DOCTORAL DISSERTATION

Structural studies of *Marseilleviridae* virus

Chihara, Akane

The Graduate University for Advanced Studies, SOKENDAI

School of Life Science

Department of Physiological Sciences

INDEX

Acknowledgements

First of all, I would like to thank all the people who have helped and inspired me in my doctoral study. Especially, I would like to express my deepest gratitude to Dr. Kazuyoshi Murata for his valuable support and advice as a supervisor in the research contents and the writing of the science paper and this thesis. He also took care of my health throughout my PhD training.

I thank Dr. Kaoru Mitsuoka and Dr. Naoko Kajimura (Research center for Ultra-High Voltage Electron Microscopy, Osaka University) for teaching and helping me about high voltage electron microscope techniques. I also thank Dr. Masaharu Takemura (Tokyo University of Science) for technical advice on *Acanthamoeba castellanii* culture and giant virus infection. I thank Dr. Yoshihiro Kubo and Dr. Yumiko Yoshimura (National Institute for Physiological Sciences) for their advice in my study and support in my daily life. I thank Dr. Toshinobu Suzaki for his support in my entrance into SOKENDAI.

I am very grateful to all members in the laboratory, Dr. Chihong Song, Dr. Raymond N. Burton-Smith, Dr. Yoko Kayama, Dr. Lin Chen, Dr. Minemasa Hida, Ms. Sachiko Yamada, Mr. Mitsuru Ikeda, Ms. Yoshie Kawaguchi, Mr. Ryoto Watanabe, Mr. Haruka Kawabata. I deeply appreciate all those who have supported and advised me in my study and daily life.

Finally, I would like to give my greatest thanks to my best friend and husband, Kentaro Kawawaki. Despite his busy work schedule, he always encouraged and supported me. His deep understanding and constant support enabled me to complete this work.

Summary

The “giant viruses” are exceptionally large physical size viruses, which are larger than small bacteria. They also have a much larger genome (>100 kilobases (kb)) than other viruses and contain many genes (> 50 genes) not found in other viruses. Giant viral particles are composed of a lipid bilayer covering viral DNA and a protein shell called capsid or tegument. Capsid and tegument show various structures. Although the recent development of cryo-electron microscopy (cryo-EM) has revealed the structures of many protein complexes and viruses at near-atomic resolution, structural studies of giant virus are scarce. It is because of the large particle size (> 200 nm) that is not suitable for observation with general cryo-EM with an accelerating voltage of 300 kV or less.

In this study, I use high-voltage cryo-electron microscope (cryo-HVEM) for single particle analysis (SPA) to investigate the structure of a giant virus at a higher resolution, and elucidate how the giant virus capsid is stably maintained and functionally used for viral infection. Theoretically, using cryo-EM with higher accelerating voltages, images of large specimens can be obtained at higher resolution.

I used a 1,000 kV cryo-HVEM (JEM-1000EES, JEOL) installed in Osaka University, for single particle structural analysis (SPA) of an icosahedral giant virus, tokyovirus (~250 nm), which is a species of *Marseilleviridae*. Tokyovirus was cultured with a *Acanthamoeba castellanii* cells and harvested. Then, purified virus particles were plunged frozen on holy-carbon supported EM grids for cryo-EM observation. SPA of cryo-HVEM images was performed by RELION 3.1 software. A homology model of the major capsid protein (MCP) was built based on the amino acid sequence and fitted into the cryo-EM map to assess additional protein components.

The structure of tokyovirus was reconstructed at 7.7 Å resolution from 1,182 particles from totally 304 images by imposing icosahedral symmetry. Tokyovirus mainly consisted of four layers surrounding central viral DNA; externally, major capsid protein (MCP) arrays, minor capsid protein (mCP) network, scaffold protein component (ScPC), and internal membrane. The internal membrane exhibited a characteristic structure extruding under the 5-fold vertices. The tokyovirus capsid exhibited a novel mCP network, supporting the outer MCP, which were structurally classified into eight protein components. Based on their function, these components are named lattice protein component (LtC), support protein component (SuC), cement protein component (CmC), zipper protein component (ZpC), glue protein component (GlC), and pentasymmetron protein components α, β, γ (PC-α, β, γ). In addition, newly identified scaffold protein component (ScPC) makes a bridge between 5-fold vertices, stabilizing mCPs and internal nuclear membrane. A homology model of tokyovirus MCP fitted into the cryo-EM map reveled that a cap structure exists above the MCP trimer.

In this study, I used 1,000 kV cryo-HVEM to investigate the structural of tokyovirus, overcoming the resolution limit imposed by the depth-of-field effect of very large objects. The 3D structure at 7.7 Å resolution showed the highest resolution of a giant virus larger than 200 nm without the block-based reconstruction technique. The mCP network of tokyovirus was complex compared to other structurally know giant viruses such as PBCV-1 and ASFV. Since the tokyovirus capsid (T = 309) is considerably larger than that of PBCV-1 (T = 169) and ASFV (T = 277), it is likely that the larger capsid is maintained by the more complex system of mCPs than these viruses. ScPCs running between the 5-fold vertices were directly interacting with the internal membrane extrusion at both ends. ScPCs may played a role in forming the characteristic membrane extrusion located under

the 5-fold vertices of this virus family icosahedral capsid. The fitted MCP model into the cryo-EM map reveled that a cap structure exists above the MCP trimer. The cap structure was suggested to be glycosylated and play a role in host cell recognition, particularly inducing the bunch formation in other lineages of the *Marseilleviridae* family.

I performed SPA of the entire tokyovirus particle using 1,000 kV cryo-HVEM and revealed the 3D structure of tokyovirus at 7.7 Å resolution. The result showed that 1,000 kV cryo-HVEM has a potential for structural analysis of large biological specimens such as giant viruses. The 1,000 kV cryo-HVEM SPA is technically not sufficient for hardware and software for higher resolution at present, but solving current problems will make it a useful tool for analyzing the structure of the large biological specimens in the near future.

Abbreviations

ASFV : African swine fever virus

CmC : Cement protein component

Cryo-EM : Cryo-electron microscopy

CTF : Contrast transfer function

GBP : Galactose binding protein

GlC : Glue protein component

hpi : hour(s) post infection

HVEM : High-voltage electron microscopy

IM : Internal membrane

LtC : Lattice protein component

MCP : Major capsid protein

mCP : Minor capsid protein

NC : Nucleocapsid

NCLDV : Nucleo-cytoplasmic large DNA virus

PAS staining : Periodic acid Schiff staining

PBCV-1 : *Paramecium bursaria* chlorella virus 1

PC-α, β and γ : Pentasymmetron component -α, β and γ

SBF-SEM : Serial block-face scanning electron microscopy

ScPC : Scaffold protein component

SGIV : Singapore grouper iridovirus

SPA : Single particle analysis

SuC : Support protein component

TEM : Transmission electron microscopy

T number : Triangulation number

ZpC : Zipper protein component

Introduction

Giant virus

A “giant virus” is a virus that has an exceptionally large physical size, which is larger than a small bacterium (Brandes and Linial, 2019). It also have a much larger genome (>100 kilobases (kb)) than other viruses and contain many genes (> 50 genes) not found in other viruses (Koonin and Yutin, 2019). Giant viruses consist of a lipid bilayer that surrounds the viral DNA and a protein shell that further surround this, called capsid or tegument (Aylward et al., 2021). Capsid and tegument exhibit different structures depending on the species.

First giant virus was discovered serendipitously in 1992 during an investigation of an outbreak of pneumonia in a hospital in Bradford, England (Rowbotham, 1980). It was initially identified as a bacterium because its fibrous surface was Gram-positively stained. However, a preliminary study of the genome showed that it lacked a number of genes required by independently living. It was named “mimivirus” (for "mimicking microbe"), reflecting its large size and apparent Gram-staining properties (La Scola et al., 2003). Since the discovery of mimivirus, many giant viruses have been reported around the world.

These giant virus are now taxonomically classified into the phylum *Nucleocytoviricota* (Aylward et al., 2021), but have historically been referred to be nucleo-cytoplasmic large DNA virus (NCLDV). NCLDV is an expansive clade of large viruses that possess double-stranded DNA and target varying host eukaryotes as a host. NCLDV currently consists of several families, including the *Asfarviridae, Ascoviridae, Iridoviridae, Marseilleviridae,*

Mimiviridae, Phycodnaviridae, and *Poxviridae*, and unclassified viruses such as cedratviruses , faustoviruses, medusaviruses, molliviruses, orpheoviruses, pacmanviruses, pandoraviruses, and pithoviruses (Iyer et al., 2001) (Figure 1). Furthermore, a new order, Megavirales, has been proposed based on the shared characteristics of these viruses (Colson et al., 2013c).

Most studies of giant viruses are focused on phylogenetic analyses based on genomic information, and structural studies of giant virus are very few.

Marseilleviridae

Marseilleviridae is a family of the new order of NCLDVs (Colson et al., 2013a), which have a highly complex genome of ~360 kb and a large particle size of ~250 nm. Currently, the known hosts for *Marseilleviridae* viruses are *Acanthamoeba* spp. The prevalence of these viruses in the environment and in animals and humans remains to be determined (Colson et al., 2013a). The first member of *Marseilleviridae* , Marseillevirus, was isolated in 2007 by culturing a water sample from a cooling tower in Paris, France, on *Acanthamoeba polyphaga* (Boyer et al., 2009). Since the discovery of the founder strain, many family members of *Marseilleviridae* have been isolated not only from aquatic environments, but also from organisms ranging from humans to insects (Colson et al., 2013a). Currently known major species in *Marseilleviridae* are Marseillevirus(Boyer et al., 2009), Melbournevirus (Doutre et al., 2014), Cannes 8 virus (Aherfi et al., 2013), Senegalvirus (Lagier et al., 2012), Tokyovirus (Takemura, 2016), Lausannevirus (Thomas et al., 2011), Port-miou virus (Doutre et al., 2015), Noumeavirus (Fabre et al., 2017), Kurlavirus (Chatterjee and Kondabagil, 2017), Tunisvirus (Aherfi et al., 2014),

Insectomime virus (Boughalmi et al., 2013), Brazilian marseillevirus (Dornas et al., 2016), and Golden marseillevirus (Dos Santos et al., 2016) (Figure 2). Several studies have also reported the presence of *Marseilleviridae* virus in humans (Colson et al., 2013b; Popgeorgiev et al., 2013; La Scola, 2014; Aherfi et al., 2016). However, there is controversy for *Marseilleviridae* virus infections of humans as other studies have shown no evidence (Sauvage et al., 2014; Macera et al., 2020). The family *Marseilleviridae* is currently classified into five lineages, from A to E (Chatterjee and Kondabagil, 2017; Fabre et al., 2017; Aoki et al., 2019). Melbournevirus, one of the *Marseilleviridae* belonging to lineage A, has initially been analyzed by cryo-electron microscopy (cryo-EM) single particle analysis (SPA) at 26 Å resolution (Okamoto et al., 2018). It shows the icosahedral capsid with a triangulation number of T = 309, the characteristic internal membrane inside the capsid that extrudes just underneath the 5-fold vertices, and the unique large density body inside the nucleoid.

Tokyovirus and other *Marseilleviridae* viruses in Japan

Tokyovirus is the first *Marseilleviridae* isolated in Asia in 2016, and is classified in lineage A (Takemura, 2016). It was isolated from a water/soil sample from Arakawa River in Tokyo, Japan, using a culture system of *Acanthamoeba castellanii.* In 2019, further *Marseilleviridae* viruses were discovered in Japan and named kyotovirus (lineage A), hokutovirus (lineage B), and kashiwazakivirus (lineage B), respectively (Aoki et al., 2019). Hokutovirus and kashiwazakivirus that belong to lineage B show two characteristic features. One is that they induce a cell aggregation named “bunch formation” in the early stage of infection (Figure 3). The bunch formation of the infected host cells

is inhibited by galactose (Aoki et al., 2021). The bunch formation by hokutovirus infected amoeba cells may be mediated through cell-cell interactions involving a galactose recognition process, which is supported by the members of *Marseilleviridae* containing a galactose-binding protein (GBP) gene in their genomes. The alignment of the amino acid sequences of several marseillevirus GBP proteins revealed differences in the respective N-terminal regions between lineage B and lineage A viruses.

The other is that the progeny viruses localize on the infected host cell surface (Figure 4). The relationship between the bunch formation and the viral particle localization on the host cell surfaces remains unclear.

Structural biology

Structural biology is the study of the molecular structure and dynamics of biological molecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function. Only genome information will not provide insight into the protein functioning, because proteins function as three-dimensional structures rather than one-dimensional chains of amino acids. Structural biology has become a major field in the 21st century, as it can provide the structural information of biological molecules. Structural biology not only contributes to basic science, but also aids drug development. The major methods of structural biology include X-ray crystallography, nuclear magnetic resonance (NMR), cryo-EM. These methods can reveal protein structures at atomic resolution.

Cryo-electron microscopy

Cryo-electron microscopy is a method for observing proteins and other biomolecules by irradiating them with electron beams while they are generally cooled by liquid nitrogen (-196°C). This method is suitable for structural analysis of proteins which are difficult to crystallize (for example large protein complexes and membrane proteins) because there is no need to crystallize the purified biomolecules.

In 1995, it was theoretically demonstrated that this new method could analyze the structure of molecules down to 100 kDa (Henderson, 1995). Since then, improvements in electron microscope hardware, image processing software and computing have laid the basis for what is known today as “the resolution revolution” in cryo-EM (Kühlbrandt, 2014). In 2017, the Nobel Prize in Chemistry was awarded to three researchers who majorly contributed to this development (Shen, 2018). Since then, cryo-EM has attracted attention as a structural analysis technique, and the number of reports on structural analysis using cryo-EM is increasing rapidly.

Single particle analysis

Single particle analysis (SPA) is an analysis method that extracts biomolecular images from two dimensions (2D) cryo-EM images, and averages and reconstructs them in three dimensions (3D) (Doerr, 2016). Currently, many structures of proteins and viruses have been resolved by SPA at higher resolution and determined their atomic structures. SPA has also been applied to giant viruses, such as mimivirus (Xiao et al., 2009), melbournevirus (Okamoto et al., 2018), Singapore grouper iridovirus (SGIV) (Pintilie et al., 2019), Faustovirus (Klose et al., 2016), *Chilo iridescent* virus (CIV) (Yan et al., 2009),

Cafeteria roenbergensis virus (CroV) (Xiao et al., 2017), but these have not reported at higher resolution (< 8 Å resolution). It is due to the large thickness of the vitreous ice (> 200 nm) that embedded with the giant virus particle.

SPA of giant viruses has often used cryo-EM with an accelerating voltage of 300 kV. At an acceleration voltage of 300 kV, the electron beam cannot penetrate the large and thick giant virus samples well, and multiple and inelastic scattering of electrons increases resulting in a lower signal-to-noise ratio. Furthermore, at an acceleration voltage of 300 kV, the depth of field is only a few hundred nanometers, which makes it impossible to focus on the entire giant virus particle and thus to acquire higher resolution images.

However, in 2019, the structure of *Paramecium bursaria* chlorella virus 1 (PBCV-1) at 3.5 Å resolution (Fang et al., 2019) and the structure of African swine fever virus (ASFV) at 4.1 Å resolution (Wang et al., 2019) were reported by the use of SPA, respectively. These SPA have used cryo-EM with an accelerating voltage of 300 kV, but utilized a hybrid methodology, termed “block-based reconstruction” (Zhu et al., 2018) to achieve these resolutions. The method focusses on sub-sections (“blocks”) of the virus particle to permit localized defocus refinement, resulting in reconstructions at higher resolution.

High voltage electron microscopy

In this study, I focus on the single particle reconstruction of the entire tokyovirus using 1,000 kV cryo-high-voltage electron microscopy (HVEM). HVEM was originally developed to extend attainable resolution using shorter electron wavelengths (Swann et al., 1974). Major usage is currently focused on thick specimens, which lower acceleration voltages (< 300kV) are unable to penetrate (Tanaka et al., 2020). Cryo-HVEM on

biological samples has not previously been reported for single particle analysis, and only a few examples using tomography have been reported (Murata et al., 2016; Okamoto et al., 2017; Murata and Kaneko, 2018).

