1C>]. Source data are provided with this paper.

6.9 Figures

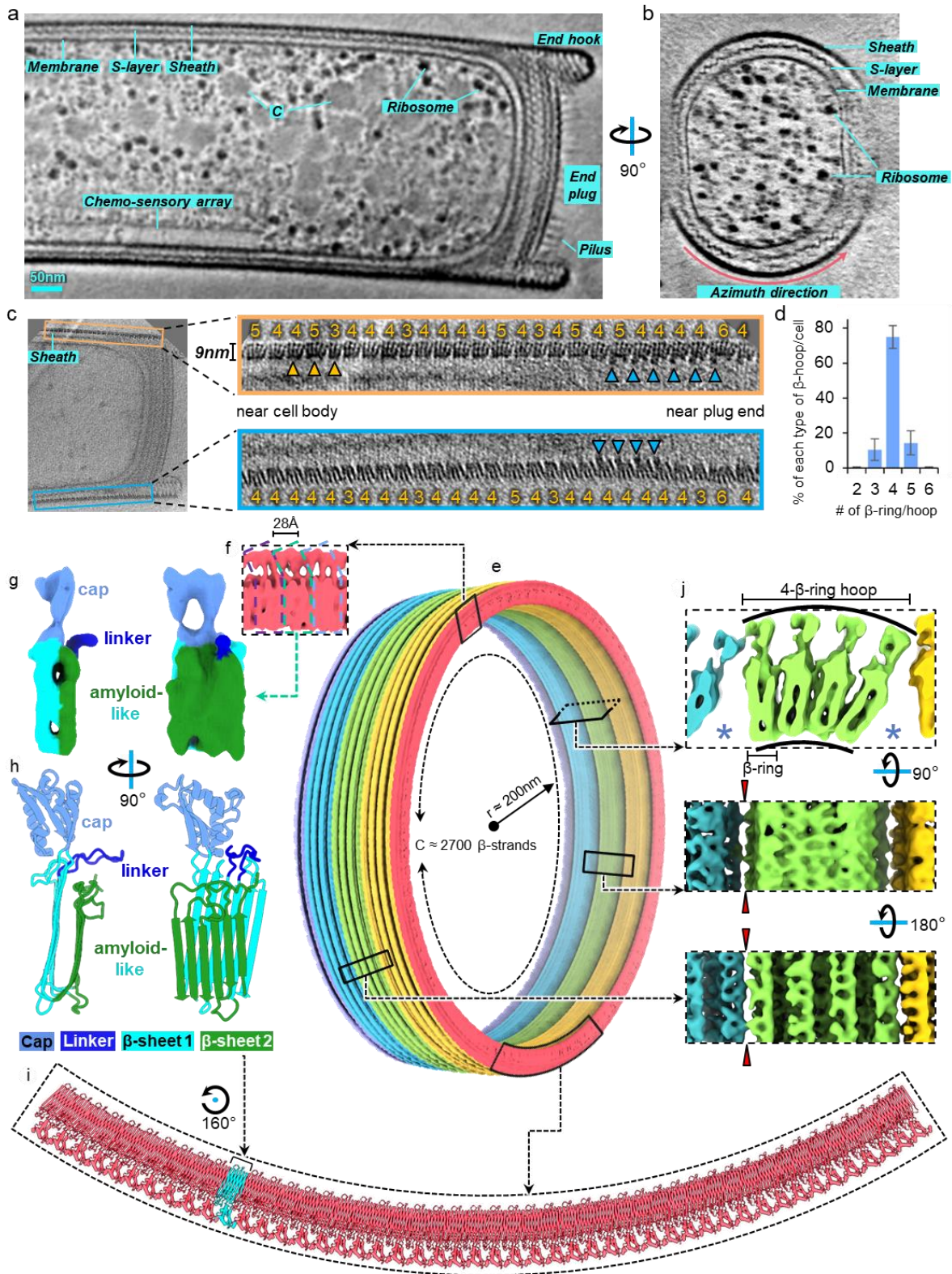


Figure 6-1. In situ structure of *M. hungatei* sheath layer.

a, b, Two orthogonal cross-sections of a representative cryoET tomogram at lower

magnification, showing the overall shape, size, distribution of the sheath. Intra-cellular structures, such as ribosomes, chemo-sensory array, and membrane-less condensate (C), can be recognized inside the cell. In total, 23 tomograms were collected in this session. **c**, Cross-section of a higher magnification tomogram of the region near the plug at one end of a cell. β -hoops (yellow arrowheads) with 3, 4, 5, and 6 β -rings can be identified in the zoom-in views of the sheath layer and are labeled with 3, 4, 5 and 6, respectively. Blue arrowheads point to extra densities attached to the β -hoops near the plug. **d**, Relative frequency histogram of different types of β -hoops in each cell, overlaid by the corresponding data points. The histogram data are presented as mean values over $n = 10$ tomograms examined in total. β -hoops with 2 or 6 β -rings are very rare (less than 1%). **e**, Model of a partial sheath layer containing 3 full and 2 half β -hoops colored differently, with circumference (C) equals to the length of around 2700 β -strands and radius (r) around 200 nm. The boxes mark the regions and viewing angles of the structures detailed in **f**, **i**, and **j**. **f**, Subtomogram average showing the structure that repeats every 28 Å along the outer circumference. **g**, **h**, CryoET density map (**g**) and AlphaFold model (**h**) of the SH monomer. **i**, Atomic model of a partial β -ring with one SH monomer highlighted in cyan. **j**, Three orthogonal (cross-section, interior and exterior) views of the subtomogram average of sheath layer with clefts between adjacent β -hoops and 'holes' (indicated by red arrowheads) at the inter- β -hoop interface. Source data are provided as a Source Data file for **d**. Scale bar = 50 nm in **a**, **b**, **c**.

