Molecular architecture of the SARS-CoV-2 virus

Hangping Yao[1,2#], Yutong Song[3,4#], Yong Chen[3,4#], Nanping Wu[1,2#], Jialu Xu[3,4,5#], Chujie Sun[3,4,5], Jiaxing Zhang[3,4], Tianhao Weng[1,2], Zheyuan Zhang[3,4], Zhigang Wu[1,2], Linfang Cheng[1,2], Danrong Shi[1,2], Xiangyun Lu[1,2], Jianlin Lei[3,4], Max Crispin[6], Yigong Shi[3,4,5,7,8], Lanjuan Li[1,2*] and Sai Li[3,4,5*]

[1]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases and [2]National Clinical Research Center for Infectious Diseases, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China;

[3]School of Life Sciences, Tsinghua University, [4]Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure and [5]Tsinghua University-Peking University Joint Center for Life Sciences, Beijing 100084, China;

[6]School of Biological Sciences, University of Southampton, Southampton SO17 1BJ, UK;

[7]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang, China;

[8]Institute of Biology, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China.

[#]These authors contributed equally to this work.

[*]To whom correspondence should be addressed: S. Li (sai@tsinghua.edu.cn) or L. Li (ljli@zju.edu.cn)

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped virus responsible for the COVID-19 pandemic. Despite recent advances in the structural elucidation of SARS-CoV-2 proteins and the complexes of the spike (S) proteins with the cellular receptor ACE2 or neutralizing antibodies, detailed architecture of the intact virus remains to be unveiled. Here we report the molecular assembly of the authentic SARS-CoV-2 virus using cryo-electron tomography (cryo-ET) and subtomogram averaging (STA). Native structures of the S proteins in both pre- and postfusion conformations were determined to average resolutions of 9-11 Å. Compositions of the N-linked glycans from the native spikes were analyzed by mass spectrometry, which revealed highly similar overall processing states of the native glycans to that of the recombinant glycoprotein glycans. The *in situ* architecture of the ribonucleoproteins (RNP) and its higher-order assemblies were revealed. These characterizations have revealed the architecture of the SARS-CoV-2 virus to an unprecedented resolution, and shed lights on how the virus packs its ~30 Kb long single-segmented RNA in the ~80 nm diameter lumen. Overall, the results unveiled the molecular architecture and assembly of the SARS-CoV-2 in native context.

As of July 5th, 2020, a total of over 11 million cases of COVID-19 were reported and more than half a million lives were claimed globally (https://covid19.who.int). The causative pathogen, SARS-CoV-2, is a novel β-coronavirus (*1-4*). SARS-CoV-2 encodes at least 29 proteins in its (+) RNA genome, four of which are structural proteins: the spike (S), membrane (M), enveloped (E) and nucleocapsid (N) proteins.

The ~600 kDa, trimeric S protein, one of the largest known class-I fusion proteins, is heavily glycosylated with 66 N-linked glycans (*5-7*). Each S protomer comprises the S1 and S2 subunits, and a single transmembrane (TM) segment (*5*). The S protein binds to the cellular surface receptor angiotensin-converting enzyme-2 (ACE2) through the receptor binding domain (RBD), an essential step for membrane fusion (*1, 4, 8-11*). The activation of S requires cleavage of S1/S2 by furin-like protease and undergoes the conformational change from prefusion to postfusion (*12-16*). Several prefusion conformations have been resolved for the S protein, wherein the three RBDs display distinct orientations, "up" or "down" (*5, 7*). The receptor binding sites expose, only when the RBDs adopt an 'up' conformation. Only the "RBD down" and "one RBD up" conformations have been observed in recombinantly expressed S proteins of SARS-CoV-2 (*5, 7*). This is different from SARS-CoV or MERS-CoV, wherein three RBD up conformation have also been observed (*17, 18*). Upon activation, S undergoes dramatic structural rearrangements involving shedding its S1 subunit and inserting its fusion peptide (FP) into the target cell membrane (*19*). The postfusion S protein transforms to needle-shaped, having three helixes entwining coaxially with its five glycans arranged almost equidistantly along the central axis (*19*).