As the acceleration voltage is increased, the wavelength of the emitted electron becomes shorter and the penetration of a sample improve. Further, increasing the accelerating voltage can increase the depth of field, and improve the (electron) optical conditions in thick samples (Downing and Glaeser, 2018). The resolution achievable with a given acceleration voltage can be simply calculated by the following formula, assuming all points within a given thickness can be considered equally focused :

$$d = \sqrt{(2\lambda t)} \qquad \text{(Eq. 1)}$$

Where d is the resolution, t is the sample thickness (in this case, particle diameter), and λ is the electron wavelength (Downing and Glaeser, 2018). Using equation (1), for example, for a 250 nm thick sample, at an acceleration voltage of 300 kV, the electron wavelength is 1.97 pm, so the theoretical resolution limit is ~9.9 Å. Extending this, at an accelerating voltage of 1,000 kV, the electron wavelength is 0.87 pm, so the theoretical resolution limit is calculated to improve to 6.6 Å (Figure 5A).

Further, considering phase error in the contrast transfer function (CTF) of the signal (DeRosier, 2000) is another estimation for calculating maximum attainable resolution (Figure 5B), and can be calculated by:

$$d = \sqrt{(\lambda t)} / ((2 \times 0.7)) \quad \text{(Eq. 2)}$$

In this case, the attainable resolutions are improved to 5.9 Å at 300 kV and 4.0 Å at 1,000 kV, respectively (Figure 5B). From these estimations, it was expected that the structure of the entire giant virus particle with a maximum diameter of 250 nm could be obtained at higher resolution using a 1,000 kV cryo-HVEM rather than using the block-based approaches.

Structure of icosahedral virus

In 1960s, a study of negative stained samples of *Sericesthis* iridescent virus (SIV) suggested that the icosahedral viral capsid was composed well-ordered assemblies of protein subunits named capsomer (Wrigley, 1969). The capsomers form pentamers at the vertexes of the icosahedral viral capsid and hexamers elsewhere (Rossmann and Johnson, 1989). The size of icosahedral viruses varies and it is defined with "triangulation (T) number" based on the capsomer's array (Caspar and Klug, 1962). The two lattice parameters (h and k) are positive integers that specify between the 5-fold vertices on the icosahedral capsid and T number is calculated by equation (3) below.

$$T = h^2 + hk + k^2 \qquad \text{(Eq. 3)}$$

The icosahedral virus capsid consists of a trimer of the major capsid protein (MCP) with a double jelly-roll structure, except at the vertices (Chelvanayagam et al., 1992; Cheng and Brooks, 2013). The MCP trimers coat the capsid surface as pseudo-hexamers. The icosahedral giant viruses also adopt the same protein assembly. However, the giant virus capsid is divided into two sub-capsid clusters called "trisymmetron",

“pentasymmetron”, each forming the surfaces around the 3-, and 5-fold icosahedral axes, respectively (Caspar and Klug, 1962) (Figure 6). T number of the icosahedral giant viruses exceeds hundred, and the size is adjusted by changing the size of trisymmetron rather than pentasymmetron.

Materials and Methods

Cultivation of *Acanthamoeba castellanii*

Acanthamoeba castellanii cells cultured in PYG medium (2% w/v proteose peptone, 0.1% w/v yeast extract, 4 mM $MgSO_4$, 0.4 mM $CaCl_2$, 0.05 mM $Fe(NH_4)_2(SO_4)_2$, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, 100 mM sucrose, pH 6.5) at 25 °C and dark condition.

Virus infection and collection

Tokyovirus (Takemura, 2016) was originally provided by Professor Masaharu Takemura, Tokyo University of Science. Viruses were infected and harvest as described previously (Doutre et al., 2014; Okamoto et al., 2018). The infected culture fluid was collected and centrifuged for 10 min at 1,500 g, 4 °C to remove dead cells, and then the supernatant was centrifuged for 35 min at 10,000 g, 4 °C to collect the virus particles.

Purification of tokyovirus

The pellet including virus particles was suspended in 1 mL of phosphate buffer solution and loaded onto a 10-60% sucrose gradient. The sample tube was centrifuged before further centrifugation for 90 min at 8,000 g, 4 °C. The resultant concentrated band was extracted and dialyzed in phosphate buffer solution before a further round of centrifugation with the same condition. The collected pellet was suspended in PBS before cryo-EM grids were prepared.

Sample preparation for cryo-EM

An aliquot (2.5 μL) of the purified tokyovirus particles was placed onto R 1.2/1.3 Quantifoil grids (Quantifoil Micro Tools GmbH, GERMANY) that were glow-discharged using a plasma ion bombarder (PIB-10, Vacuum Device Inc., JAPAN) beforehand. This grid was then blotted (blot time: 10 s, blot force: 10) and plunge-frozen using a Vitrobot Mark IV (Thermo Fisher Scientific, US) at 95% humidity and 4 °C.

Data acquisition with cryo-HVEM

A total of 304 micrographs were manually collected using a JEM-1000EES (JEOL Inc., JAPAN) equipped with an autoloader stage and K2 Summit camera (Gatan Inc., US) , which was optimized for 1,000 kV during a total of seven sessions using three grids. Micrograph movie frames were collected in super resolution mode in the K2 camera at a magnification equivalent to 1.456 Å/pixel with a target defocus of 2-4 μm. Each exposure time was 32 seconds, at a frame interval of 0.2 seconds for a total of 160 frames per micrograph. The frames were stacked using EMAN2 (Tang et al., 2007). The detail conditions are shown in Table 1.

Image processing

Micrograph movies were imported into a beta build of RELION 3.1 (Zivanov et al., 2018). Motion correction was performed by MotionCor2(Zheng et al., 2017) using 5×5 patches with B-factor blurring of 500 $Å^2$. CTF estimation of the images was carried out using CTFFIND 4.1.10 (Rohou and Grigorieff, 2015) with the following parameters:

lower defocus limit, 2,500Å; upper defocus limit, 80,000Å; step size, 100; exhaustive search. After the first run using a box size of 512, good CTF fits were selected, and failed fits were re-run using a larger FFT box size. It was repeated until an FFT box size of 2,048. Other parameters were used with the default settings. Micrographs that still failed to find a good CTF fit were discarded. The "good" CTF fit image is defined as the image that shows distinguishable Thon rings and can be fitted to the simulated CTF curve. Finally, a total of 156 micrographs were selected, and 1,529 particles were manually picked from them. In the initial image process, the particles 4× downsampled from 1.431 to 5.824 Å/pixel (2,400×2,400 pixels boxes become 600×600 pixels boxes) were 2D-classified into 40 classes with a circular mask diameter of 2,600Å, angular sampling of 2° (automatically increased to 3.75° due to use of GPU acceleration) and a search range of 7 pixels. The selected 1,458 particles were carried in good classes were further 2D classified with 1° angular sampling (automatically increased to 1.825°) and "ignore CTF until first peak" option, resulting in 1,419 particles in further good classes. All processing in 3D was carried out with icosahedral symmetry imposed. An initial model was generated with the stochastic gradient descent algorithm in RELION 3.1. The particles classified into 5 3D classes, using 1.8° angular sampling for 25 iterations were further processed using 0.5° angular sampling for a further 25 iterations. Then, two best classes consisted of 1,297 particles were selected. 3D refinement was carried out using a spherical mask. Then, particles were re-extracted and re-centered with 3× downsampling (4.368 Å/pixel) and 2x downsampling (2.912 Å/pixel) continuously, and refined continuously. Magnification anisotropy refinement and CTF refinement carried out somewhat improved the attainable resolution. Furthermore, 3D classification was carried out with alignment disabled using a mask blocked the disordered internal volume (viral DNA) and

selected the final best class consisted of 1,182 particles was selected. Particle polishing showed negligible effect. Final resolution of post-processing depended heavily on the mask used: the capsid-only mask resulted in a 7.7 Å resolution, but the mask including the scaffold proteins and internal membrane lowed the resolution to 8.7 Å. A soft spherical mask gave a 9.4 Å resolution. Local resolution was calculated using the blocres module of Bsoft (Heymann and Belnap, 2007) with no mask applied. The full pathway through the SPA 3D reconstruction is summarized in Figure 7.

Segmentation of 3D reconstructed map

Each structural layer of tokyovirus is extracted by applying the spherical masks of each layer to the 3D reconstruction map. Each mask is created based on the radius from the center of the virus particle using EMAN2 software (Tang et al., 2007). Extracted layers (MCP, mCP, ScPC and IM) were colored and visualized using UCSF Chimera (Pettersen et al., 2004).

Segmentations for mCP components

Segmentations for each mCP component was performed with *Segger* interface (Pintilie and Chiu, 2012), a plugin to UCSF Chimera (Pettersen et al., 2004). The cryo-EM map of the mCPs was automatically divided into small blobs and the small blobs are manually grouped based on their repeated structures and locations. The exact boundaries between the mCP components are not clear, but the individual eight mCP components, named LtC, SuC, ZpC, CmC, GlC, and PC-α, β, and γ, were segmented from the cryo-EM map.

Homology search of tokyovirus MCP sequence

The amino acid sequence of tokyovirus MCP (Accession No. YP_009255130.1) was used as the query sequence, and homology search was performed using the BLASTP algorithm (Altschul et al., 1990) with PDB as a reference database. As a result, the MCP sequences of SGIV (Accession No. YP_164167) (Song et al., 2004) and PBCV-1 (Accession No. NP_048787.1) (Yanai-Balser et al., 2010) were detected with high scores (Table 2).

Amino acid sequence alignment of MCP

The amino acid sequence of tokyovirus MCP was aligned by PROMAL3D (Pei et al., 2008) for those of SGIV (PDB ID: 60JN) (Pintilie et al., 2019) and PBCV-1 (PDB ID: 5TIP) (De Castro et al., 2018), which were detected with to highly homology from the homology search.

Prediction and comparison of the secondary structure of MCP

The secondary structure of tokyovirus MCP was predicted by PSIPRED (Buchan and Jones, 2019) for the amino acid sequence of tokyovirus MCP (Accession No. YP_009255130.1). Then, the secondary structure of tokyovirus MCP was compared with PBCV-1 (PDB ID: 5TIP) (De Castro et al., 2018) and SGIV (PDB ID: 60JN) (Pintilie et al., 2019) on three-dimensions using UCSF Chimera (Pettersen et al., 2004).

Model building of MCP and extraction of the cap region

The SWISS-MODEL (Waterhouse et al., 2018) and I-TASSER (Yang and Zhang, 2015) servers were used to generate a homology model of tokyovirus MCP using the MCP model of PBCV-1 (PDB ID: 5TIP) (De Castro et al., 2018) as a template. The homology model was rigid body fit into a tokyovirus MCP trimer volume extracted from the 3-fold symmetry axis of the cryo-EM map, and regions of the model which lay outside of the volume were adjusted by COOT (Emsley and Cowtan, 2004) and energetically refined by PHENIX (Adams et al., 2010). Then, *molmap* in UCSF Chimera was used to generate a 15 Å molecular map of the MCP trimer, and the map was subjected to the *relion_mask_create* module of RELION 3.1 (Zivanov et al., 2018) to generate a binary mask with an extension of 5 pixels and a soft edge of 5 pixels. The mask of the MCP trimer was used to extract the cap region on the MCP trimer.

Visualization of data

The 3D reconstructions of tokyovirus were visualized using *relion_display* module of RELION 3.1 (Zivanov et al., 2018) or UCSF Chimera (Pettersen et al., 2004) .

Sample preparation for SBM-SEM

For SBF-SEM, prefixed *Acanthamoeba castellanii* cells that were infected with hokutovirus (8 hpi) provided by Professor Masaharu Takemura, Tokyo University of Science. The cells were then washed 2 times with phosphate buffer solution and postfixed with 1.5 % solution of potassium ferricyanide ($K_3[Fe(CN)_6]$) and 2 % osmium tetroxide

(OsO_4) in the same buffer at 4 ℃ for 1 hour. The post-fixed samples were washed 3 times with Milli-Q water and incubated in Thiocarbohydrazide (TCH) solution at room temperature for 20 min. The treated samples were washed 3 times with Milli-Q water and further fixed with 2 % OsO_4 at room temperature for 30 min. Then, the double-fixed samples were subjected with 1% uranyl acetate at 4 ℃ overnight and treated with aspartic acid solution at 60℃ for 30 min. The samples after sufficiently washing with Milli-Q water were dehydrated by a graded ethanol series (50-100%) at 4℃. Finally, the dehydrated samples were infiltrated with epoxy resin (Durcupan^{TM}ACM, single component A- D; SIGMA-Aldrich, US) and the resin was polymerized at 60 ℃ for 3 days.

Acquisition of SBF-SEM image stacks

The resin blocks containing the cell samples were manually trimmed with a razor blade and the blocks were cut off into a small piece and glued on the sample rivet with Conductive Epoxy (Durcupan^{TM}ACM, single component A- D; SIGMA-Aldrich, US). The specimens on the rivet were further trimmed to the smaller size (approximately 1.2 x 1.2 mm) with a razor blade and all the lateral surfaces of the specimen block was coated with conductive silver paint (Electron Microscopy Sciences, US). The conductive epoxy and silver paint served as a medium to dissipate the electric charge accumulated on the block surface as a result of irradiation by the electron beam during SEM imaging.

The scanning electron microscope (MERLIN, Carl Zeiss Co.,Ltd., GERMANY) equipped with a field-emission gun, a Gemini electron column, a 3View in-chamber ultramicrotome system (Gatan Inc., US), and a back-scattered electron detector was used, which was originally developed at the Max Planck Institute for Medical Research,

Heidelberg, Germany (Biazik et al., 2015). The image stack was acquired in an automated manner by using Gatan Digital Micrograph software (Gatan Inc., US) after adjusting the SEM imaging conditions and microtoming parameters. The electron microscope was operated at an acceleration voltage of 1.2 kV with a standard 30 μm aperture in the normal current mode and in the high-vacuum mode. All images were taken with the following scanning settings: dwell time = 3.0 μs; image size = 8,192 × 8,192 pixels (21.06 × 21.06 μm on the specimen; pixel size = 8 nm). Once a SEM image had been acquired, a 70-nm-thick layer was removed from the block surface and the freshly exposed surface was imaged again. This image acquisition cycle was repeated 1,000 times.

Alignment of the image stack, segmentation, and 3D representation

Serial images recorded at 8,192 × 8,192 pixels were subjected to band-pass filtration (15–1,000 nm) to reduce noise and to enhance the contrast. After 2× binning of the images, the image stack was automatically aligned (registration) in translation mode of the 'Register Virtual Stack Slices' in the Fiji/ImageJ software package (Schindelin et al., 2012). The aligned images were saved as a new stack file in the MRC format (Crowther et al., 1996) by using the IMOD software (Kremer et al., 1996). Individual cells and hokutoviruses were manually segmented by AMIRA software (FEI Visualization Science Group, US). This software was also used to generate the final figures.

SDS-PAGE and Periodic acid Schiff staining

A sample of the purified tokyovirus was treated with SDS-PAGE sample buffer (2332330/AE-1430 EzApply, ATTO Corp., JAPAN) and boiled for 5 minutes. 10 μg of each denatured sample was loaded into a well (2331635/CHR12.5L c, ATTO Corp., JAPAN). 7 μL of standard molecular marker (2332340/AE-1440, ATTO Corp., JAPAN) was also loaded to estimate the molecular weights of the resulting protein bands. Electrophoresis was performed at constant current of 10.5 mA with an electrophoresis buffer (2332323/WSE-7055, ATTO Corp., JAPAN) using an Electrophoresis system (2322240/WSE-1010, ATTO Corp., JAPAN). The resultant gel was stained with a CBB regent (2332370/AE-1340, ATTO Corp., JAPAN). Periodic acid Schiff staining was carried out according to the protocol described in the PAS stain kit (GlycoGel Stain Kit 24693-1, Polysciences, Inc., US)

Results

Image acquisition with 1,000 kV cryo-HVEM

304 micrographs were obtained with a 1,000 kV cryo-HVEM installed in Osaka University (Figure 8). Detailed conditions of image acquisition are described in Materials and Methods, and Table 1. The size of sample imposes limits which require careful balance between a lower magnification to increase particles imaged per micrograph and a high enough magnification to achieve worthwhile resolutions in reconstructions. Tokyovirus particle with a maximum diameter of 250 nm is very large specimen for SPA, requiring lower magnification to include multiple particles in a single image. However, at lower magnification, a sufficient amount of beam intensity cannot be obtained in the microscope. The exposure time of 32 seconds was used to obtained sufficient dose for higher resolution.