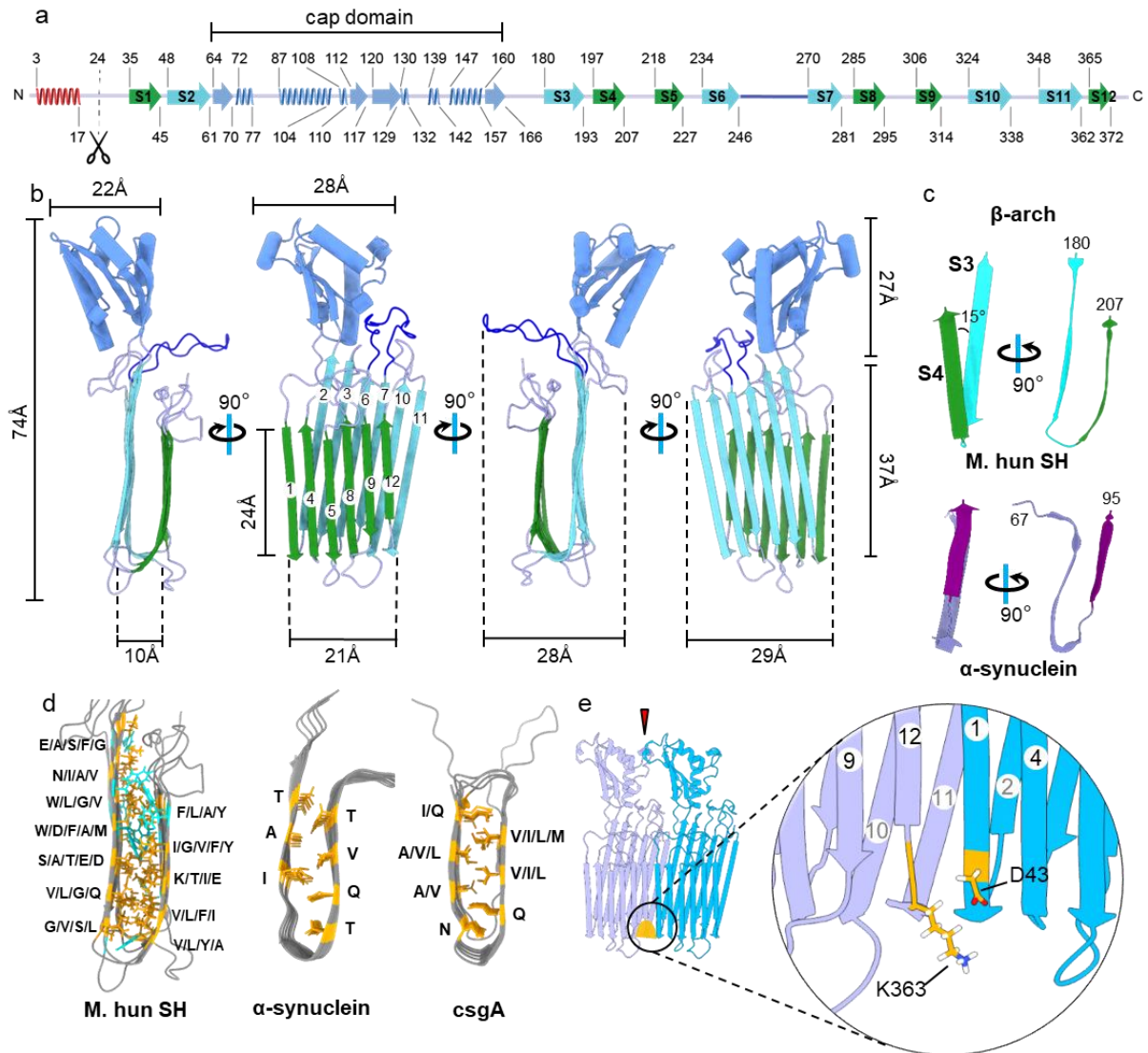


Figure 6- 2. Amyloid properties of SH monomer.

a, b, Structure of the SH monomer secondary structures displayed either along its amino-acid sequence (**a**) or as ribbon diagram in 3D (**b**). Different structural elements are colored consistently between (**a**) and (**b**), and the truncated N-terminal signal peptide (red) is shown only in (**a**). **c,** Comparison of the β -arch in SH and in α -synuclein (PDB: 7LC9). **d,** Comparison of side chains between parallel β -sheets among *M. hungatei* SH, α -synuclein, and curli major subunit *csgA*. Aromatic amino acids are colored in cyan and others in orange. **e,** Atomic model of two adjacent SH monomers along the outer circumference, with putative cap-cap interaction pointed by arrowhead. Inset shows the disruption on β -sheet 2 (yellow) at the interface between

K363 and D43 of the two neighboring monomers.

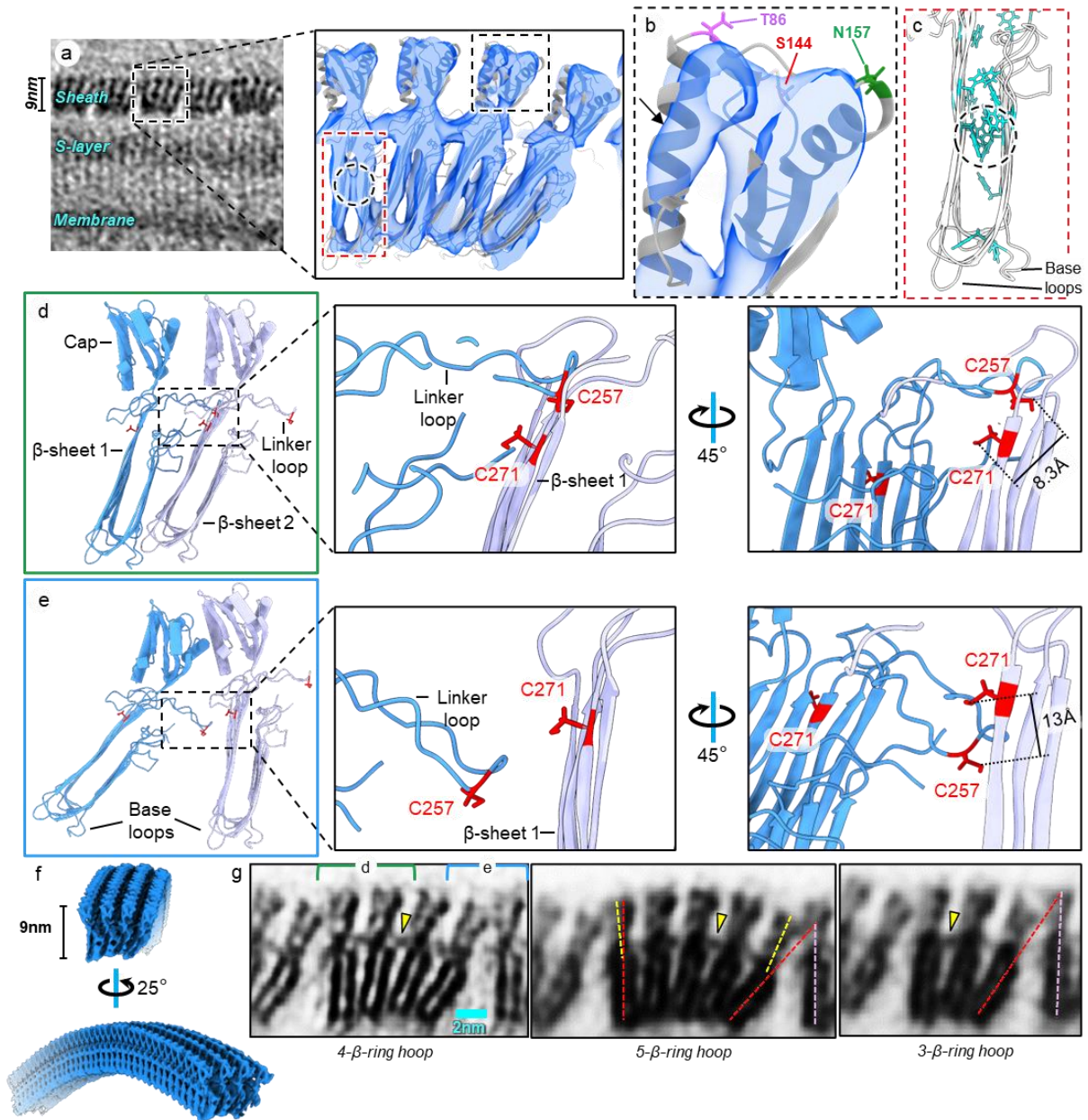


Figure 6- 3. Atomic model of sheath β -hoop.

a, Typical distribution of *M. hungatei* cell envelope and zoom-in subtomogram average of the 4- β -ring hoop docked with predicted SH monomer models. The black and red boxes mark the regions of the structures enlarged in panels **b**, **c**, respectively. **b**, Enlargement of the cap region in **a**, with its longest helix, measured ~ 23 Å in length, indicated by an arrow. **c**, All aromatic residues between β -sheet 1 and 2 are colored in cyan, the circled region indicates a cluster of

aromatic residues, whose location overlaps with those densities in the subtomogram average that connects two β -sheets in the middle. **d, e**, Interfaces of two neighboring SH subunits along cell growth axis either within a β -hoop (**d**) or between two β -hoops (**e**), their locations in the sheath are indicated in **g**. The insets show the putative disulfide bonds connecting adjacent β -rings. **f**, Model of a segment of 4- β -ring hoop, filtered to 10Å resolution and shown as shaded surface. **g**, Subtomogram averages of 4-, 3-, and 5- β -ring hoops, with the intra-hoop SH connection indicated by yellow arrowheads. The angles formed by red and pink dashed lines show the variation near the β -hoop interfaces caused by the number of β -rings in a β -hoop; the angles formed by yellow and red dashed lines suggest flexible connection between the cap and the β -sheets domain.