Coronavirus has the largest genome among all RNA viruses. It is enigmatic how the N protein oligomerizes, organizes, and packs the ~30 Kb long single-stranded RNA in the viral lumen. Early negative-staining electron microscopy of coronaviruses showed single-strand helical RNPs with a diameter of ~15 nm (*20*). Cryo-ET of SARS-CoV revealed RNPs organized into lattices underneath the envelope at ~4-5 nm resolution (*21*). However, such ultrastructure was not observed in the Mouse hepatitis virus (MHV), the prototypic β-coronavirus (*22*). No molecular model exists for the coronavirus RNP, and little is known about the architecture, assembly and RNA packaging of the RNPs of other (+) RNA viruses.

To address these questions, we combined cryo-ET and STA for imaging analysis of 2,294 intact virions propagated from an early viral strain (*23*). Here we report the architecture and assembly of the authentic SARS-CoV-2 virus.

Results and Discussion

Molecular landscape of the SARS-CoV-2 virus.

SARS-CoV-2 virions (ID: ZJU_5) were collected on January 22nd, 2020, from a patient with severe symptoms, and were propagated in Vero cells. The patient was infected during a conference with attendees from Wuhan (*23*). Intact virions in (fig. S1A), or isolated from (Fig. 1A) the supernatant were imaged by cryo-EM, both samples showed pleomorphic enveloped particles. We modelled 2,294 virions as ellipsoids by meshing their lipid envelopes, measuring average diameters of 64.8±11.8 nm, 85.9±9.4 nm and 96.6±11.8 nm for the short, medium and long axis of the envelope respectively. S proteins (Fig. 1A, fig. S1) and RNPs (fig. S4) represented the most distinctive features of the virus. A comprehensive analysis of the data revealed an average of 26 RNPs and 26 prefusion S on each virion (Fig. 1C). The spike population per virus is comparable to HIV (*24*), but ~5 times less than the Lassa virus (LASV) (*25*) and 10 times less than the Influenza virus (*26*). Given the crowded packaging inside the virus, the actual number of RNPs per virus was estimated to be 20-30% more than the ones manually identified.

Two conformations of the prefusion S, namely the RBD down conformation and one RBD up conformation from inactivated authentic SARS-CoV-2 virions were classified and reconstructed to 9.5 Å and 10.9 Å resolution by STA, with local resolution reaching 8.6 Å (fig. S3). The paraformaldehyde fixation was proven to have little impact on the structures at this resolution (*25, 27*). The proportion of RBD down conformation among all prefusion spikes was estimated to be 54% per virion (Fig. 1C). The membrane proximal stalk of S represented the poorest resolved region with a local resolution of ~21 Å, showing no trace of the TM or membrane in the structure (Fig. 1D). When the tomogram slices were scrutinized, spike populations of either standing perpendicular to or leaning towards the envelope were observed, suggesting that the TM was averaged out in the map. Interestingly, a third population of Y-shaped spike pairs with combined stems were observed (figs. S1B-C).

Refined orientations of the prefusion S showed they rotate around their stalks almost freely outside the envelope, with the highest possibility of leaning 40° relative to the normal axis of the envelope (Figs. 1B, D). The rotational freedom of spikes is allowed by its low population density, which is prominently distinct from other enveloped viruses possessing class-I fusion

proteins (*24-26*). This unique feature may facilitate the virus engagement with the cellular receptor ACE2, which exists in dimeric form and may bind to two spikes simultaneously (*8*). Although having more freedom in exploring the surrounding environment, the sparsely packed spikes on the viral envelope are also more vulnerable to neutralizing antibodies that bind the otherwise less accessible domains (*28*) or glycan holes (*17*).

Native structures of S in the prefusion conformation.

The native structures of S in both prefusion conformations were compared to the rigidly fitted recombinant protein structures (PDB: 6VXX and 6VYB) (*7*). Five N-linked glycans are visible on both the RBD down and the one RBD up conformations, of which N717 and N1098 were best resolved, and N282, N801 and N1134 were left with larger densities (Fig. 2B, fig. S2A). The difference in glycan density between our native structures and the cryo-EM structures (*5, 7*) can result from the local resolution variation due to the flexible nature of the glycans, the difference in the expression systems or the existence/absence of the stem region and the TM.