In this study, the “super-resolution” mode (7,420 × 6,780 pixels) of the K2 Summit camera was used for data acquisition, keeping the pixel spacing less than 1.5 Å/pixel at low magnification (Table 1). Super-resolution mode can detect the electrons with sub-pixel precision by virtually dividing 1 pixel into 4 pixels. Super-resolution mode is essential for large giant virus particles to maintain the resolution of the micrographs and the number of particles in the micrograph, though the detector performance suffers slightly (Figure 9).

Overall structure of tokyovirus

Figure 10 shows the 7.7 Å resolution overall structure of tokyovirus reconstructed by RELION 3.1 (Zivanov et al., 2018) with icosahedral symmetry imposed. At the outermost tokyovirus particle, MCP trimers were arranged in an icosahedral form with T = 309 (h = 7, k = 13) (Figure 10A). Inside of the MCP trimers array, layers of minor capsid proteins (mCPs) and scaffold protein components (ScPCs) exist in that order, and further inside the internal membrane stores the viral DNA (Figure 10C, D). Compared to the MCP trimers array and mCPs, densities of ScPCs are smeared, showing structural flexibility of this component (Figure 10D). Figure 10E shows the estimated local resolution of the tokyovirus particle sliced for internal visualization. The highest resolution was shown in the interface between the MCP trimers array and mCPs.

Structures of tokyovirus

Structure of tokyovirus was viewed by careful filtering and segmentation of the reconstruction volume (Figure 11). MCP trimers cover the entire surface of the viral capsid except for the 5-fold vertices (right blue in Figure 11) where pentagonal structures (called “pentons”) plug the hole in the vertices. The viral surface is formed with 20 trisymmetrons and 12 pentasymmetrons. Trisymmetron and pentasymmetron include 136 MCP trimers and 30 MCP trimers, respectively. mCPs form a hexagonal network under the MCP trimer in trisymmetron and pentasymmetron. Under the mCPs, ScPC forms anti-parallel chains between pentasymmetrons along the trisymmetron interfaces (yellow in Figures 12). The internal membrane shows a characteristic structure due to its ScPC, with the extrusions under the 5-fold vertices (Figure 10B, grey in Figure 11).

Minor capsid proteins

The mCPs of tokyovirus were segmented and extracted. As a result, eight types of protein components based on their structural features and arrangements were classified (Figure 12).

Since structural proteins other than MCP have not been identified in tokyovirus, these protein components do not directly specify individual mCPs and their genome. These eight mCP components were named Lattice protein component (LtC) (Figure 12B, 1 to 9), Cement protein component (CmC) (Figure 12C, I to V), Support protein component (SuC) (Figure 12D), Zipper protein component (ZpC) (Figure 12E, a to i), and Glue protein component (GlC) (Figure 12F, A to N), three pentasymmetron protein components of α, β, and γ (PC-α, β, and γ) (Figure 12G), in addition to penton in icosahedral vertices (Figure 12H), respectively. The mCP hexagonal network of trisymmetron formed by mCPs can be separated into three trapezoidal units consisted of five components, LtC, SuC, ZpC, CmC, and GlC, and these trapezoids are connected by rotating 120° around the 3-fold rotation axes. These trisymmetrons are connected to each other via ZpC and GlC, and connected to the pentasymmetrons via PC-β and PC-γ at the apexes of the trisymmetron.

The LtC forms a wavy chain along the gap of MCP trimers (orange to red in Figure 12A). The longest chain consists of five repeats of LtCs and the shortest chain consists of two repeats of LtCs (Figure 12B). This combination of the wavy chains builds the trapezoidal lattice unit, and further creates a base of the trisymmetron by supporting the MCP trimers. The CmCs repeats five times in the trapezoidal unit (Figure 12C). Within the trisymmetron, the CmCs connect the three trapezoidal units by interfacing the ends of five repeated LtCs (sky blue in Figure 12A). The SuC is included in only one the

trapezoidal unit (royal blue in Figure 12A). Under the hexagonal network structure formed by LtCs, the SuC creates a bridge between several LtCs (Figure 12A). The ZpCs repeated nine times in the trapezoidal unit (magenta in Figure 12A). They fill the gaps between LtCs at the trisymmetron interfaces. The GlCs repeated 14 times in the trapezoidal unit (green in Figure 12A). They are present at the boundary between adjacent trisymmetrons and pentasymmetrons, and they appear to glue them. Two protein components, ZpC and GlC (magenta and green in Figure 12A), are involved in connecting adjacent trisymmetrons and run directly above the ScPC array (yellow on Figure 12A).

In pentasymmetron, PC- α , β and γ , each exists in each asymmetric unit, support the MCP trimers array like the mCP components in trisymmetron (purple, light green, and cyan in Figure 12A). PC-α is located near the center of pentasymmetron and supports the pentons (dark pink in Figure 12A) placed in the center of the 5-fold vertices. PC-β and γ are located at the edges of pentasymmetron and play a role in connecting to neighboring trisymmetrons.

Scaffold protein component

ScPC is a novel component found in giant viruses, which is placed between mCP and internal membrane (yellow in Figure 11 and 13). ScPC has a large head and long tail structure (Figure 13B), which is arranged along the edge of trisymmetron (Figures 12A). The two ScPCs are connected antiparallel by their heads and tails, with both heads located at the edge of the pentasymmetron (Figure 13A). The adjacent ScPC pairs are connected to each other by SuC (royal blue in Figure 13A) at the end of the ScPC. The ScPCs form a framework of the internal membrane, creating the characteristic extrusion of the membrane. They interact with both the internal membrane and the mCPs (Figure 14).

Internal membrane and its interaction with ScPCs

The internal membrane exhibits a characteristic structure that extrudes under the 5-fold vertices. In the extrusion, the membrane forms multiple lipid layers (asterisk in Figure 14D). The ScPC array and SuCs directly interact with the extrusion of internal membrane, they may play a role in distorting the membrane into this extruded shape at the 5-fold vertices (Figure 14). Figure 14 shows a weak density between the mCPs and the membrane extrusion connecting the capsid and the internal membrane extrusion, but it is not clear what it is. The mCPs of pentasymmetron are thicker than those of trisymmetron (Figure 14D). It may have a DNA transfer function on the capsid.

Major capsid protein

The primary structure of tokyovirus MCP shows higher scores against SGIV (PDBID: 6OJN) (Pintilie et al., 2019) and PBCV-1 (PDBID: 5TIP) (De Castro et al., 2018), with high query cover and moderate identities (Table 2). Protein structure predictions suggest that the structure of the tokyovirus MCP, similar to that of other giant virus MCPs, has a double jelly-roll structure composed of two sets of eight β-strands (B1 to I2 in Figure 15). A homology model of tokyovirus MCP was generated based on the MCP model of PBCV-1 using SWISSMODEL server (Waterhouse et al., 2018), and compared to both MCP models of PBCV-1 and SGIV (Figure 16). In tokyovirus MCP, the loops connecting the external parts of jelly-roll 1 or jell-roll 2 are relatively long. The HI1 loop is particularly elongated with the 23-residue sequence. These loops suggest that MCP creates a high wall when forming MCP trimer on the viral surface.

Cap structure on the MCP trimer

To build the MCP trimer model, the MCP trimer volume was extracted from the center of trisymmetron and the homology model was rigid-body-fitted into the cryo-EM map. Regions of the model poorly fitted in the volume were manually corrected using COOT (Emsley and Cowtan, 2004) and energetically refined with PHENIX (Adams et al., 2010) (Figure 17B). As a result, an empty “cap” structure was newly identified on the MCP trimer (red in Figure 17C). The strong density in the MCP trimer volume suggested that other proteins are located above the MCP trimer.

Glycoprotein in tokyovirus

Periodic acid Schiff (PAS) staining (Aterman and Norkin, 1963) was applied to the SDS-PAGE gel from the purified tokyovirus particles. Tokyovirus MCP monomer, that is 52 kDa protein, was not stained by PAS staining (Figure 18A), but another protein at ~14 kDa was stained. This result showed that tokyovirus MCP is not glycosylated but another ~14 kDa glycoprotein is likely to be included in the virus.

Bunch formation by the lineage B *Marseiileviridae* viruses

Serial slice images of bunched *Acanthamoeba* cells caused by infection with lineage B hokutovirus (Aoki et al., 2019) were acquired by using SBF-SEM, and 3D images were reconstructed (Figure 19A, B). *Acanthamoeba* cells are clustered against central one cell (Figure 19C, transparent green). Furthermore, hokutovirus particles were distributed locally in the interfaces between the viral infected cell (Figure 19C, transparent green) and others (Figure 19C, red, blue, and transparent purple). The result suggested that cell-to-cell interactions were caused via hokutovirus particles.

Discussion

In this study, 1,000 kV cryo-HVEM was applied to structural analysis of tokyovirus to fully reconstruct the viral particle, overcoming the resolution limit imposed by the depth-of-field effect for exceptionally large object. The 3D structure at 7.7 Å resolution (Figure 10) showed the highest resolution of a giant virus larger than 200 nm without the block-based reconstruction technique (Zhu et al., 2018) or similar methods (Wang et al., 2019). As a result, the novel structures of the mCP network including the ScPC array were identified (Figure11 - 13), where the ScPC arrays bridging the 5-fold vertices directly interact with internal membrane, performing the characteristic membrane extrusion of the *Marseilleviridae* viruses (Figure 14). For the MCP trimers array, a unique cap structure was newly identified on the MCP trimer (Figure 17). The cap structure was suggested to be glycosylated and used to identify the host cell, particularly inducing the bunch formation in other lineages of the *Marseilleviridae* family.

Minor capsid protein network stabilizes the major capsid protein array

mCP networks have been specifically observed in icosahedral giant viruses, which have now been reported in PBCV-1 (Fang et al., 2019) and ASFV (Wang et al., 2019). It has been suggested that the mCP network is required to maintain the large capsids of giant viruses. Tokyovirus exhibits a more complex network arrangement of mCPs from PBCV-1 and AFSV. It consists of trapezoidal arrays of LtCs connected with CmC in trisymmetron (Figure 11). Two protein components of GlC and ZpC surrounding the

trisymmetron connect the neighboring trisymmetrons and connect to the pentasymmetrons. SuC interacting with each trapezoidal lattice on the interior surface connects the capsid to the internal membrane. Since the tokyovirus capsid (T = 309) is considerably larger than that of PBCV-1 (T = 169) and ASFV (T = 277), it is likely that the capsid is maintained by a more complex combination of multiple mCPs than these viruses.

Scaffold protein component

The array of "scaffold" protein component (ScPC) between the mCP network and the internal membrane (yellow in Figure 10 - 12) were newly identified in tokyovirus. The antiparallel pair of ScPCs running along the trisymmetron interface create a bridge between each pentasymmetron by interacting with the pentasymmetron components at both ends. These features are quite analogous to the "tape-measure" proteins (Xian et al., 2021) reported in PBCV-1 and ASFV structures (Fang et al., 2019; Wang et al., 2019), which is the long filamentous mCP named P2 in PBCV-1 and M1249L in ASFV and is proposed to determine the particle size. It suggests that the tokyovirus capsid may be constructed by a mechanism different from that of PBCV-1 and ASFV. The ScPC bridges can impose constraints on the capsid dimensions like tape measure proteins. Further structural investigations at higher resolution are necessary to elucidate its potential role as a tape measure in the future.

The extrusion of the internal membrane

Melbournevirus, one of *Marseilleviridae*, was first reported to possess a very

characteristic extrusion of the internal membrane under the 5-fold vertices (Okamoto et al., 2018). The 7.7 Å tokyovirus structure revealed that ScPC and SuC were directly interact with the internal membrane (Figure 13). In addition to these, the thick protein component in pentasymmetron (PC-α, β, γ) may play a role in forming the inner membrane extrusion. Special structures in the 5-fold vertices are also reported in other icosahedral giant viruses, for example, the single spike of PBCV-1 (Milrot et al., 2017) and the stargate structure of mimivirus (Zauberman et al., 2008). Viral DNA can be transferred to the host cell through these structures. *Marseilleviridae* viruses also presumably transport viral DNA through 5-fold vertices, and the special structures at 5-fold vertices may aid in stable transport of viral DNA. *Marseilleviridae* viruses have the special structure on all twelve vertices, whereas PBCV-1 and mimivirus have the special structure on one of the twelve vertices. The structures in *Marseilleviridae* viruses can efficiently transport viral DNA through multiple vertices. This may allow rapidly replication of *Marseilleviridae* virus.

Cap structure of MCP trimer may recognize the host cell

The “cap” structure above the MCP trimer was suggested to be a ~14 kDa glycoprotein by PAS staining (Figure 18). However, the band was relatively weak compared to that of MCP in the Coomassie stained gel (Fig 18A). In the case of PBCV-1, sugar chains directly bound to MCP (De Castro et al., 2018) are used for host cell recognition. Previous study has reported that lineage B hokutovirus and kashiwazakivirus in the *Marseilleviridae* induce a bunch formation of the host amoeba (Aoki et al., 2019). However, this bunch formation was inhibited by galactose (Aoki et al., 2021). It suggests that the bunch

formation by the *Marseilleviridae* viruses is associated with a galactose recognition process. It was also supported by the fact that the members of *Marseilleviridae* contain a GBP gene in their genomes (Aoki et al., 2021). Amino acid sequence comparisons of several marseillevirus GBP proteins revealed differences in their respective N-terminal regions between lineage B and lineage A viruses , which may have contributed to the different bunch-inducing abilities observed. Lineage B hokutovirus and kashiwazakivirus are known for progeny viruses to localize on the host cell surface after infection. The three-dimensional localization of the progeny hokutoviruses in the bunch formation, which was investigated by SBF-SEM (Figure 19), suggested that the bunch formation occurs via viral particles bound on the cell surface. Based on these observations, it is hypothesized that the glycosylated cap structure on the MCP trimer can contribute to the bunch formation of virus-infected amoeba by binding the viruses on the amoeba cell, and that this cap structure may differs between lineages, which determines the presence or absence of bunch formation. The glycoprotein forming the cap structure may directly play a role in these bunch formation and adhesion to the cell, potentially to increase transmission efficiency, though it needs further study to clarify the precise function of the glycoprotein.

Conclusion and Future Perspectives

In order to answer the questions in discussion section, higher resolution structures of these viruses and their comparisons with other lineages of *Marseilleviridae* viruses are necessary. The current resolution of 7.7 Å of tokyovirus has not reached the theoretical maximum of 4.0 Å for the ~250 nm particles (Figure 5B). There were several factors that

limited the current resolution of tokyovirus, but the main limiting factor for current resolution is likely the amount of data that can be collected. The 1,000 kV cryo-HVEM uses the manual collection of micrographs at present. The 1,182 particles from 156 micrographs used in this study were an extremely small number compared to other reported SPA of giant viruses using 300 kV microscopes. For example, PBCV-1 with a maximum diameter of 190 nm was reconstructed at 4.4 Å reconstruction with ~13,000 particles from 5,624 micrographs (Fang et al., 2019) and ASFV with a 250 nm maximum diameter of the outer capsid were reconstructed at 14.1 Å reconstruction with 16,266 particles from 17,135 micrographs (Liu et al., 2019), and at 8.8 Å reconstruction with 63,348 particles from 64,852 micrographs (Wang et al., 2019). Another limitation on the attainable resolution lies in software. Some elements of general-purpose cryo-EM SPA image processing software do not support HVEM accelerating voltages. The software support will be needed for high resolution SPA of giant viruses using 1,000 kV microscopes in the future.

In conclusion, using 1,000 kV cryo-HVEM provides benefits for larger particles. The decreased defocus gradient across a particle aids high resolution whole particle reconstruction. Sample damage from the electron beam is also decreased, which is of particular importance with biological samples. Furthermore, increased beam penetration permits improved visualization of internal structure, as used in cryo-electron tomography with pithovirus, an amphora-shaped giant virus having ~800 nm thickness (Okamoto et al., 2017). Beyond the more widely available 300 kV cryo-EM, it permits study of finer details of even larger viruses, which lower acceleration voltages are unable to penetrate. Finally, the reconstruction methodology is simpler than that of block-based reconstruction.