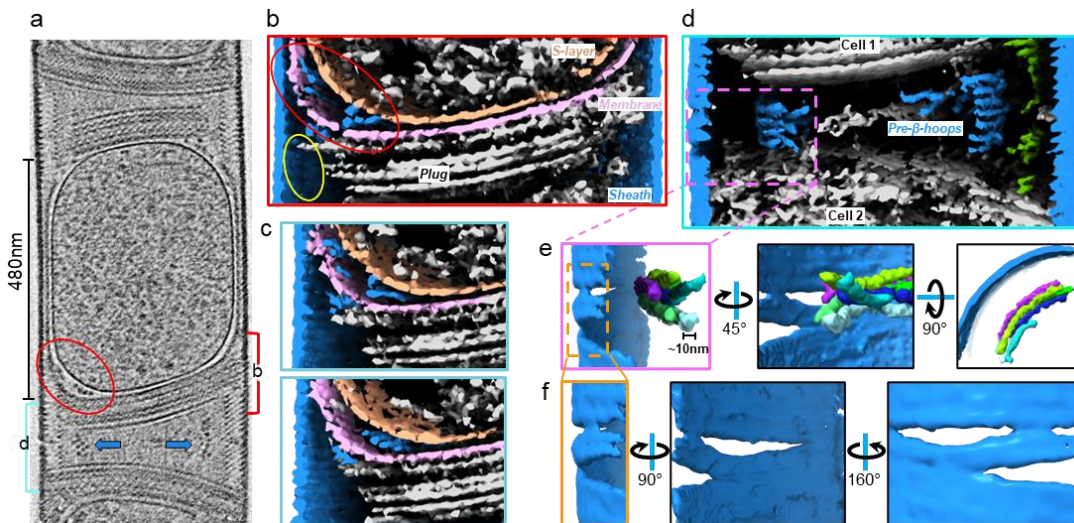


Figure 6- 4. In situ structure of an immature cell.

a, Tomogram of an immature *M. hungatei* cell with a cell length around 480nm, tilted incomplete plugs, pocket area (red ellipse) located between the membrane and S-layer. Filamentous densities (indicated by blue arrows) both appear inside the pocket area and between plugs from neighboring cells. The upper and lower boundaries of segmented structure displayed in **b** and **d** are indicated by red and cyan brackets, respectively. In total, 37 tomograms were collected in this session. **b-f**, Cross-section views at different z of 3D-rendered map generated from the

tomogram in **a**, showing filamentous or putative pre- β -hoop densities (color in blue) either inside the pocket near the plug breakage (yellow ellipse) on the left (**b**, **c**), or in the interstitial space between plugs (**d**). Putative pre- β -hoop densities with similar curvature as outer sheath are colored individually (**e**), and their location are close to the temporary sheath opening region (**f**).

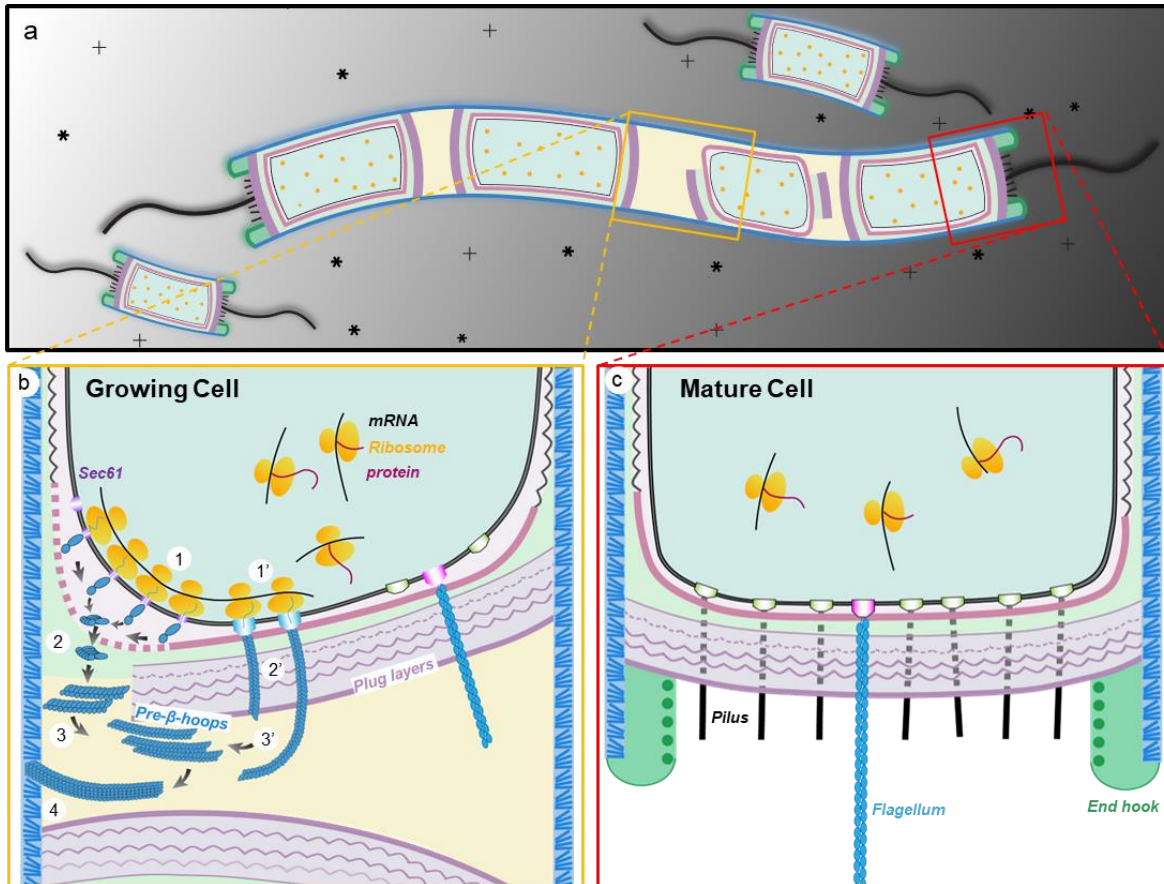
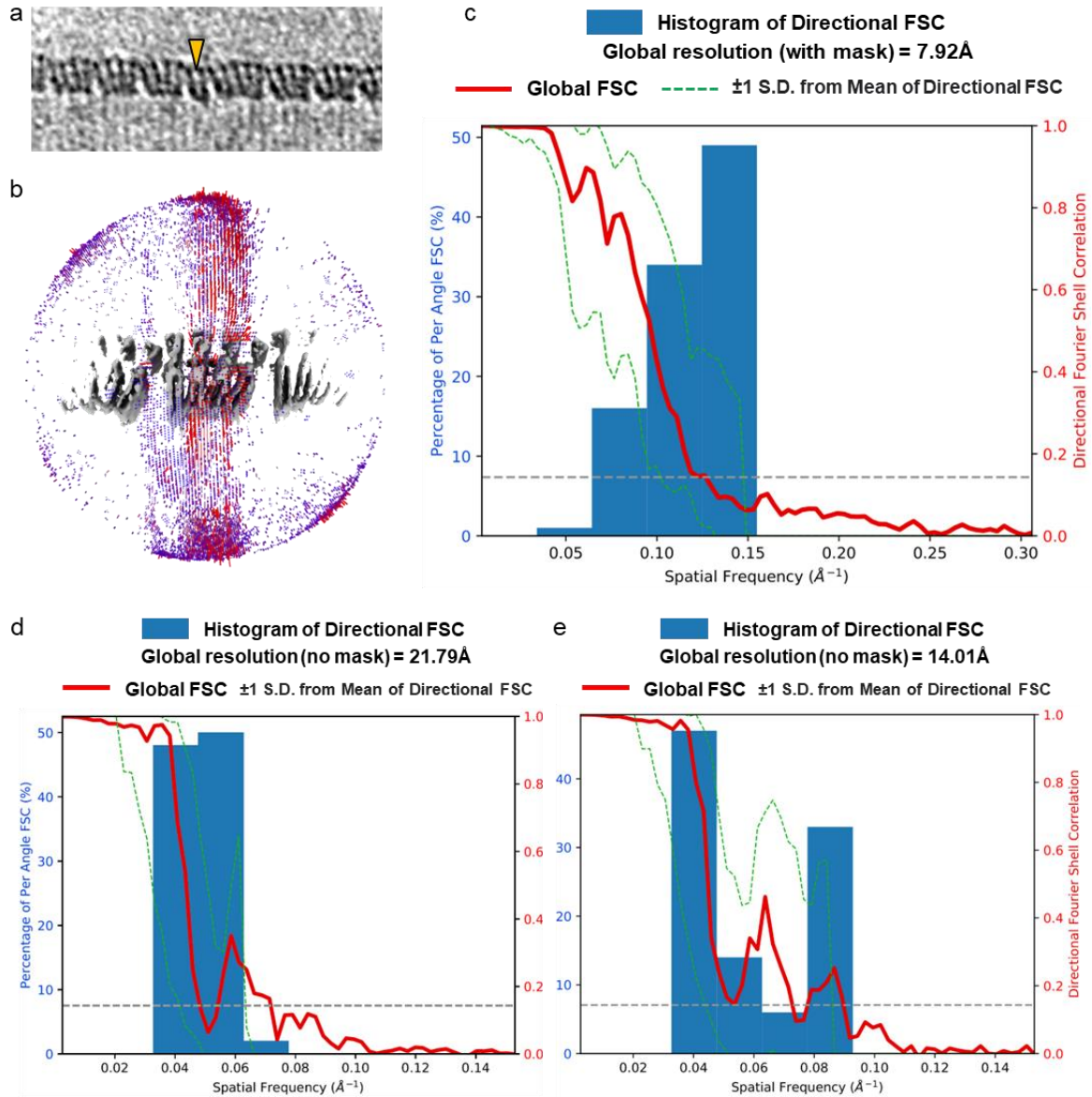