We determined the native glycan identity by analyzing the virus sample using mass spectrometry (MS). The viral particles, with or without PNGase digestion, were resolved on SDS-PAGE. After PNGase treatment, the S1 and S2 subunits were reduced by ~30 KDa and ~20 KDa, respectively, in weight (Fig. 2C). The bands corresponding to S1 and S2 before PNGase treatment were analyzed by MS to reveal the glycan compositions at each of the 22 glycosylation sites (Fig. 2D, figs. S2B-C). The overall processing states of the native glycans are highly similar to that of the recombinant glycoprotein glycans (*6*). Populations of under-processed oligomannose-type glycans are found at the same sites as seen in the recombinant material, including at N234 where the glycan is suggested to have a structural role (*29*). However, as is observed at many sites on HIV (*30, 31*), the virus exhibits somewhat lower levels of oligomannose-type glycosylation compared to the recombinant, soluble mimetic. Overall, the presence of substantial population of complex-type glycosylation suggests that the budding route of SARS-CoV-2 into the lumen of endoplasmic reticulum-Golgi intermediate compartments (ERGIC) is not an impediment to glycan maturation and is consistent with both analysis of SARS-CoV glycans (*32*) and the identification of neutralizing antibodies targeting the fucose at the N343 glycan on SARS-CoV-2 (*33*). Furthermore, the lower levels of oligomannose-type glycans compared to HIV and LASV are also consistent with lower glycan density (*34, 35*). The recapitulation of the main features of native viral glycosylation by

soluble, trimeric recombinant S glycoprotein is encouraging for vaccine strategies utilizing recombinant S protein.

Native structure of S in postfusion conformation.

Apart from the triangular prefusion spikes, needle-like densities were occasionally observed on the viral envelope (fig. S1). A 15.3 Å resolution structure was solved and well fitted with PDB: 6M3W, suggesting a postfusion conformation of the S protein. The structure measures 22.5 nm in length, 6 nm in width and stands perpendicular to the viral envelope (Fig. 3D). Five glycan densities (N1098, N1134, N1173 and N1194) are displayed equidistantly on the spike (Fig. 3D), reminiscent to the recombinant postfusion structure (*19*). In comparison to the prefusion conformation, the fixation in orientation to the envelope suggests a dramatic conformational reordering of the stem region to achieve the postfusion conformation. The postfusion S were found only on a small fraction of the viruses (18% of all viruses, each carried an average of 5 postfusion S), hindering us from reaching higher resolutions.

Interestingly, viruses possessing less spikes in total tend to have higher proportion of postfusion spikes (Fig. 3C). It is possible that the virus sparsely packed with spikes are more vulnerable to antibodies, which were reported to trigger detachment of S1 from the SARS-CoV spike and transform it into a postfusion conformation (*17*). Statistics on the refined coordinates (Figs. 3A-C) suggested a tendency of the postfusion S to bundle on the viral surface. One such bundle was reconstructed, showing four postfusion S spaced by ~10 nm on the viral surface (Fig. 3B), compared to ~15 nm average distance between the nearest prefusion S. We speculated that the postfusion S observed on the virus were products of occasional, spontaneous dissociation of S1 (*19*), which was cleaved by host proteinases. They may also come from syncytium naturally formed on infected cells *(36)*, when budding progeny virions carried a few residual postfusion spikes from the cell surface.

Architecture and assembly of the RNPs in intact virions.

It remains enigmatic how coronaviruses pack the ~30 Kb RNA within the ~80 nm diameter viral lumen; if the RNPs are ordered relative to each other to avoid RNA entangling, knotting or even damaging; or if the RNPs are involved in virus assembly. When raw tomogram slices were inspected, densely packed, persimmon-like densities were discernible throughout the virus lumen, some of which appeared to be locally ordered (fig. S4A). Combining previous

cryo-ET observation on coronaviruses (*22*) and SDS-PAGE/MS analysis (Fig. 2), the densities most likely belonged to the RNPs.