Block-based reconstruction is a powerful technique, but as the name implies, it breaks the whole structure down into blocks. Stitching these back together is called a composite map, and while these are gaining in popularity, do have some inherent dangers. For example, FSC calculations cannot be performed on composite maps without artifact, and map sharpening can reveal overlap artefacts between individual maps. Block-based reconstruction was tested on data sets from 1,000 kV cryo-HVEM, but it did not improve the resolution. Although the exact reasons are unknown, our hypothesis is that the relatively low number of particles in the final reconstruction and the lower curvature of the Ewald sphere at 1,000 kV (when compared to 300 kV) decreases the effectiveness of block-based reconstruction. However, block-based reconstruction methods may have some advantages in cryo-HVEM given sufficient particle numbers. This is because it can overcome minor distortions of very large symmetric particles, yielding improved resolutions. Furthermore, for objects with low symmetry the direct method using higher accelerating electrons is more available than block-based reconstruction .

References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213-221.

Aherfi S, Pagnier I, Fournous G, Raoult D, La Scola B, Colson P (2013) Complete genome sequence of Cannes 8 virus, a new member of the proposed family "Marseilleviridae". Virus Genes 47:550-555.

Aherfi S, Boughalmi M, Pagnier I, Fournous G, La Scola B, Raoult D, Colson P (2014) Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae. Arch Virol 159:2349-2358.

Aherfi S, Colson P, Audoly G, Nappez C, Xerri L, Valensi A, Million M, Lepidi H, Costello R, Raoult D (2016) Marseillevirus in lymphoma: a giant in the lymph node. Lancet Infect Dis 16:e225-e234.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Aoki K, Fukaya S, Takahashi H, Kobayashi M, Sasaki K, Takemura M (2021) Marseilleviridae Lineage B Diversity and Bunch Formation Inhibited by Galactose. Microbes Environ 36.

Aoki K, Hagiwara R, Akashi M, Sasaki K, Murata K, Ogata H, Takemura M (2019) Fifteen Marseilleviruses Newly Isolated From Three Water Samples in Japan Reveal Local Diversity of. Front Microbiol 10:1152.

Aterman K, Norkin S (1963) The periodic acid-Schiff reaction. Nature 197:1306.

Aylward FO, Moniruzzaman M, Ha AD, Koonin EV (2021) A phylogenomic framework for charting the diversity and evolution of giant viruses. PLoS Biol 19:e3001430.

Biazik J, Vihinen H, Anwar T, Jokitalo E, Eskelinen EL (2015) The versatile electron microscope: an ultrastructural overview of autophagy. Methods 75:44-53.

Boughalmi M, Pagnier I, Aherfi S, Colson P, Raoult D, La Scola B (2013) First isolation of a Marseillevirus in the Diptera Syrphidae Eristalis tenax. Intervirology 56:386-394.

Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D (2009) Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc Natl Acad Sci U S A 106:21848-21853.

Brandes N, Linial M (2019) Giant Viruses-Big Surprises. Viruses 11.

Buchan DWA, Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res 47:W402-w407.

Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1-24.

Chatterjee A, Kondabagil K (2017) Complete genome sequence of Kurlavirus, a novel member of the family Marseilleviridae isolated in Mumbai, India. Arch Virol 162:3243-3245.

Chelvanayagam G, Heringa J, Argos P (1992) Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. J Mol Biol 228:220-242.

Cheng S, Brooks CL, 3rd (2013) Viral capsid proteins are segregated in structural fold space. PLoS Comput Biol 9:e1002905.

Chihara A, Burton-Smith RN, Kajimura N, Mitsuoka K, Okamoto K, Song C, Murata K (2022) A novel capsid protein network allows the characteristic internal membrane structure of Marseilleviridae giant viruses. Sci Rep 12:21428.

Colson P, Pagnier I, Yoosuf N, Fournous G, La Scola B, Raoult D (2013a) “Marseilleviridae”, a new family of giant viruses infecting amoebae.

Colson P, Fancello L, Gimenez G, Armougom F, Desnues C, Fournous G, Yoosuf N, Million M, La Scola B, Raoult D (2013b) Evidence of the megavirome in humans. J Clin Virol 57:191-200.

Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, Cheng XW, Federici BA, Van Etten JL, Koonin EV, La Scola B, Raoult D (2013c) "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol 158:2517-2521.

Crowther RA, Henderson R, Smith JM (1996) MRC image processing programs. J Struct Biol 116:9-16.

De Castro C, Klose T, Speciale I, Lanzetta R, Molinaro A, Van Etten JL, Rossmann MG (2018) Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches. Proc Natl Acad Sci U S A 115:E44-e52.

DeRosier DJ (2000) Correction of high-resolution data for curvature of the Ewald sphere. Ultramicroscopy 81:83-98.

Doerr A (2016) Single-particle cryo-electron microscopy. Nat Methods 13:23.

Dornas FP, Assis FL, Aherfi S, Arantes T, Abrahão JS, Colson P, La Scola B (2016) A Brazilian Marseillevirus Is the Founding Member of a Lineage in Family Marseilleviridae. Viruses 8:76.

Dos Santos RN, Campos FS, Medeiros de Albuquerque NR, Finoketti F, Côrrea RA, Cano-Ortiz L, Assis FL, Arantes TS, Roehe PM, Franco AC (2016) A new marseillevirus isolated in Southern Brazil from Limnoperna fortunei. Sci Rep 6:35237.

Doutre G, Philippe N, Abergel C, Claverie JM (2014) Genome analysis of the first Marseilleviridae representative from Australia indicates that most of its genes contribute to virus fitness. J Virol 88:14340-14349.

Doutre G, Arfib B, Rochette P, Claverie JM, Bonin P, Abergel C (2015) Complete Genome Sequence of a New Member of the Marseilleviridae Recovered from the Brackish Submarine Spring in the Cassis Port-Miou Calanque, France. Genome Announc 3.

Downing KH, Glaeser RM (2018) Estimating the effect of finite depth of field in single-particle cryo-EM. Ultramicroscopy 184:94-99.

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

Fabre E, Jeudy S, Santini S, Legendre M, Trauchessec M, Couté Y, Claverie JM, Abergel C (2017) Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun 8:15087.

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

Henderson R (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys 28:171-193.

Heymann JB, Belnap DM (2007) Bsoft: image processing and molecular modeling for electron microscopy. J Struct Biol 157:3-18.

Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 75:11720-11734.

Klose T, Reteno DG, Benamar S, Hollerbach A, Colson P, La Scola B, Rossmann MG (2016) Structure of faustovirus, a large dsDNA virus. Proc Natl Acad Sci U S A 113:6206-6211.

Koonin EV, Yutin N (2019) Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. Adv Virus Res 103:167-202.

Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH (2020) Global Organization and Proposed Megataxonomy of the Virus World. Microbiol Mol Biol Rev 84.

Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. J Struct Biol 116:71-76.

Kühlbrandt W (2014) The resolution revolution. Science 343:1443-1444.

La Scola B (2014) Looking at protists as a source of pathogenic viruses. Microb Pathog 77:131-135.

La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie JM, Raoult D (2003) A giant virus in amoebae. Science 299:2033.

Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, Trape JF, Koonin EV, La Scola B, Raoult D (2012) Microbial

culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185-1193.

Liu S, Luo Y, Wang Y, Li S, Zhao Z, Bi Y, Sun J, Peng R, Song H, Zhu D, Sun Y, Zhang L, Wang W, Qi J, Yan J, Shi Y, Zhang X, Wang P, Qiu HJ, Gao GF (2019) Cryo-EM Structure of the African Swine Fever Virus. Cell Host Microbe 26:836-843.e833.

Macera L, Spezia PG, Focosi D, Mazzetti P, Antonelli G, Pistello M, Maggi F (2020) Lack of Marseillevirus DNA in immunocompetent and immunocompromised Italian patients. J Med Virol 92:187-190.

Milrot E, Shimoni E, Dadosh T, Rechav K, Unger T, Van Etten JL, Minsky A (2017) Structural studies demonstrating a bacteriophage-like replication cycle of the eukaryote-infecting Paramecium bursaria chlorella virus-1. PLoS Pathog 13:e1006562.

Murata K, Kaneko Y (2018) Visualization of DNA Compaction in Cyanobacteria by High-voltage Cryo-electron Tomography. J Vis Exp.

Murata K, Hagiwara S, Kimori Y, Kaneko Y (2016) Ultrastructure of compacted DNA in cyanobacteria by high-voltage cryo-electron tomography. Sci Rep 6:34934.

Okamoto K, Miyazaki N, Song C, Maia F, Reddy HKN, Abergel C, Claverie JM, Hajdu J, Svenda M, Murata K (2017) Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. Sci Rep 7:13291.

Okamoto K, Miyazaki N, Reddy HKN, Hantke MF, Maia F, Larsson DSD, Abergel C, Claverie JM, Hajdu J, Murata K, Svenda M (2018) Cryo-EM structure of a Marseilleviridae virus particle reveals a large internal microassembly. Virology 516:239-245.

Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295-2300.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612.

Pintilie G, Chiu W (2012) Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. Biopolymers 97:742-760.

Pintilie G, Chen DH, Tran BN, Jakana J, Wu J, Hew CL, Chiu W (2019) Segmentation and Comparative Modeling in an 8.6-Å Cryo-EM Map of the Singapore Grouper Iridovirus. Structure 27:1561-1569.e1564.

Popgeorgiev N, Boyer M, Fancello L, Monteil S, Robert C, Rivet R, Nappez C, Azza S, Chiaroni J, Raoult D, Desnues C (2013) Marseillevirus-like virus recovered from blood donated by asymptomatic humans. J Infect Dis 208:1042-1050.

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

Rossmann MG, Johnson JE (1989) Icosahedral RNA virus structure. Annu Rev Biochem 58:533-573.

Rowbotham TJ (1980) Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. J Clin Pathol 33:1179-1183.

Sauvage V, Livartowski A, Boizeau L, Servant-Delmas A, Lionnet F, Lefrère JJ, Laperche S (2014) No evidence of Marseillevirus-like virus presence in blood donors and recipients of multiple blood transfusions. J Infect Dis 210:2017-2018.

Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519-530.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676-682.

Shen PS (2018) The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. Anal Bioanal Chem 410:2053-2057.

Song WJ, Qin QW, Qiu J, Huang CH, Wang F, Hew CL (2004) Functional genomics analysis of Singapore grouper iridovirus: complete sequence determination and proteomic analysis. J Virol 78:12576-12590.

Swann PR, Humphreys CJ, Goringe MJ (1974) High voltage electron microscopy: proceedings of the third International Conference. In. London, New York, Academic Press.

Takemura M (2016) Morphological and Taxonomic Properties of Tokyovirus, the First Marseilleviridae Member Isolated from Japan. Microbes Environ 31:442-448.

Tanaka N, Fujita T, Takahashi Y, Yamasaki J, Murata K, Arai S (2020) Progress in environmental high-voltage transmission electron microscopy for nanomaterials. Philos Trans A Math Phys Eng Sci 378:20190602.

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157:38-46.

Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, Croxatto A, Greub G (2011) Lausannevirus, a giant amoebal virus encoding histone doublets. Environ Microbiol 13:1454-1466.

Wang N, Zhao D, Wang J, Zhang Y, Wang M, Gao Y, Li F, Bu Z, Rao Z, Wang X (2019) Architecture of African swine fever virus and implications for viral assembly. Science 366:640-644.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296-w303.

Wrigley NG (1969) An electron microscope study of the structure of Sericesthis iridescent virus. J Gen Virol 5:123-134.

Xian Y, Avila R, Pant A, Yang Z, Xiao C (2021) The Role of Tape Measure Protein in Nucleocytoplasmic Large DNA Virus Capsid Assembly. Viral Immunol 34:41-48.

Xiao C, Fischer MG, Bolotaulo DM, Ulloa-Rondeau N, Avila GA, Suttle CA (2017) Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel assembly pathway for giant viruses. Sci Rep 7:5484.

Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG (2009) Structural studies of the giant mimivirus. PLoS Biol 7:e92.

Yan X, Yu Z, Zhang P, Battisti AJ, Holdaway HA, Chipman PR, Bajaj C, Bergoin M, Rossmann MG, Baker TS (2009) The capsid proteins of a large, icosahedral dsDNA virus. J Mol Biol 385:1287-1299.

Yanai-Balser GM, Duncan GA, Eudy JD, Wang D, Li X, Agarkova IV, Dunigan DD, Van Etten JL (2010) Microarray analysis of Paramecium bursaria chlorella virus 1 transcription. J Virol 84:532-542.

Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43:W174-181.

Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A (2008) Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. PLoS Biol 6:e114.

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14:331-332.

Zhu D, Wang X, Fang Q, Van Etten JL, Rossmann MG, Rao Z, Zhang X (2018) Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. Nat Commun 9:1552.

Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7.

Figures

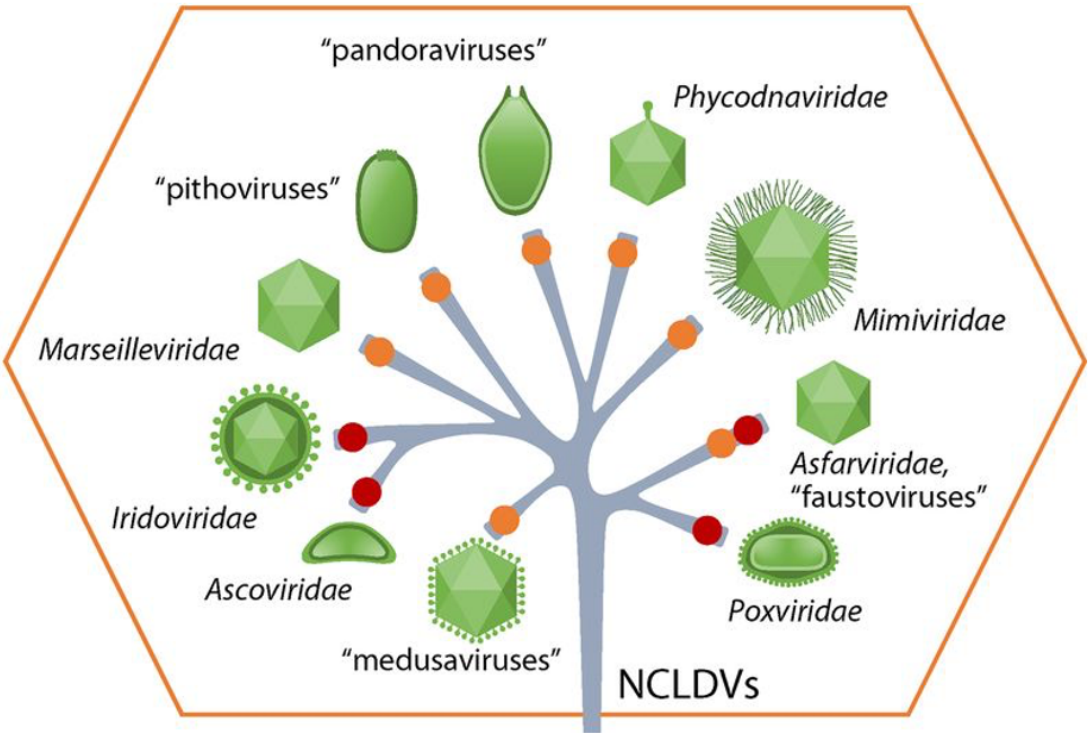

Figure 1 The phylogenetic tree of NCLDVs (Koonin et al., 2020). Illustrations of these virion shapes are included. Most NCLDVs have icosahedral capsids, but some NCLDVs exhibit characteristic shapes. For example, the *poxviridae* are brick-shaped, while pithovirus and pandoravirus are amphora-shaped. The *Ascoviridae* are bacilliform, ovoid, and allantoid.