Figure 6- 5. Illustration of proposed sheath biogenesis of *M. hungatei* cell.

a, Schematic illustration of the living environment of *M. hungatei*. The cell filament in the middle contains 1 immature and 3 mature cells. **b**, Proposed model of sheath biogenesis depicting four stages: **1**, SH monomer production and pre- β -hoop oligomerization; **2**, pre- β -hoop translocation through either the S-layer breakage or secretion systems on the plug into the pocket and interstitial spaces between plugs; **3**, pre- β -hoop assembly into β -hoops; and **4**, β -hoop insertion into the sheath layer. **c**, Organization of the plug region of a mature end. An alternative

interpretation of the filamentous densities is the archaeal flagella, as indicated in **b** and **c**, with approximate the same diameter of a 4- β -ring hoop (~10 nm).

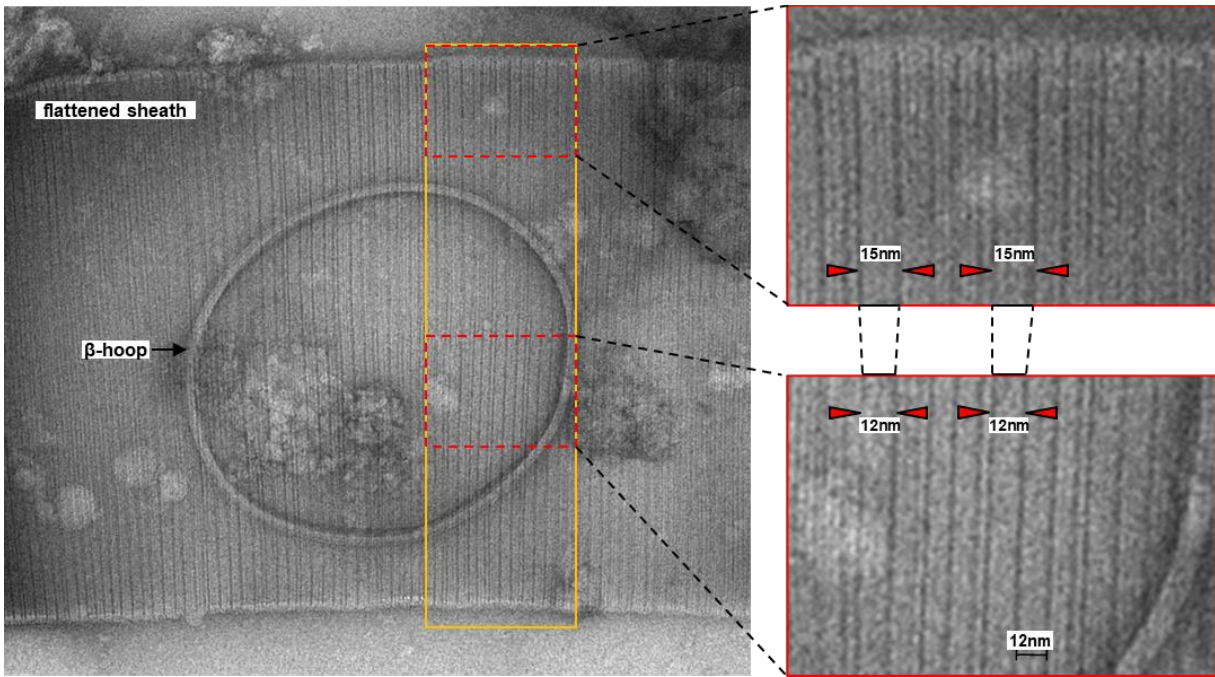
6.10 Supplemental Figures



Supplemental Figure 6- 8. Raw cryoET tomogram and quality evaluation for subtomogram averages.