In total 18,500 RNPs were picked in the viral lumen, and initially aligned using a sphere as the template and a large spherical mask. A persimmon-like conformation with little structural feature emerged adjacent to the density for lipid bilayer (fig. S4D), suggesting that significant number of RNPs were membrane proximal. Alignment using a tight spherical mask revealed a 13.1 Å resolution reverse G-shaped architecture of the RNP, measuring 15 nm in diameter and 16 nm in height (Fig. 4B). Its shape is in good comparison to the *in situ* structure of the chikungunya viral RNP, which is also positive stranded (*37*), but different from the MHV RNP released using detergent (*38*).

The map was segmented into five head-to-tail reverse L-shaped densities, each fitted with a pair of N protein (N_NTD: 6WKP, N_CTD: 6WJI) dimerized by the N-CTD (*39*) (fig. S5C). We analyzed the electrostatic potential distribution on the surface of the decamer, and suggested a tentative structural model of RNA winded RNP (figs. S5B, D). Interestingly, an early observation on the MHV RNPs showed ~15 nm diameter helices with five subunits per turn (*20*). Due to the limited resolution and little previous structural knowledge about (+) RNA virus RNPs, our model shall be interpreted with caution.

Further 2D classification of the RNPs revealed three classes: 1) closely packed against the envelope, 2) hexagonally and 3) triangularly packed RNPs (Fig. 4A). Following 3D refinement, a membrane proximal, "eggs-in-a-nest" shaped RNP assembly (referred to as the "hexon"), and a membrane-free, tetrahedrally packed assembly (referred to as the "tetrahedron") emerged (Fig. 4C, figs. S4E-G). Projection of the two class averages back onto their refined coordinates revealed that the majority of hexons came from spherical virions, while more tetrahedrons from ellipsoidal virions (Fig. 4D). This was quantified by statistics: ellipsoidal virions tend to pack more RNP tetrahedrons (Fig. 4E). Furthermore, the spacing between two neighboring RNPs (~16 nm) is the same for both the tetrahedron and hexons, and some tetrahedrons could assemble into hexons when projected onto their *in situ* coordinates (fig. S4B).

Our results suggest that the RNP triangle is a key and basic packing unit throughout the virus. The RNPs were reported to form a lattice underneath the envelope (*21*). It remains unanswered if the ultrastructures of RNPs are assembled by RNA, the transmembrane M or E proteins, RNP itself, or multiples of the above.

Solving RNPs to subnanometer resolution was hindered by the crowding of RNPs against each other (figs. S4E-G). Furthermore, structural features of the RNPs on higher order assemblies smeared, possibly due to the symmetry mis-match between individual RNPs and the assembly (Figs. 4C-D). No virus with strictly ordered RNPs throughout the lumen was found by projections. We conclude that the native RNPs are highly heterogeneous, densely packed yet locally ordered in the virus, possibly interacting with the RNA in a beads-on-a-string stoichiometry.

Observation of virus-like particles surrounding SARS-CoV-2 infected cells

Vero cells were infected with SARS-CoV-2 followed by chemical fixation four days post-infection. Cryo-ET captured a significant number of electron dense, envelope-free particles across the cell periphery, or releasing from the cells (Figs. 5A-D). Concavities were observed on the cell membrane near the dense particles. Occasionally, enveloped particles containing RNP densities were seen among the released dense particles. Such particles were extensively observed around the infected Vero cells, as well as frequently seen in purified viral samples (23% of all tomograms) (Figs. 5E-F). The identity of the particles is yet to be characterized to validate whether they are the Defective Interfering (DI) particles of the SARS-CoV-2. The DI particles were found to contain a single-stranded RNA of ~5.2×10^6 in molecular weight (compared to ~5.4×10^6 in standard virus) for MHV (*40*). The minor deletion in genome and massive quantity of the DI particles disguise the authentic viruses from been detected. Our observation has the potential to explain for the asymptomatic infection on the molecular level.

Our work provides molecular insights not only on the structures of spikes in the pre- and postfusion conformations and the RNPs, as well as the glycan compositions in a native context, but also on how they assemble on the authentic virus. We also demonstrated the potential of observing virus activities across the periphery of infected cells directly by cryo-ET.