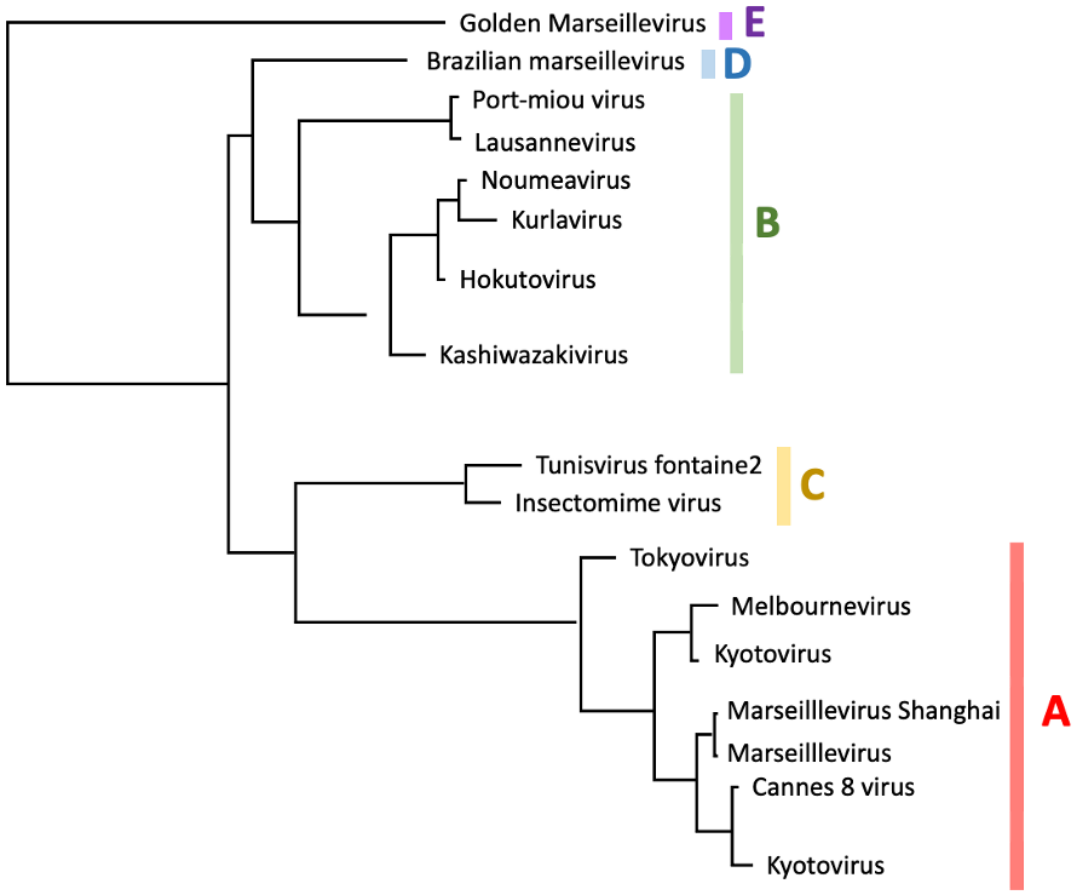

Figure 2 The phylogenetic tree of the family *Marseilleviridae*. The tree was based on MCP gene nucleotide alignment derived from the full-length alignment. The color bars indicate lineage A to E.

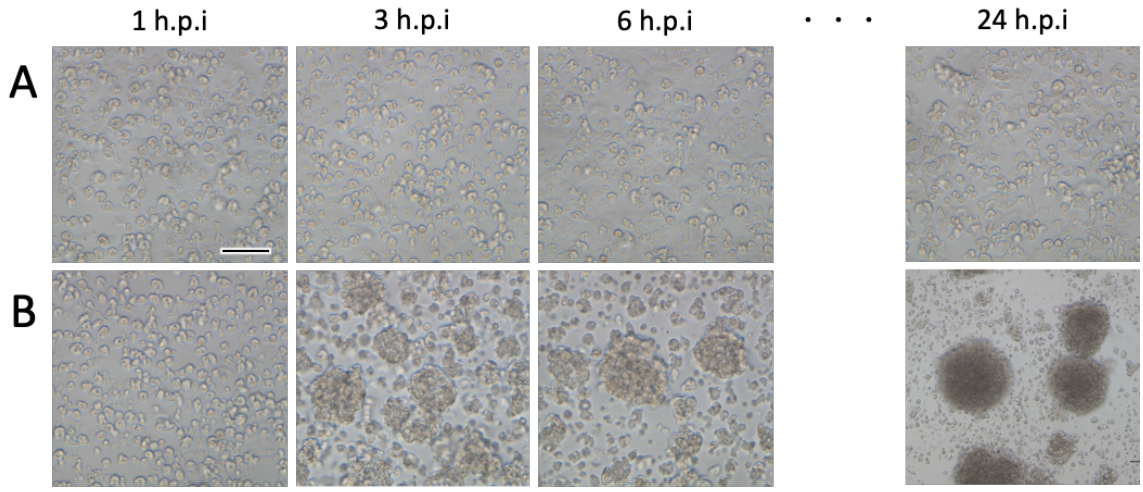

Figure 3 Light microscopic observation of bunch formation in *Acanthamoeba castellanii* cells. A) The amoeba cells infected with tokyovirus belonging to lineage A didn't show a bunch formation (1, 3, 6, 24 hpi). B) The amoeba cells infected with hokutovirus belonging to lineage B showed a bunch formation at 3 hpi. Scale bar = 100 μm.

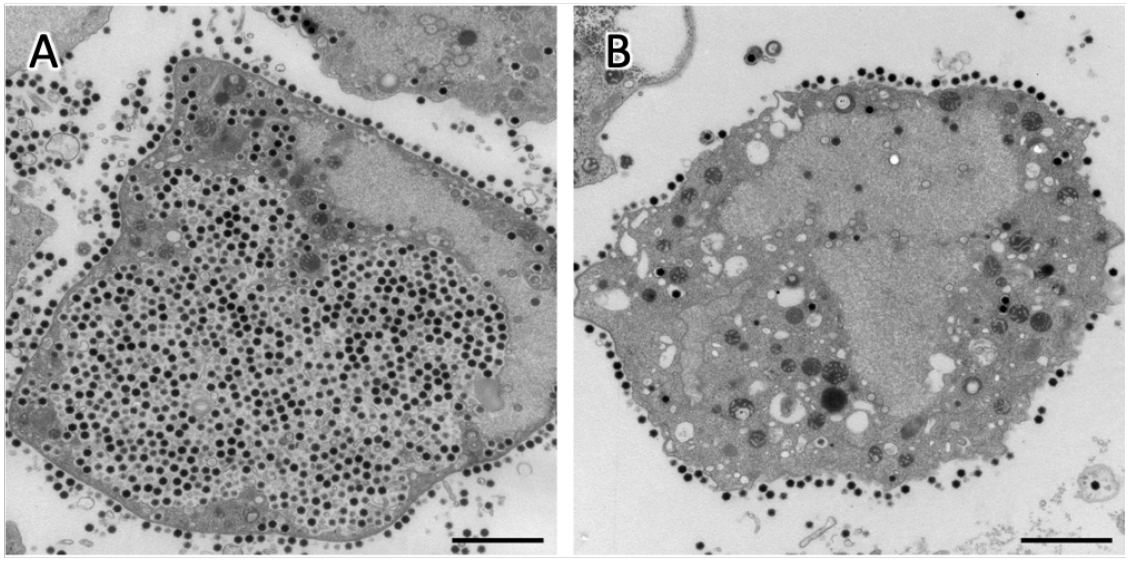

Figure 4 TEM images of ultrathin sections of the amoeba cells infected with hokutovirus (A) and kashiwazakivirus (B) (Aoki et al., 2019). Both show that the progeny viruses coat the host cell surface. Scale bar = 2μm.

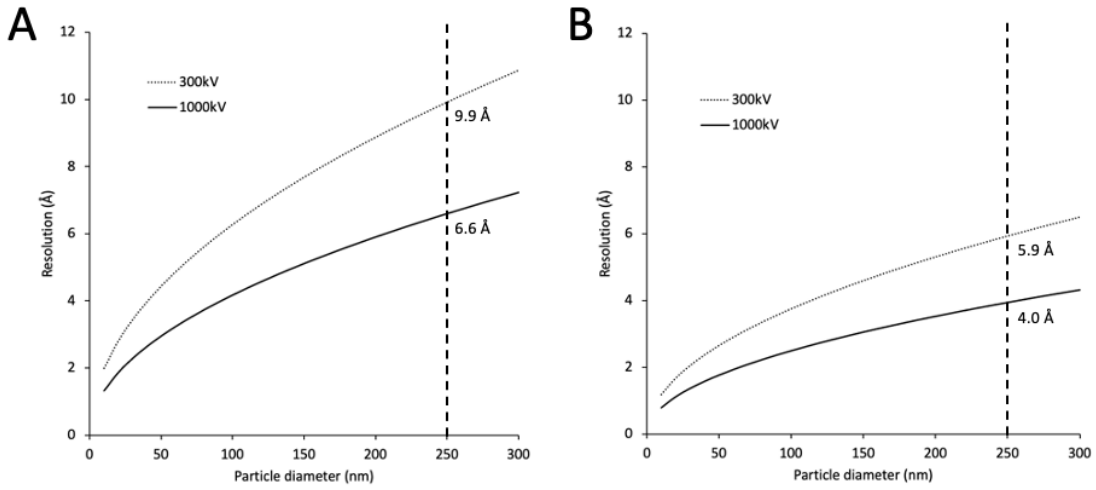

Figure 5 Depth of field effect in TEM. A) Theoretical resolution limits assuming a uniform focus (Eq. 1) are plotted against particle (object) size at accelerating voltages of 300 kV and 1,000 kV, respectively. B) Theoretical resolution limits allowing a phase error (Eq. 2) are plotted against particle (object) size at accelerating voltages of 300 kV and 1,000 kV, respectively. The dashed lines indicate the diameter of tokyovirus (250 nm), showing the resolution limits of tokyovirus theoretically estimated by the two methods at accelerating voltages of 300 kV and 1,000 kV, respectively.

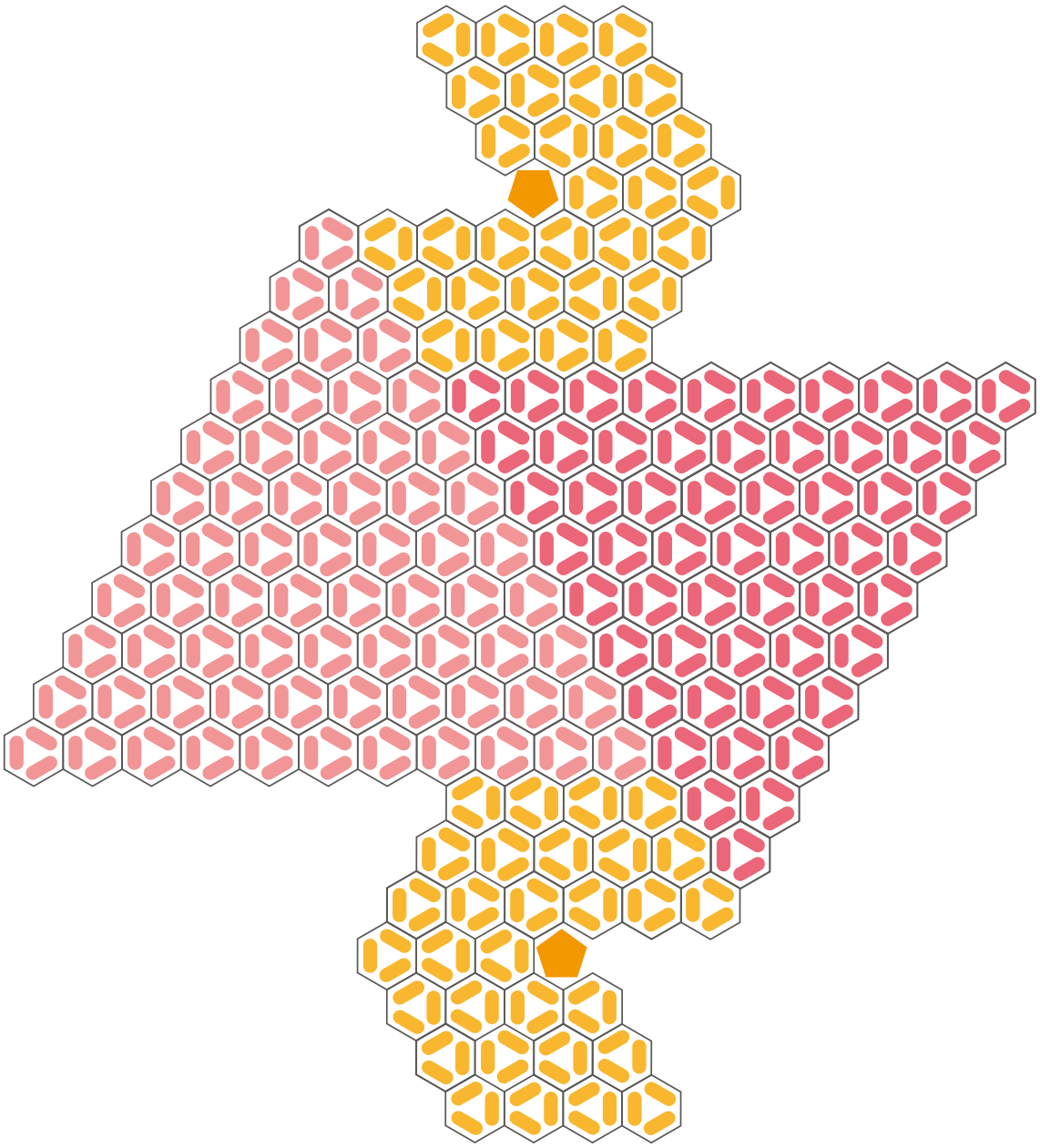

Figure 6 The schematic diagram of tri-, and pentasymmetrons. Hexagons indicate the MCP trimer. The MCP trimer array forms a triangular sheet (called trisymmetron colored pink) and a pentagonal vertex (called pentasymmetron colored yellow). These tri- and pentasymmetrons are commonly used to cover the icosahedral giant virus capsids. The size of the trisymmetron varies with the particle size.

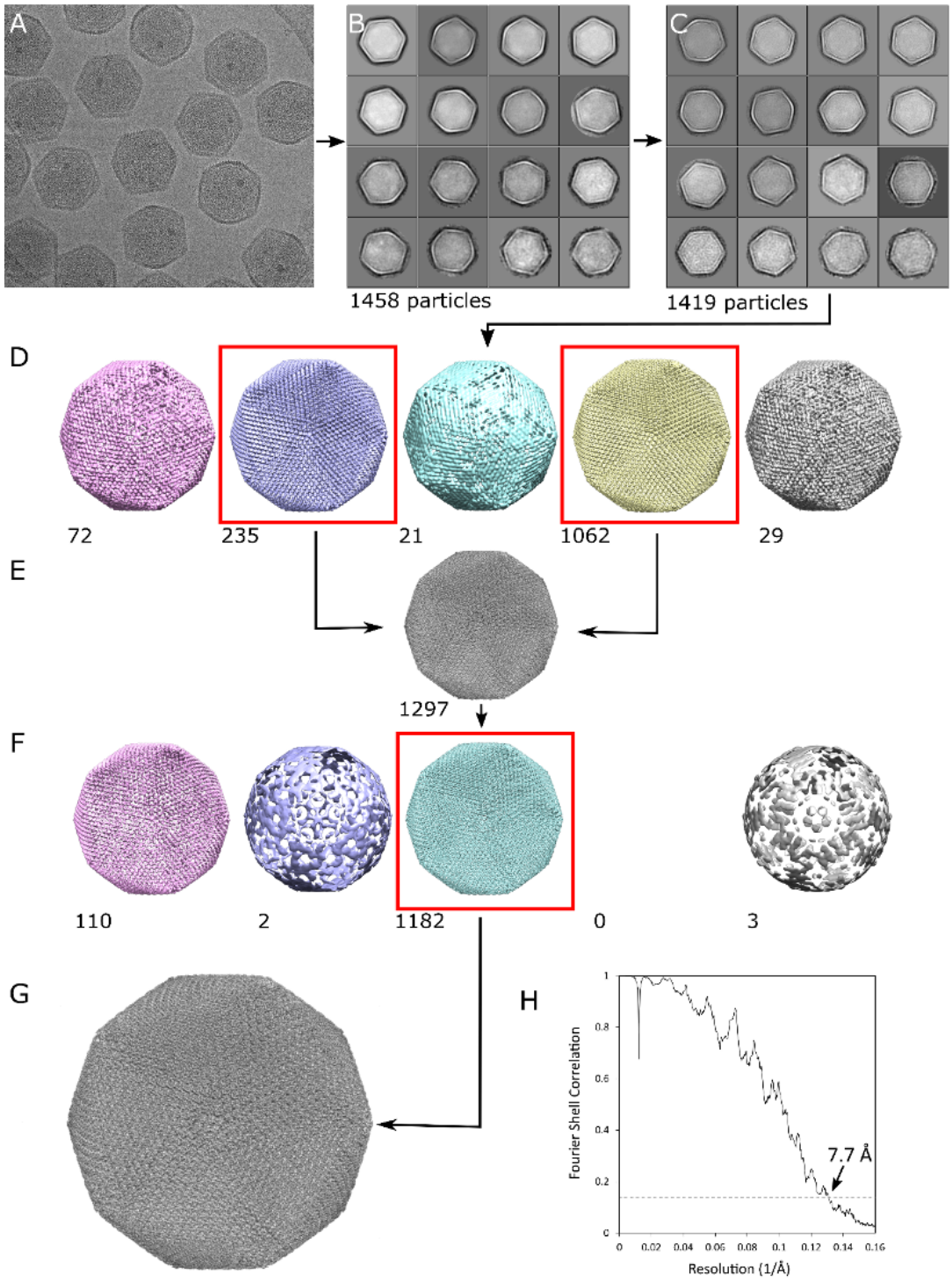
A
B
C
1458 particles
1419 particles
D
72
235
21
1062
29
E
1297
F
110
2
1182
0
3
G
H
Fourier Shell Correlation
7.7 Å
Resolution (1/Å)

Figure 7 The work flow of the 1,000 kV cryo-HVEM SPA of tokyovirus using RELION 3.1 (Zivanov et al., 2018). A) A representative micrograph. B) First round of 2D classification. C) Second round of 2D classification. D) 3D classification into five classes with the clearest two chosen for initial 3D refinement. E) Initial 3D refinement. F) After CTF refinement cycles, the reconstruction was again classified into five classes, with the highest resolution one chosen for final 3D refinement. G) Final 3D reconstruction. Numbers by a 3D class indicate particle count. Red boxes indicate classes brought forward through processing. H) Gold standard FSC (Scheres, 2012) of the final 3D reconstruction.