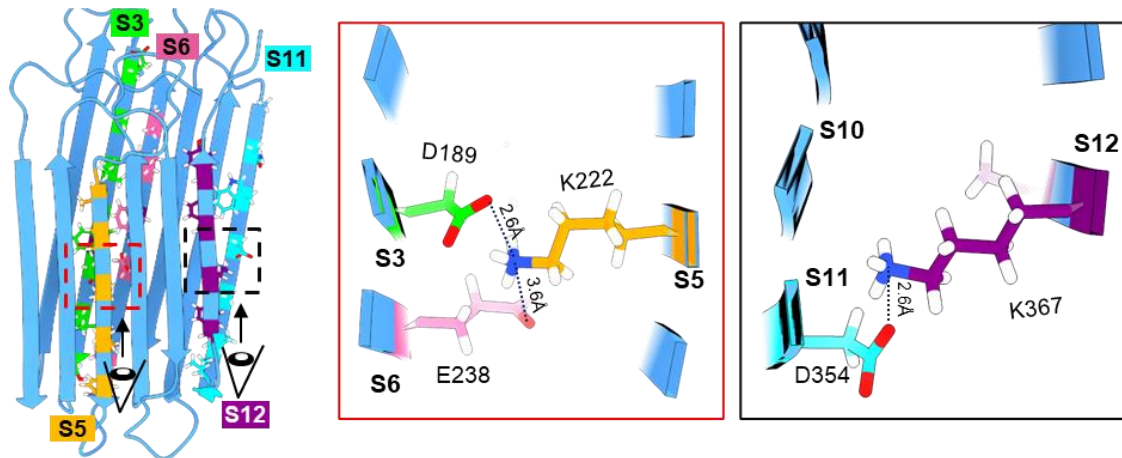
a, A density slice of cryoET tomogram with a 2- β -ring hoop indicated by the arrowhead. **b**, Particle orientation distribution. **c-e**, Directional Fourier shell correlation (FSC) curves for the

subtomogram averages of 4- β -ring hoop (c), 3- β -ring hoop (d), 5- β -ring hoop (e). Scale bar = 10nm in a.



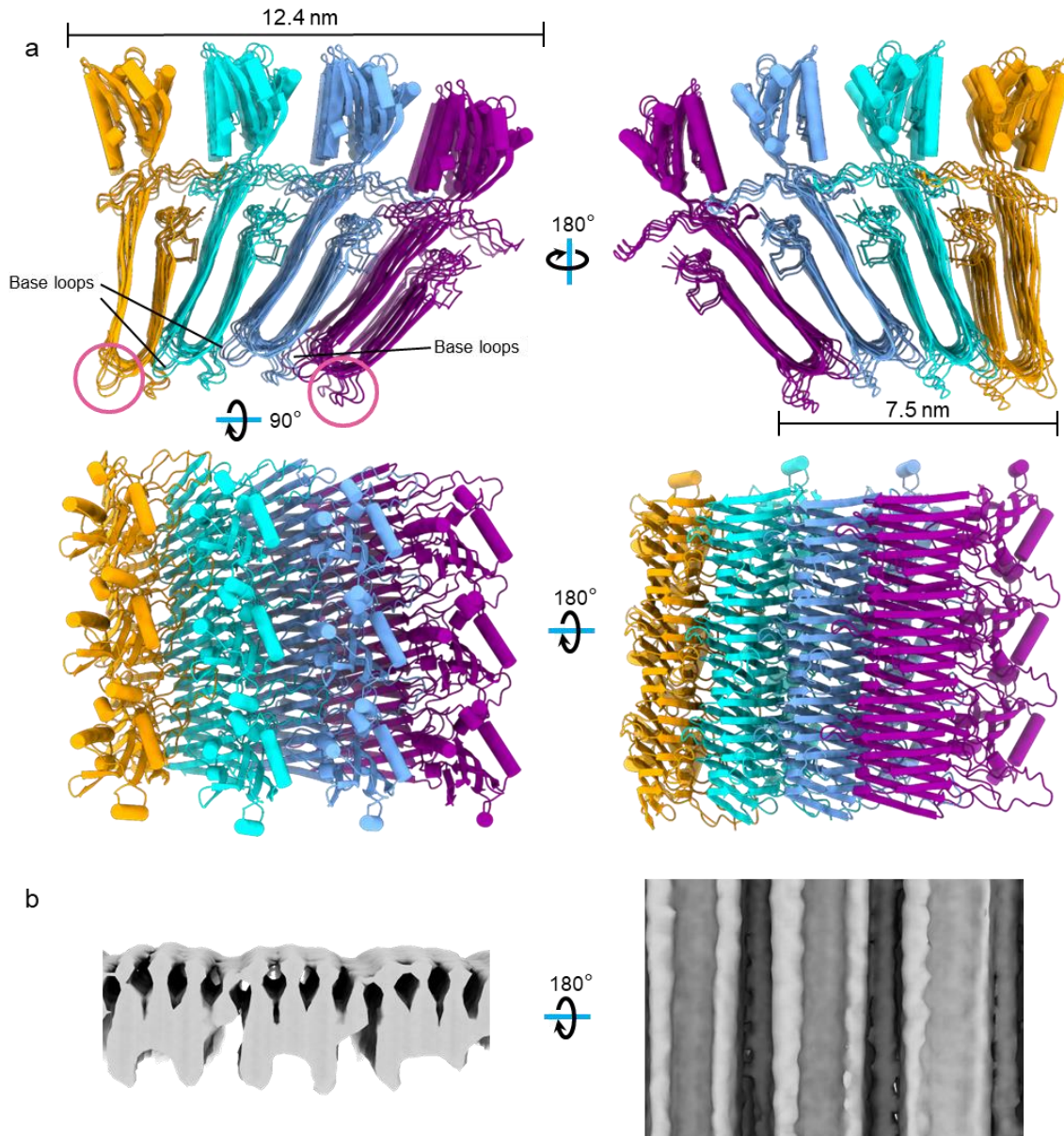
Supplemental Figure 6- 9. Negative stain TEM image of a region of a “ghost” *M. hungatei* cell showing overlapped hoop and flattened sheath segment

The yellow-boxed region in the left panel highlights a segment of sheath layer within which β -rings switch membership between neighboring β -hoops, as exemplified by those in the two insets. The stain-filled vertical dark strips (pointed to by red triangles) are the gaps between neighboring β -hoops. As indicated by its variable width, a β -hoop could change from a 5- β -ring-hoop (~15nm in width) to a 4- β -ring-hoop (~12nm in width).



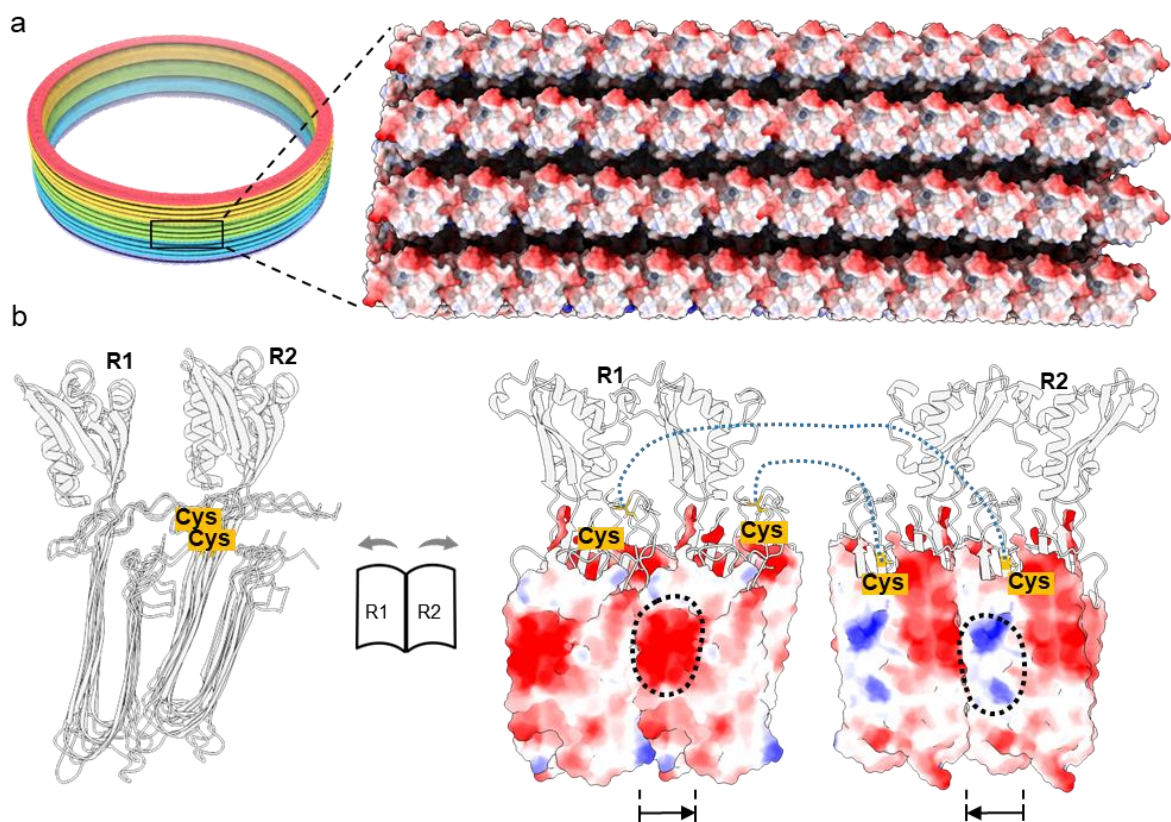
Supplemental Figure 6- 10. Inner connection between β -sheet 1 and 2.

Salt bridges (D189-K222 or E238-K222, and D354-K367) connect from β -sheet 1 to β -sheet 2 in the amyloid-like domain. The region and viewing orient of insets are indicated in the left guide image.



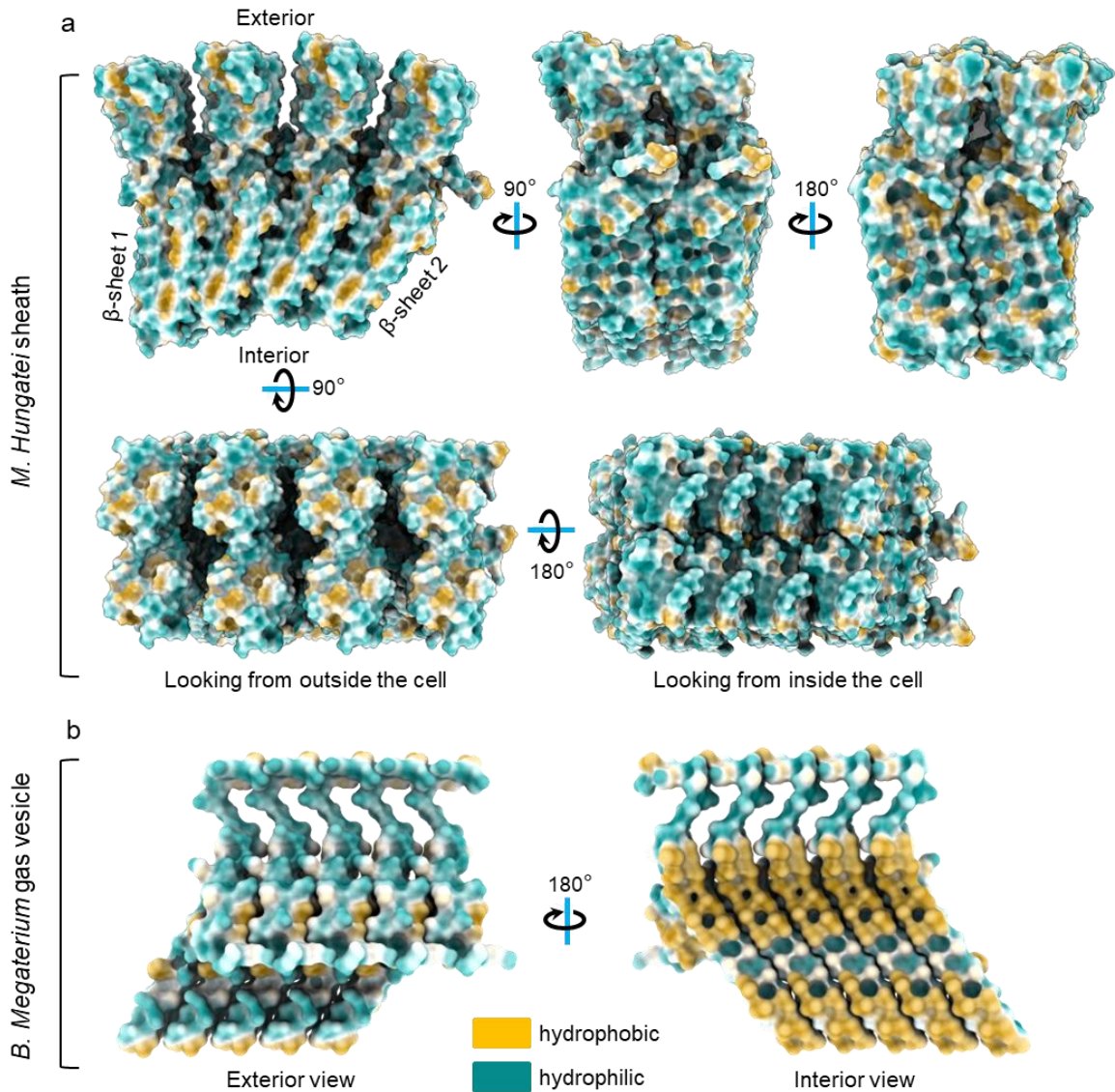
Supplemental Figure 6- 11. Atomic model of the 4- β -ring hoop assembly and subtomogram average of β -hoop with extra densities near the plug end.

a, Orthogonal views of a 4- β -ring hoop segment are shown as ribbon and colored by β -rings. Each β -ring segment contains 3 SH subunits. The exposed base loops on are circled both sides of the inter-hoop cleft. **b**, Orthogonal views of the subtomogram average of β -hoop near the plug end and with extra attached densities. The extra densities are indicated by pink arrowheads and their locations are correspond to the free base loops as circled in **a**.



Supplemental Figure 6- 12. Surface charge of β -hoop.

a, A 4- β -ring hoop segment, with extensive negative (red) electrostatic potential at exterior surface. **b**, An open book view of neighboring β -rings (R1 and R2) in a β -hoop, showing the distribution of the complementary electrostatic potential at their interface (red and blue corresponds to negative and positive charges, respectively) and potential disulfide bonds as indicated by the dotted lines.



Supplemental Figure 6- 13. Hydrophobicity of β -hoop.

a, Orthogonal surface views of a 4- β -ring hoop segment containing 8 SH subunits, colored by hydrophobicity. **b**, a segment of the *Bacillus megaterium* gas vesicles (PDB: 7R1C), colored by hydrophobicity, shows its strong hydrophobic interior and hydrophilic exterior.