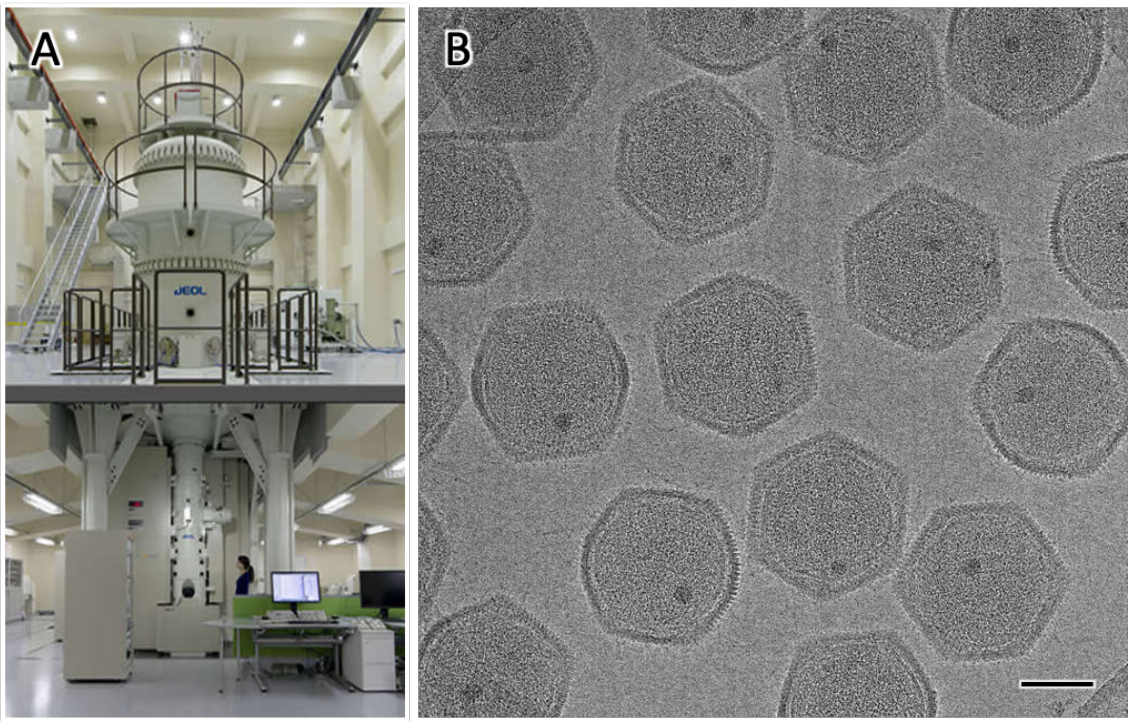

Figure 8 1,000 kV high voltage electron microscopy. A) The 1,000 kV cryo-high voltage electron microscope (cryo-HVEM) (JEM-1000EES, JEOL) installed in Osaka University (http://www.uhvem.osaka-u.ac.jp/jp/feature.html) B) A raw image of tokyovirus collected by 1,000 kV cryo-HVEM.

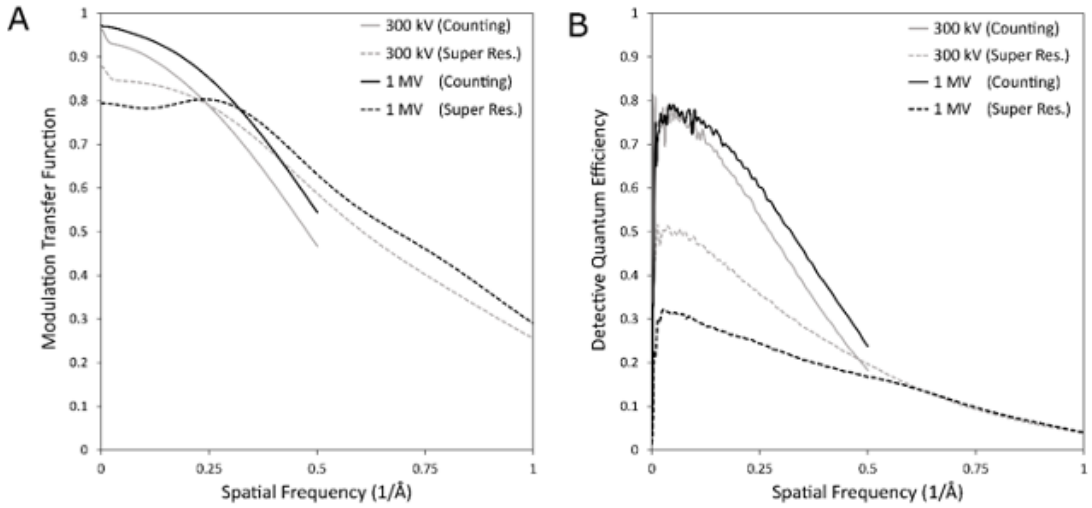

Figure 9 Performance of the detector depending on accelerating voltage. A) Modulation transfer function (MTF) curves for the 1,000 kV cryo-HVEM (JEOL JEM-1000EES) equipped with a K2 Summit direct electron detector (Gatan Inc.) using electron counting and super resolution modes (solid and dashed lines, respectively) (Chihara et al., 2022). Data of a 300 kV microscope are included for comparison, using a Titan Krios G2 (Thermo Fisher Scientific) equipped with a K2 Summit direct electron detector (Gatan Inc.). B) Detective quantum efficiency (DQE) curves for the same conditions as (A).

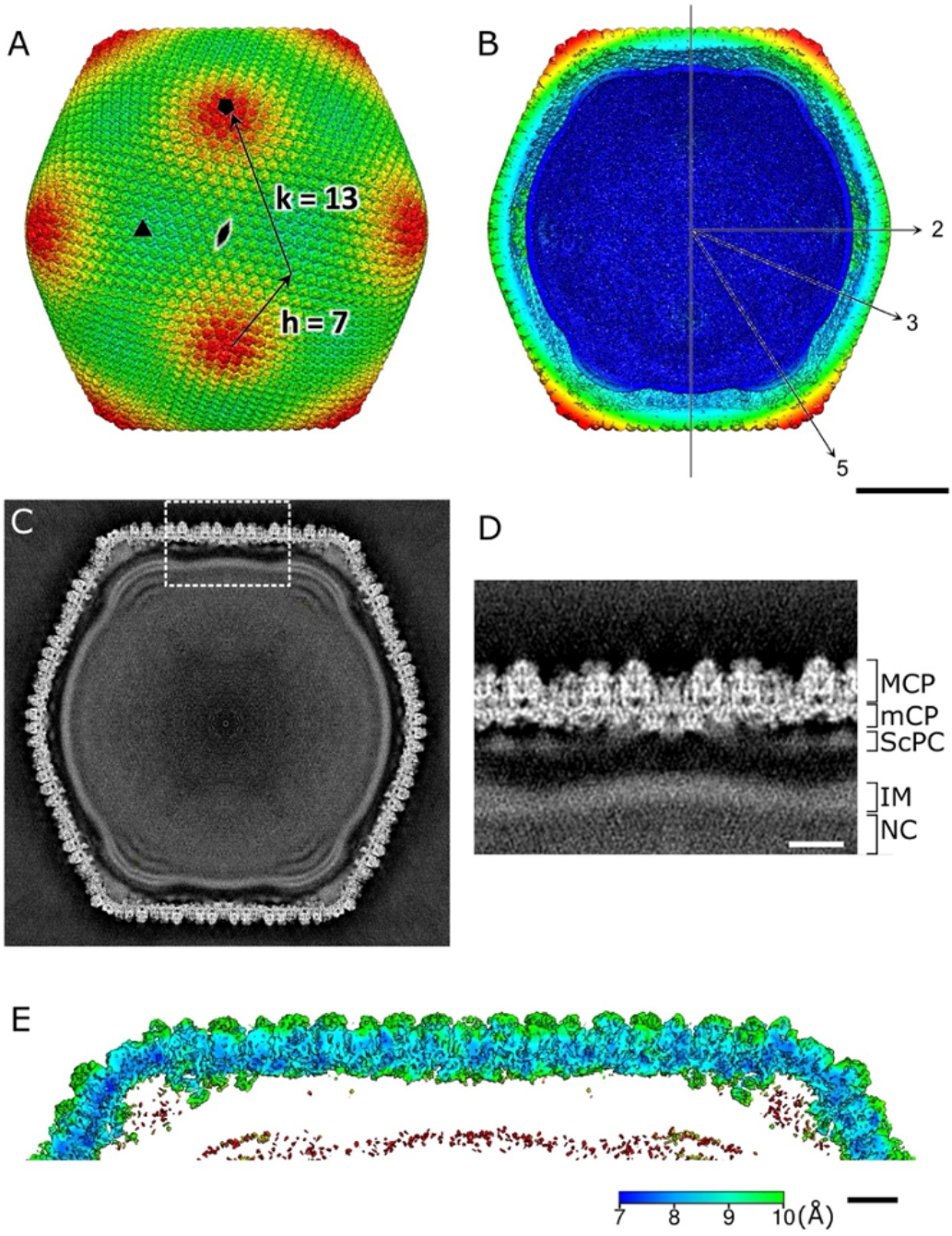
A
k = 13
h = 7
B
2
3
5
C
D
MCP
mCP
ScPC
IM
NC
E
7
8
9
10(Å)

Figure 10 A 3D reconstruction of tokyovirus at 7.7 Å resolution. A) An isosurface view, coloured by radius, and showing the 5-fold (black pentagon), 3-fold (black triangle) and 2-fold (black double-teardrop) symmetry axes. h, k indexes indicating the T=309 icosahedron are included. B) A cross-section view, with symmetry axes shown with arrows. (A, B) are colored by radius in UCSF Chimera [ref] with the following parameters: blue, 910 Å; turquoise, 1,010 Å; green, 1080 Å; yellow, 1,125 Å; red, 1200 Å. C) A central slice of the 3D reconstruction. D) Focussed view of the marked box in (C), showing delineation between major capsid protein (MCP), minor capsid protein (mCP), scaffold protein component (ScPC), internal membrane (IM), and nucleocapsid (NC). E) Local resolution of the capsid, estimated by the *blocres* module of Bsoft (Heymann and Belnap, 2007), focussing on the upper capsid edge, sliced to permit visualisation of internal density. Scale bars: A-C 50 nm, D-E 10 nm.

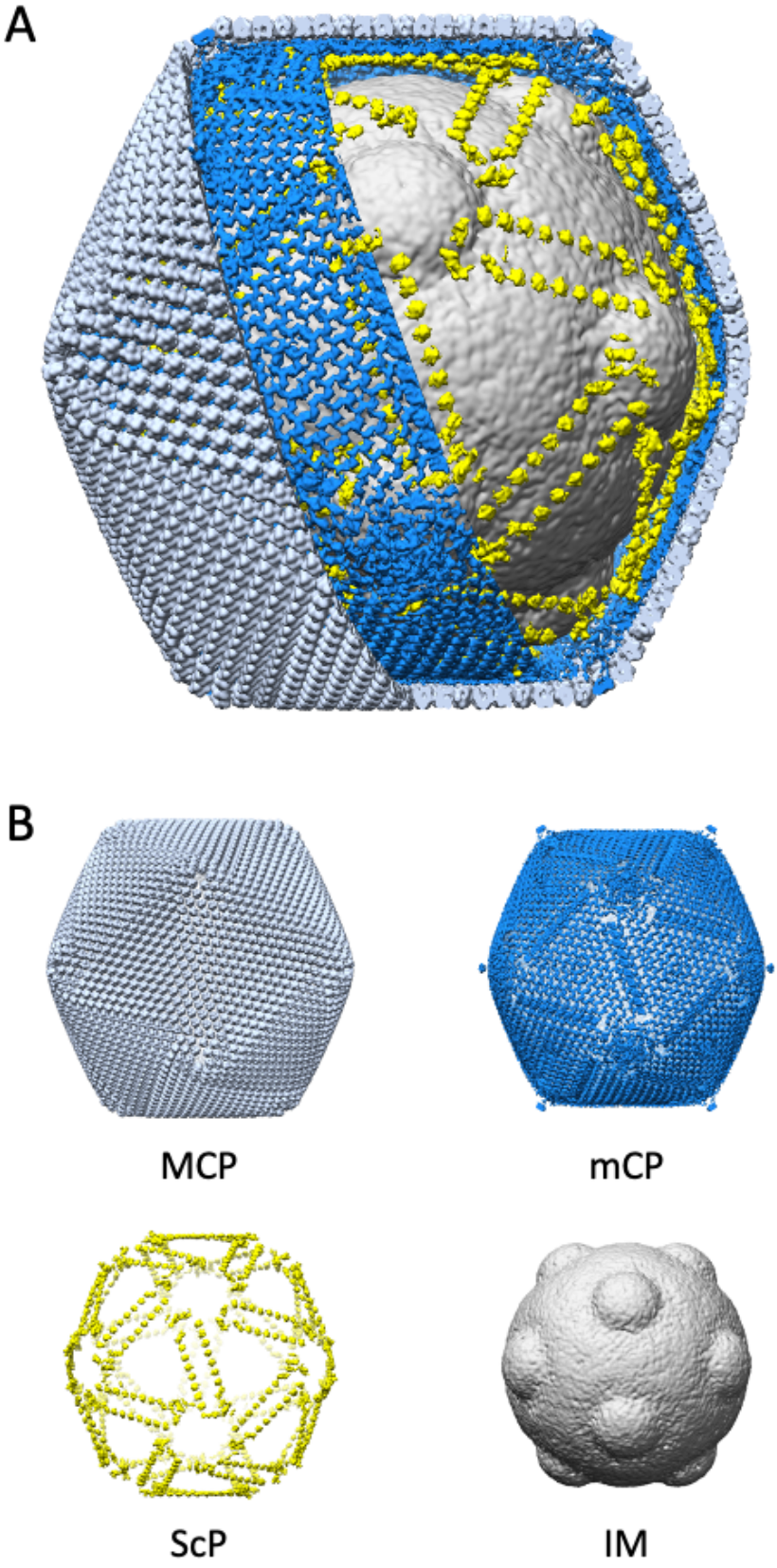
A
B
MCP
mCP
ScP
IM

Figure 11 Segmentation of tokyovirus structure. A) The complete tokyovirus virion cut out to show each component. Individual components of the virion; major capsid protein (MCP) layer (light blue), minor capsid protein (mCP) layer (light blue), scaffold protein component (ScPC) array (yellow), and internal membrane (IM) (grey) are indicated. The map is low-pass-filtered to improve visualisation clarity. B) Each component is displayed separately.