6.11 Table

Table 1: CryoET data collection and processing statistics

| | High Mag. (Whole cell) | Low Mag. (Whole cell) | Low Mag. (Immature cell) |
|---|-----------------------------|--------------------------|-----------------------------|
| Data collection | | | |
| Microscope | Titan Krios | Titan Krios | Titan Krios |
| Voltage (kV) | 300 | 300 | 300 |
| Volta Phase Plate (VPP) ⁵⁴ | No | Yes | No |
| Total Electron exposure (e ⁻ /Å ²) | 110 | 110 | 100 |
| Slit width (eV) | 20 | 20 | 20 |
| Detector | K2 | K2 | 4 mega pixel CCD |
| Defocus range (μm) | -1.5 to -4.0 | N.D. | -3.0 to -6.0 |
| Pixel size (Å) | 1.634 | 4.050 | 6.102 |
| Software | SerialEM ⁴⁵ | SerialEM ⁴⁵ | Batch Tomography |
| Tilt-series range | ±60° | ±60° | ±70° |
| Tilt-series increment | ±3° | ±3° | ±2° |
| Tilt-series scheme | bi-directional | bi- directional | continuous |
| Tilt-series used | 10 | 1/23 | 1/37 |
| Data processing | | | |
| Software: tilt-series alignment | IMOD ⁴⁷ | IMOD ⁴⁷ | IMOD ⁴⁷ |
| Software: final reconstruction | Relion4.0 ⁵⁰ | N.D. | N.D. |
| Initial particle images: 3, 4, 5-β-ring hoop (no.) | 3,323, 20,137, and 4,237 | N.D. | N.D. |
| Final particle images: 3, 4, 5-β-ring hoop (no.) | 2,971, 17,359, and 3,609 | N.D. | N.D. |
| Final Box-size (px) | 120, 160, 120 | N.D. | N.D. |
| Pixel size final reconstruction (Å) | 3.268, 1.634, 3.268 | N.D. | N.D. |
| Symmetry imposed | C1 | N.D. | N.D. |
| Map resolution: 3, 4, 5-β-ring hoop (Å) | 21.8, 7.9, and 14.0 | N.D. | N.D. |
| FSC threshold | 0.143 | N.D. | N.D. |

N.D.: not determined

Table 6- 2. CryoET data collection and processing statistics

6.12 References

- 1 Ferry, J. G., Smith, P. H. & Wolfe, R. S. Methanospirillum, a New Genus of Methanogenic Bacteria, and Characterization of Methanospirillum hungatii sp.nov. *International Journal of Systematic and Evolutionary Microbiology* **24**, 465-469, doi:<https://doi.org/10.1099/00207713-24-4-465> (1974).
- 2 Hedderich, R. & Whitman, W. B. Physiology and biochemistry of the methane-producing Archaea. *The prokaryotes* **2**, 1050-1079 (2006).
- 3 Toso, D. B., Henstra, A. M., Gunsalus, R. P. & Zhou, Z. H. Structural, mass and elemental analyses of storage granules in methanogenic archaeal cells. *Environ Microbiol* **13**, 2587-2599, doi:10.1111/j.1462-2920.2011.02531.x (2011).
- 4 Poweleit, N. *et al.* CryoEM structure of the Methanospirillum hungatei archaeellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nat Microbiol* **2**, 16222, doi:10.1038/nmicrobiol.2016.222 (2016).
- 5 Zeikus, J. G. & Bowen, V. G. Fine structure of Methanospirillum hungatii. *J Bacteriol* **121**, 373-380, doi:10.1128/jb.121.1.373-380.1975 (1975).
- 6 Patel, G. B., Roth, L. A. & Sprott, G. D. Factors Influencing Filament Length of Methanospirillum hungatii. *Microbiology* **112**, 411-415, doi:<https://doi.org/10.1099/00221287-112-2-411> (1979).
- 7 Gunsalus, R. P. *et al.* Complete genome sequence of Methanospirillum hungatei type strain JF1. *Stand Genomic Sci* **11**, 2, doi:10.1186/s40793-015-0124-8 (2016).
- 8 Shaw, P. J., Hills, G. J., Henwood, J. A., Harris, J. E. & Archer, D. B. Three-dimensional architecture of the cell sheath and septa of Methanospirillum hungatei. *J Bacteriol* **161**, 750-757, doi:10.1128/jb.161.2.750-757.1985 (1985).
- 9 Beveridge, T. J., Sprott, G. D. & Whippey, P. Ultrastructure, inferred porosity, and gram-staining character of Methanospirillum hungatei filament termini describe a unique cell

- permeability for this archaeobacterium. *J Bacteriol* **173**, 130-140, doi:10.1128/jb.173.1.130-140.1991 (1991).
- 10 Southam, G. & Beveridge, T. J. Characterization of novel, phenol-soluble polypeptides which confer rigidity to the sheath of *Methanospirillum hungatei* GP1. *J Bacteriol* **174**, 935-946, doi:10.1128/jb.174.3.935-946.1992 (1992).
- 11 Southam, G. & Beveridge, T. J. Dissolution and immunochemical analysis of the sheath of the archaeobacterium *Methanospirillum hungatei* GP1. *J Bacteriol* **173**, 6213-6222, doi:10.1128/jb.173.19.6213-6222.1991 (1991).
- 12 Southam, G. *et al.* Transmission electron microscopy, scanning tunneling microscopy, and atomic force microscopy of the cell envelope layers of the archaeobacterium *Methanospirillum hungatei* GP1. *J Bacteriol* **175**, 1946-1955, doi:10.1128/jb.175.7.1946-1955.1993 (1993).
- 13 Beveridge, T. J., Stewart, M., Doyle, R. J. & Sprott, G. D. Unusual stability of the *Methanospirillum hungatei* sheath. *J Bacteriol* **162**, 728-737, doi:10.1128/jb.162.2.728-737.1985 (1985).
- 14 Stewart, M., Beveridge, T. J. & Sprott, G. D. Crystalline order to high resolution in the sheath of *Methanospirillum hungatei*: a cross-beta structure. *J Mol Biol* **183**, 509-515, doi:10.1016/0022-2836(85)90019-1 (1985).
- 15 Sprott, G. D., Beveridge, T. J., Patel, G. B. & Ferrante, G. Sheath disassembly in *Methanospirillum hungatei* strain GP1. *Canadian Journal of Microbiology* **32**, 847-854, doi:10.1139/m86-156 (1986).
- 16 Christensen, L. F. B. *et al.* The Sheaths of *Methanospirillum* Are Made of a New Type of Amyloid Protein. *Front Microbiol* **9**, 2729, doi:10.3389/fmicb.2018.02729 (2018).

- 17 Xu, W. *et al.* Modeling and measuring the elastic properties of an archaeal surface, the sheath of *Methanospirillum hungatei*, and the implication of methane production. *J Bacteriol* **178**, 3106-3112, doi:10.1128/jb.178.11.3106-3112.1996 (1996).
- 18 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682, doi:10.1038/s41592-022-01488-1 (2022).
- 19 Li, B. *et al.* Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications* **9**, 3609, doi:10.1038/s41467-018-05971-2 (2018).
- 20 Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185-190, doi:10.1038/nature23002 (2017).
- 21 McGlinchey, R. P., Ni, X., Shadish, J. A., Jiang, J. & Lee, J. C. The N terminus of alpha-synuclein dictates fibril formation. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2023487118 (2021).
- 22 Kollmer, M. *et al.* Cryo-EM structure and polymorphism of A β amyloid fibrils purified from Alzheimer's brain tissue. *Nature Communications* **10**, 4760, doi:10.1038/s41467-019-12683-8 (2019).
- 23 Hallinan, G. I. *et al.* Structure of Tau filaments in Prion protein amyloidoses. *Acta Neuropathologica* **142**, 227-241, doi:10.1007/s00401-021-02336-w (2021).
- 24 Patel, G. B., Sprott, G. D., Humphrey, R. W. & Beveridge, T. J. Comparative analyses of the sheath structures of *Methanothrix concilii* GP6 and *Methanospirillum hungatei* strains GP1 and JF1. *Canadian Journal of Microbiology* **32**, 623-631, doi:10.1139/m86-117 (1986).
- 25 Riek, R. & Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature* **539**, 227-235, doi:10.1038/nature20416 (2016).