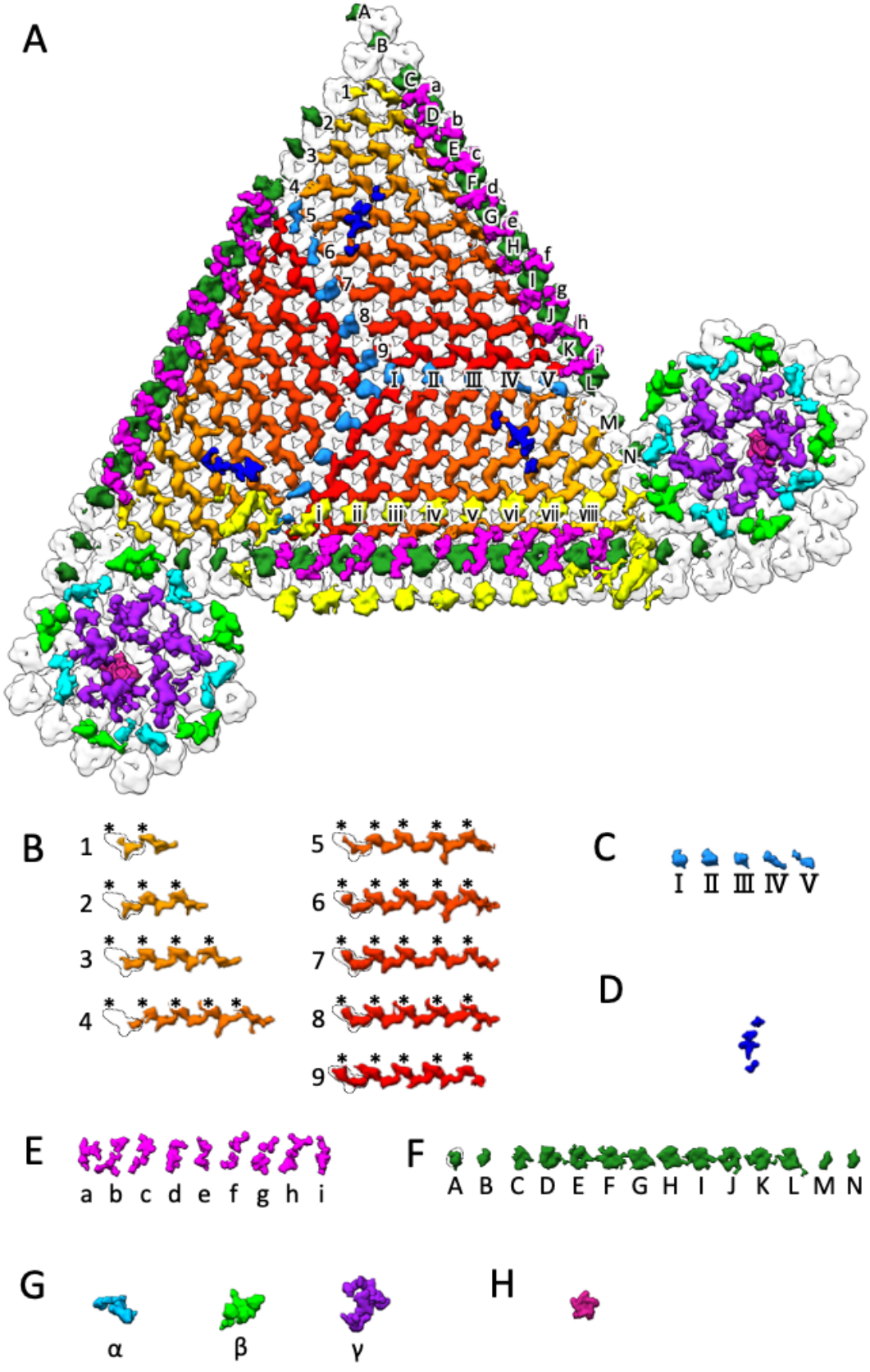
A
B
C
D
E
F
G
H
α
β
γ

Figure 12 Focussed segmentations of the minor capsid protein components and the scaffold protein components. A) A view from inside the tokyovirus capsid. mCPs were classified into 8 components based on the structures, consisting of lattice protein component (LtC) (orange to red), support protein component (SuC) (royal blue), cement protein component (CmC) (sky blue), zipper protein component (ZpC) (magenta), glue protein component (GlC) (green), three pentasymmetron protein components (PC-α, β, and γ) (cyan, light green and purple), and penton (dark pink). Scaffold protein component (ScPC) is indicated by yellow. B) LtC is displayed separately. Each waveform numbered 1 to 9. Asterisk indicates one cycle of the waveform. C) CmC is displayed separately. The number of repetitions indicate Ⅰ to Ⅴ. D) SuC is displayed separately. E) ZpC is displayed separately. The number of repetitions indicate a to i. F) GlC is displayed separately. The number of repetitions indicate A to N. G) PC-α, β, γ is displayed separately. H) the pentagon component is displayed separately.

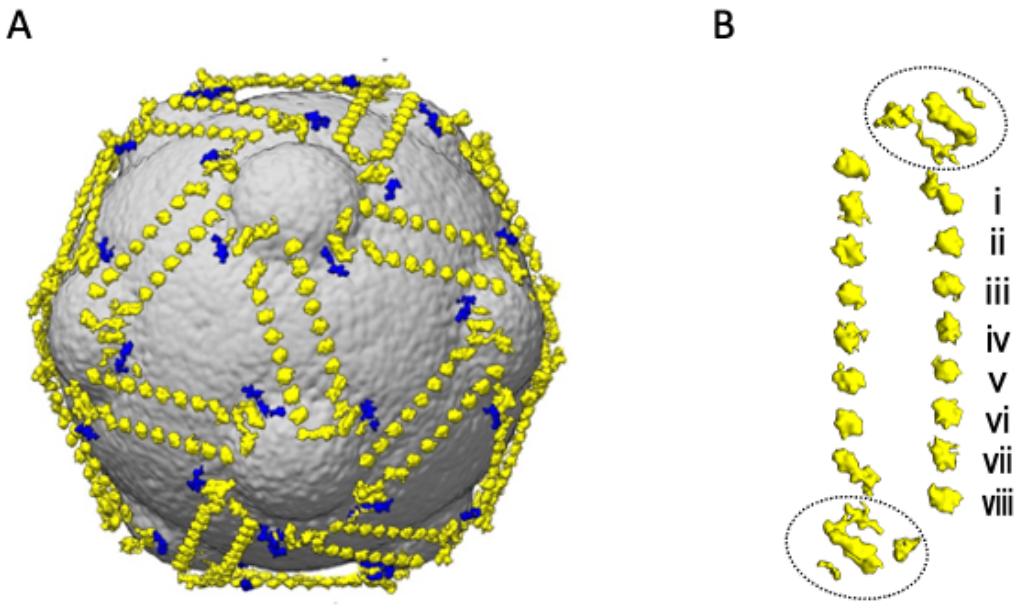

Figure 13 The scaffold protein component network between the internal membrane and capsid shell. A) The internal membrane (grey) is overlaid with the ScPCs (yellow) and SuC (royal blue) network. B) Anti-parallel structure of ScPC. Dotted circles show the large heads. The tail components are numbered i to viii.

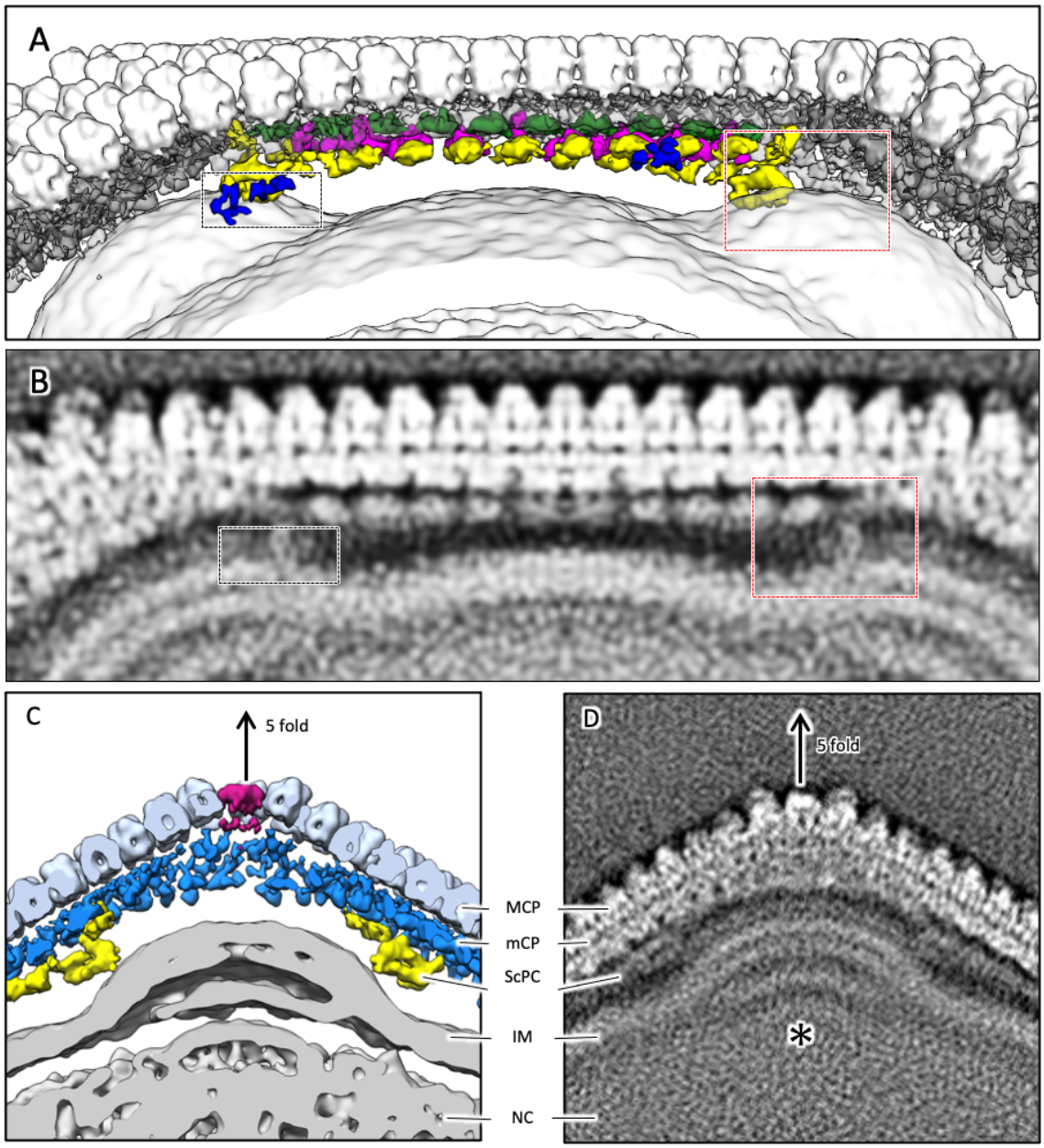
A
B
C
5 fold
MCP
mCP
ScPC
IM
NC
D
5 fold

Figure 14 Interactions between the capsid and the inner membrane. A) Isosurface view of the mCP components colored as in Figure 11 (GlC; green, ZpC; hot pink, ScPC; yellow, SuC; royal blue). The black dotted box indicated that SuPCs and SuCs directly connect with the inner membrane on the membrane exclusions. B) A 3D reconstruction image of (A). C) Isosurface view of the 5-fold axis. MCPs are colored in right blue, mCPs are colored in blue, ScPCs are colored yellow, inner membrane is colored in grey, and penton is colored in dark pink, respectively. D) A 3D reconstruction image of (C). Asterisk indicates a multilayer structure in the inner membrane exclusion.


B1
C1
D1
E1
Tokyovirus 1 MTSVICGTRVTAASGFVDLATFSDLEAYLYGGCSAVTYFVRAIKKANWFSFLPVVLRNISGLPGFGSEFSASVNRSGDYVLNTWLRVRLPLIAIRPTNAGGAINANATIRWTRNFMHNLVEKVNITFND 129
SGIV 1 ----------YPVGWFTKLPTMATRVSGNPAFGQEFSVGVPRSGDYVLNAWLTLKTPEIKLLETNRLG---ANGTVRWTKNLMHNAVEHASLTFND 83
PBCV-1 1 ----------AGGLSQLVAYGAQDVYLTGN-PQITFFKTV-YRRYTNFAIESIQQTINGSVGFGNKVSTQISRNGDLITDIVVEFVLTKG----------GNGGTTYYPAEELLQDVELEIGG 101
Conservation:
F1
G1
H1
I1
Tokyovirus 130 LIVHEFDSYWFDFNSQFNIDASKRVGYRNMIGDIPAMINPVTTGNPLG----TGEFFNLPIPLFYTEDSGLALAVSALPFNDIKINYCLRRWQDLIVLNVGVGANPPTFDDIVQVSYDAAFTLIYSSNAPAITNVETWCHYAVV 269
SGIV 84 ICAQQFNTAYLDAWTQFNMCEGKRIGYDNMIGNTSDMTNPTPAQGQDGARTLPSKNLVLPLPFFFSRDCGLALPTVVLPYNEIRINIKLRSLQELLVFQNKDTGNVIPI----------SATDIAGGLADTVEAYVYMTVGLV 216
PBCV-1 102 QRIDKHYNDWFRTYDALFRMNDDRYNYRRMTDWVN--------NELVG----AQKRFYVPLIFFFNQTPGLALPLIALQYHEVKLYFTLASQVQGVNYNGSSAIA----------GAAQPTMSVWVDYIFL 210
Conservation:
B2
C2
D2
E2
Tokyovirus 270 HNDERVKMGKNPRDMVIKQVQKVNETTINLSQL-NALVPIDIRVSHAVVGYFYAIRNSSTPGEWSNYTTEPAYA----------GLDPLEAAQLVYESTARVS- 361
SGIV 217 SNVERCAMAGTVRDMVVEQMQAAPTHIVNPQNT--NNVHVDMRFSHAVKALFFMVQNVTYKSVGSNYTCVTPVNGPG-------NTVMEP--------AMSVDPIKSASLTYENTTRLAN 318
PBCV-1 211 DTQERTRFAQLPHEYLIEQLQFTGSETATPSATTQASQNIRLNFNHPTKYLAWNFNNPTN---YGQYTALANIPGACSGAGTAAATVTTPDYGNTGTYNEQLAVLDSAKIQLNGQDRFAT 327
Conservation:
F2
G2
H2
I2
Tokyovirus 362 NGSDYYSLMVPWFWHKSIPEETGYHAYSYSLDTFASDPKGSTNYSKLTNVSNQYVPSVAAVNASAGVTNTGIPIPSATNPAVLQQNQTFQHIFRVLNFNVLRLSGGSLGLPVL- 474
SGIV 319 MGVEYYSLVQPWYFSASIPVYTGYHMYSYALNVGSVHPSGSTNYGRLTNASITVTMSPESVVAAAG---------GGNNNSGYNEPQRFALVVIAVNHNVIRIMNGSMGFPIL- 423
PBCV-1 328 RKGSYFNKVQPYQSIGGVTP-AGVYLYSFALKPAGRQPSGTCNFSRIDNATLSLTYKTCSIDATSPAAVLGN----TETVTANTATLLTALNIYAKNYNVLRIMSGMGGLAYAN 436
Conservation:

Figure 15 Sequence alignment of tokyovirus MCP versus SGIV MCP and PBCV-1 MCP. Putative β-sheets in jelly-roll 1 and jelly-roll 2 were highlighted light blue and pink, respectively. Residues conserved between all three sequences are marked with asterisks.