- 26 Pospich, S. & Raunser, S. The molecular basis of Alzheimer's plaques. *Science* **358**, 45-46, doi:doi:10.1126/science.aap8002 (2017).
- 27 Hennetin, J., Jullian, B., Steven, A. C. & Kajava, A. V. Standard conformations of beta-arches in beta-solenoid proteins. *J Mol Biol* **358**, 1094-1105, doi:10.1016/j.jmb.2006.02.039 (2006).
- 28 Kajava, A. V., Baxa, U. & Steven, A. C. Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils. *FASEB J* **24**, 1311-1319, doi:10.1096/fj.09-145979 (2010).
- 29 Olsen, A., Arnqvist, A., Hammar, M., Sukupolvi, S. & Normark, S. The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csgA*, the subunit gene of fibronectin-binding curli in *Escherichia coli*. *Mol Microbiol* **7**, 523-536, doi:10.1111/j.1365-2958.1993.tb01143.x (1993).
- 30 Dueholm, M. S. *et al.* The Tubular Sheaths Encasing *Methanosaeta thermophila* Filaments Are Functional Amyloids. *J Biol Chem* **290**, 20590-20600, doi:10.1074/jbc.M115.654780 (2015).
- 31 Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu Rev Biophys* **41**, 21-40, doi:10.1146/annurev-biophys-050511-102312 (2012).
- 32 Zhou, Z. H., Zhang, H., Jakana, J., Lu, X. Y. & Zhang, J. Q. Cytoplasmic polyhedrosis virus structure at 8 Å by electron cryomicroscopy: structural basis of capsid stability and mRNA processing regulation. *Structure* **11**, 651-663, doi:10.1016/s0969-2126(03)00091-1 (2003).
- 33 Zhou, Z. H. *et al.* Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nat Struct Biol* **8**, 868-873. (2001).

- 34 Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368-373, doi:10.1038/276368a0 (1978).
- 35 Prasad, B. V., Wang, G. J., Clerx, J. P. & Chiu, W. Three-dimensional structure of rotavirus. *J Mol Biol* **199**, 269-275, doi:10.1016/0022-2836(88)90313-0 (1988).
- 36 Baker, T. S. *et al.* Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J* **60**, 1445-1456, doi:10.1016/S0006-3495(91)82181-6 (1991).
- 37 Liu, H. *et al.* Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* **329**, 1038-1043, doi:10.1126/science.1187433 (2010).
- 38 Ge, P. & Zhou, Z. H. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proc Natl Acad Sci U S A* **108**, 9637-9642, doi:10.1073/pnas.1018104108 (2011).
- 39 Matilla-Cuenca, L. *et al.* Bacterial biofilm functionalization through Bap amyloid engineering. *npj Biofilms and Microbiomes* **8**, 62, doi:10.1038/s41522-022-00324-w (2022).
- 40 Shanmugam, N. *et al.* Microbial functional amyloids serve diverse purposes for structure, adhesion and defence. *Biophysical reviews* **11**, 287-302, doi:10.1007/s12551-019-00526-1 (2019).
- 41 Voorhees, R. M. & Hegde, R. S. Structure of the Sec61 channel opened by a signal sequence. *Science* **351**, 88-91, doi:10.1126/science.aad4992 (2016).
- 42 Baquero, D. P. *et al.* Structure and assembly of archaeal viruses. *Adv Virus Res* **108**, 127-164, doi:10.1016/bs.aivir.2020.09.004 (2020).
- 43 Dutka, P. *et al.* Structure of *Anabaena flos-aquae* gas vesicles revealed by cryo-ET. *Structure* **31**, 518-528 e516, doi:10.1016/j.str.2023.03.011 (2023).

- 44 Huber, S. T., Terziel, D., Evers, W. H., Maresca, D. & Jakobi, A. J. Cryo-EM structure of gas vesicles for buoyancy-controlled motility. *Cell* **186**, 975-986 e913, doi:10.1016/j.cell.2023.01.041 (2023).
- 45 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).
- 46 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
- 47 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76, doi:10.1006/jsbi.1996.0013 (1996).
- 48 Nicastro, D. *et al.* The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944-948, doi:10.1126/science.1128618 (2006).
- 49 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, doi:10.1016/j.jsb.2011.05.011 (2011).
- 50 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *bioRxiv*, 2021.2009.2030.462538, doi:10.1101/2021.09.30.462538 (2021).
- 51 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature methods* **14**, 793-796 (2017).
- 52 Liu, Y.-T. *et al.* Isotropic reconstruction for electron tomography with deep learning. *Nature Communications* **13**, 6482, doi:10.1038/s41467-022-33957-8 (2022).
- 53 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

- 54 Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences* **111**, 15635-15640, doi:doi:10.1073/pnas.1418377111 (2014).

Chapter 7: Conclusion

In our endeavor to explore the intricate flagellar structure of *T. brucei*, we successfully resolved its components at an impressive ~ 20 Å resolution using cryoET and STA. This achievement opened a window to the *in situ* structure, facilitating direct observation of critical proteins, such as dynein, which are the driving force behind flagellar motility. Our comprehensive examination revealed the mysterious PFR structure, showcasing the elegant simplicity of a complex structure composed primarily of only two homologous filamentous proteins. Furthermore, the insights we derived from the PFR model allowed us to propose a novel model elucidating how SSN planes support the non-planar helical wave in the *T. brucei* flagellum. We also achieved the resolution of the PACs structure, providing invaluable evidence on bridging PFRs and the axoneme, thus accommodating the periodicity mismatch in the entire flagellum.

This work addressed substantial challenges within the cryoET and STA pipeline by introducing two innovative software solutions. IsoNet masterfully mitigated the persistent issue of "missing wedge" artifacts, offering a broad application across the cryoET field. Through deep neural networks, IsoNet effectively fills the missing wedge area with meaningful signal, eliminating the elongation artifacts along the Z-axis. IsoNet's applications demonstrated its prowess in restoring missing information, enabling direct observation without missing wedge artifacts, a vital advancement for examining crowded cellular environments.

Moreover, TomoNet introduced automation into the workflow, seamlessly integrating optimized template matching and deep learning for particle picking. This versatile combination simplified the management of entire tomography projects, efficiently handling extensive tomogram datasets through an intuitive graphical interface. TomoNet's application in studying *M. hungatei*'s sheath structure exemplifies its utility as a universal solution for particle arrangement within lattice-like configurations. This application yielded a groundbreaking

resolution of 7.9 Å, unravelling the topological intricacies and subunit organization comprising the native cylindrical sheath tube. We are proud to be the first to achieve sub-nanometer resolution in the archaeal sheath layer, allowing us to establish an atomic model for the sheath's topology, establish the assembly model from sheath monomers to β -rings, then to β -hoops, and finally into the elongated sheath tube. Immature cell tomograms captured in the cryoET suggest models for sheath biogenesis, involving the oligomerization of sheath monomers into β -ring precursors, followed by translocation through the unplugged end of a dividing cell, and subsequent insertion of nascent β -hoops into the immature sheath cylinder at the junction between two daughter cells.

In summary, this work underscores the remarkable potential of cryoET and STA, enriched by innovative software solutions, in unraveling the complexities of biological structures at near-atomic resolution.