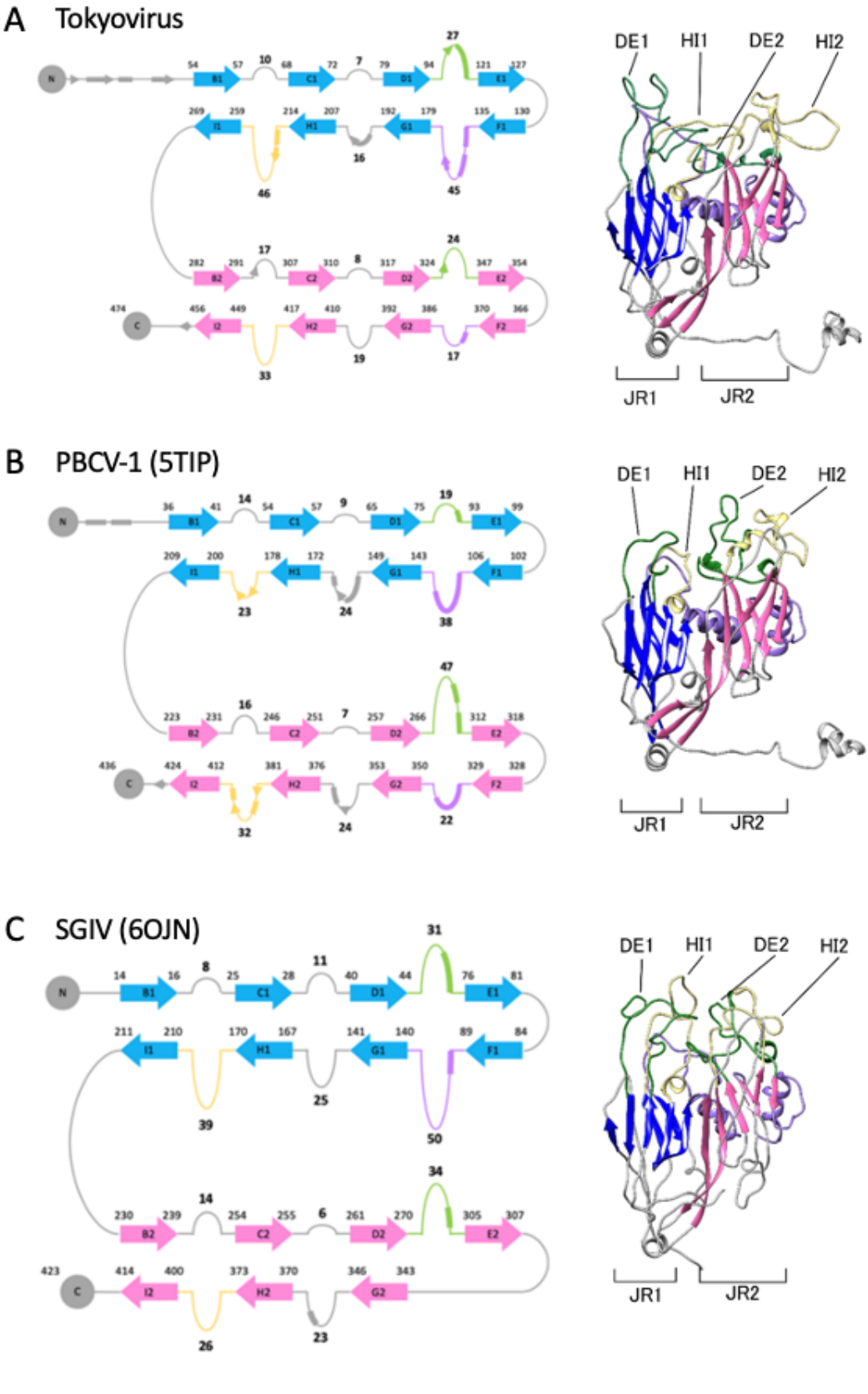
A Tokyovirus
DE1
HI1
DE2
HI2
JR1
JR2
B PBCV-1 (5TIP)
DE1
HI1
DE2
HI2
JR1
JR2
C SGIV (6OJN)
DE1
HI1
DE2
HI2
JR1
JR2

Figure 16 Comparing the secondary structure of MCP monomers. A) A ribbon diagram and model of tokyovirus MCP. Jelly roll motif 1 and 2, which consist of β-strands, are indicated by light blue arrows and pink arrows, respectively. Each β-strand is labelled B1 to I2. The exterior loops (DE, FG, HI) of MCP are colored light green, purple, and yellow. α-helices are represented as bars. Corresponding figures of PBCV-1 MCP and SGIV MCP are shown in B and C.

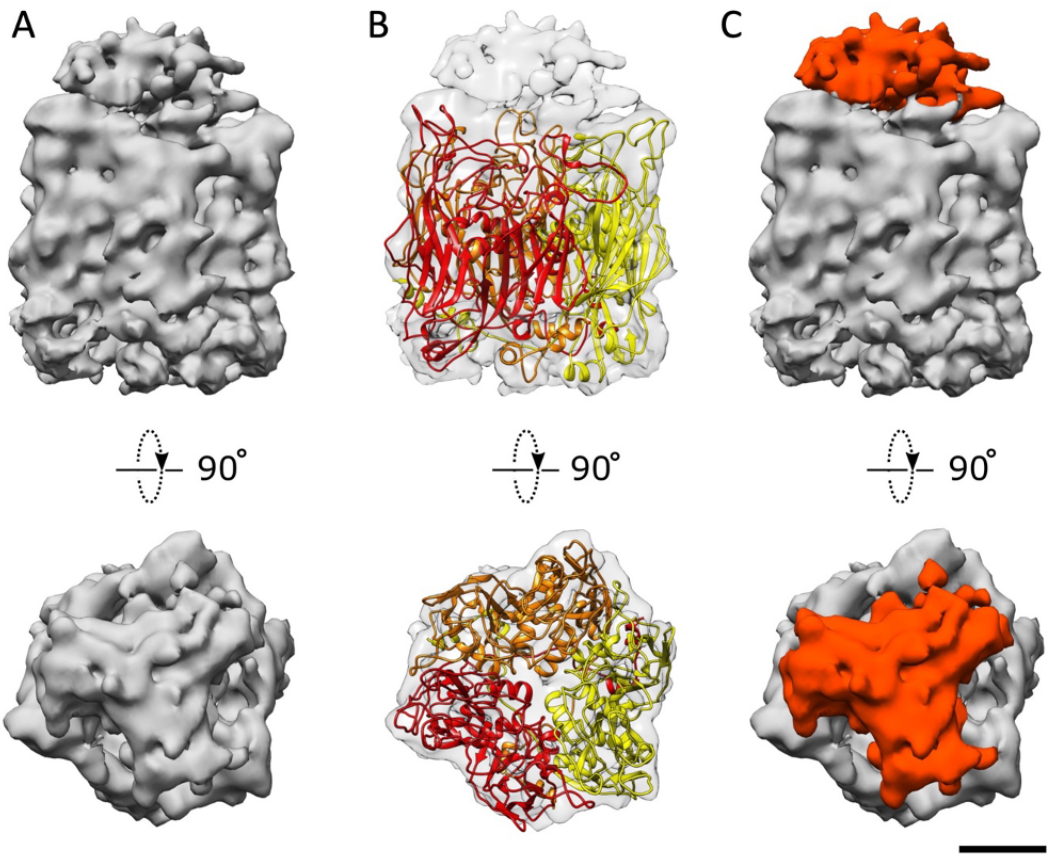

Figure 17 Cryo-EM map and fitted homology model of the tokyovirus MCP trimer. The homology model was generated by SWISS-MODEL (Waterhouse et al., 2018) and adjusted to the volume with COOT (Emsley and Cowtan, 2004) and refined with PHENIX (Adams et al., 2010). A) Side and top views of the extracted cryo-EM map of the MCP trimer. B) Side and top views of the extracted MCP trimer cryo-EM map with the homology model fitted. C) Side and top views of the extracted MCP cryo-EM map of the MCP trimer with the additional cap region is colored in red. Scale bar equals 2 nm.

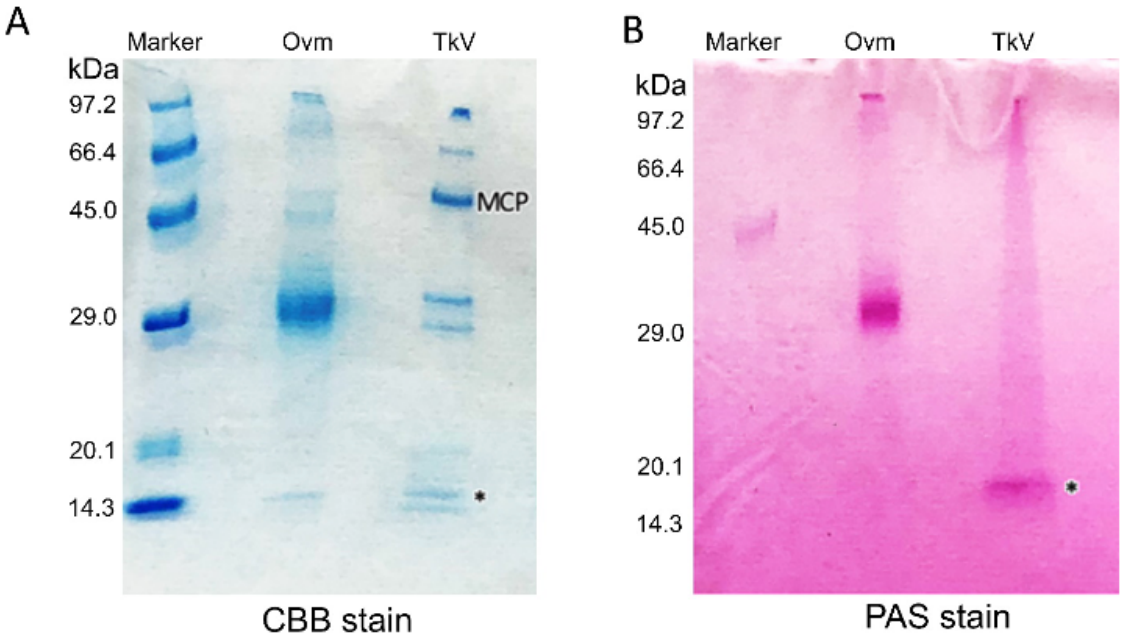

Figure 18 Identification of potential glycoproteins by Periodic acid Schiff (PAS)(Aterman and Norkin, 1963) staining. A) SDS-PAGE gel of a positive control of albumin (Ovm), and tokyovirus (TkV). A strong band for TkV MCP was identified at ~52 kDa. The gel was stained with Coomassie Brilliant Blue (CBB). B) PAS-stained gel with identical loading to that of (A). In the PAS-stained gel, no signal was detected at the same MW as the MCP, while a single band was identified for TkV at ~14 kDa (marked with asterisks on both gels). The molecular weights of each marker component are indicated on the left of each gel.

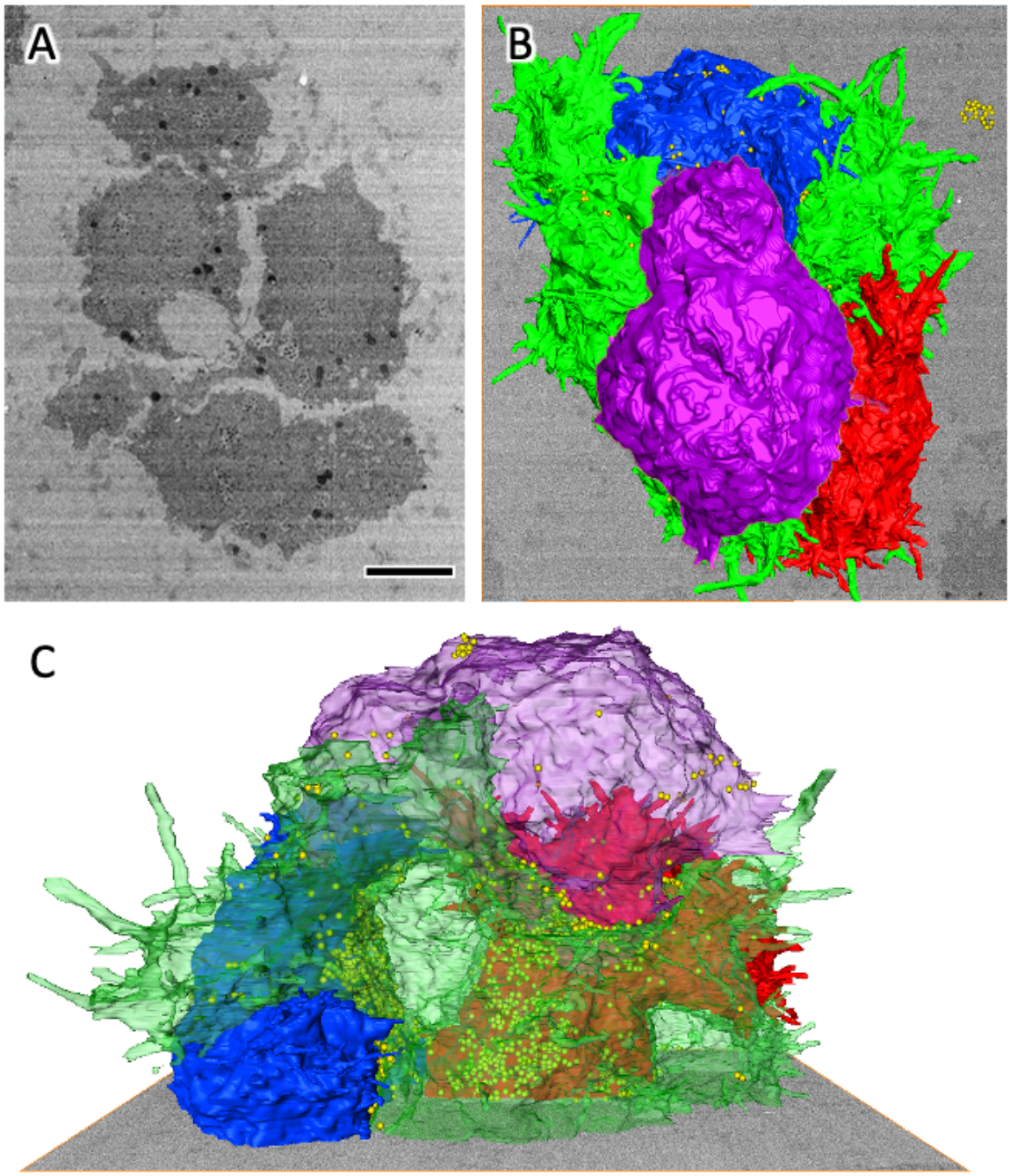
A
B
C

Figure 19 Three-dimensional representation of a bunched amoeba cells infected with hokutovirus. The map was reconstructed from serial-sectioned images by SBF-SEM. A) A slice image of the bunched cells. Scale bar equals 5 μm. B) 3D surface model describing cellular structures of bunched amoeba cells infected with hokutovirus. The cells are colored in red, blue, right green and purple, respectively, and hokutovirus colored yellow. C) Transparent images of (B) viewed from the left side.

Table 1 Cryo-EM data information

Tokyovirus data collection			
Nominal magnification	15,000×		
(corresponds to)	1.456 Å/pixel		
Microscope (accelerating voltage)	JEOL JEM-1000EES (1,000 kV)		
Spherical aberration (mm)	4.1		
Detector	Gatan K2 Summit		
Detector mode (dimensions in pixels)	Super resolution (7,420 × 6,780)		
Total exposure ($e^-/Å^2$)	~35		
Defocus target range	-1.5 – -4 μm		
Total frames per micrograph	160		
Total acquisition time per micrograph	32 seconds		
Cryo-HVEM processing			
Total selected particles	1,529		
Final particles used	1,182		
Pixel size	2.912 Å/pixel		
Symmetry imposed	I3		
Mask used	Capsid	Capsid+IM	Spherical
Map resolution (Å)	7.7	8.7	9.4
Map sharpening imposed ($Å^2$)	-100	-100	-100
EMDB accession code	30797	30798	30799

Table 2 Results of BLAST comparison of tokyovirus MCP versus other NCLDVs.

Virus	PDB ID	a.a.	BLAST search			
			E-value	Score	Query Cover (%)	Identities (%)
Iridovirus	6OJN	463	1E-105	310	99	38.21
PBCV-1	5TIP	436	5E-17	68.9	96	19.92
Faustovirus	5J7O	645	0.016	23.9	30	37.5
Phage PM2	2VVF	269	0.043	21.2	14	26.67
ASFV	6KU9	693	0.069	21.9	38	12
Sputnik virus	3J26	508	0.48	18.9	27	36
PRD-1	1CJD	394	0.93	17.3	16	37.04
Vaccinia virus	2YGB	569	2.4	16.5	14	29.63
Phage FLiP	5OAC	310	2.7	15.4	15	42.86
Adenovirus	6B1T	952	3	16.9	16	21.43
STIV	2BBD	350	4.4	15	13	22.73
Mavirus	6G45	610	4.8	15.8	12	25.93