References and Notes

1. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
2. R. Lu *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565-574 (2020).
3. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
4. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 (2020).
5. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
6. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, eabb9983 (2020).
7. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 (2020).
8. R. Yan *et al.*, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444-1448 (2020).
9. J. Lan *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220 (2020).
10. J. Shang *et al.*, Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020).
11. Q. Wang *et al.*, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904 e899 (2020).
12. G. Simmons, P. Zmora, S. Gierer, A. Heurich, S. Pohlmann, Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. *Antiviral Res.* **100**, 605-614 (2013).
13. S. Belouzard, V. C. Chu, G. R. Whittaker, Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U.S. A.* **106**, 5871-5876 (2009).
14. G. Simmons *et al.*, Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc. Natl. Acad. Sci. U.S. A.* **101**, 4240-4245 (2004).
15. W. Song, M. Gui, X. Wang, Y. Xiang, Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS. Pathog.* **14**, e1007236 (2018).
16. R. N. Kirchdoerfer *et al.*, Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 15701 (2018).
17. A. C. Walls *et al.*, Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **176**, 1026-1039 e1015 (2019).
18. J. Pallesen *et al.*, Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U.S. A.* **114**, E7348-E7357 (2017).
19. Yongfei Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. https://doi.org/10.1101/2020.05.16.099317 (2020).
20. E. O. Caul, S. I. Eggleston, Coronavirus-like particles present in simian faeces. *Vet. Rec.* **104**, 168-169 (1979).
21. B. W. Neuman *et al.*, Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J. Virol.* **80**, 7918-7928 (2006).
22. M. Barcena *et al.*, Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. *Proc. Natl. Acad. Sci. U.S. A.* **106**, 582-587 (2009).
23. Hangping Yao *et al.*, Patient-derived mutations impact pathogenicity of SARS-CoV-2. https://doi.org/10.1101/2020.1104.1114.20060160 (2020).
24. J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, S. Subramaniam, Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109-113 (2008).
25. S. Li *et al.*, Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike. *PLoS. Pathog.* **12**, e1005418 (2016).
26. A. Harris *et al.*, Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U.S. A.* **103**, 19123-19127 (2006).
27. W. Wan *et al.*, Structure and assembly of the Ebola virus nucleocapsid. *Nature* **551**, 394-397 (2017).
28. X. Chi *et al.*, A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science*, eabc6952 (2020).

29. Lorenzo Casalino *et al.*, Shielding and Beyond: The Roles of Glycans in SARS-CoV-2 Spike Protein. https://doi.org/10.1101/2020.1106.1111.146522 (2020).
30. W. B. Struwe *et al.*, Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. *Cell reports* **24**, 1958-1966 e1955 (2018).
31. L. Cao *et al.*, Differential processing of HIV envelope glycans on the virus and soluble recombinant trimer. *Nat. Commun.* **9**, 3693 (2018).
32. G. Ritchie *et al.*, Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology* **399**, 257-269 (2010).
33. D. Pinto *et al.*, Structural and functional analysis of a potent sarbecovirus neutralizing antibody. doi: https://doi.org/10.1101/2020.04.07.023903 (2020).
34. Y. Watanabe *et al.*, Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688 (2020).
35. Y. Watanabe *et al.*, Structure of the Lassa virus glycan shield provides a model for immunological resistance. *Proc. Natl. Acad. Sci. U.S. A.* **115**, 7320-7325 (2018).
36. S. Xia *et al.*, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **30**, 343-355 (2020).
37. J. Jin *et al.*, Neutralizing Antibodies Inhibit Chikungunya Virus Budding at the Plasma Membrane. *Cell Host Microbe* **24**, 417-428.e415 (2018).
38. M. Gui *et al.*, Electron microscopy studies of the coronavirus ribonucleoprotein complex. *Protein Cell* **8**, 219-224 (2017).
39. C. Y. Chen *et al.*, Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **368**, 1075-1086 (2007).
40. S. Makino, K. Fujiwara, M. M. Lai, Defective interfering particles of coronavirus. *Adv. Exp. Med. Biol.* **218**, 187-195 (1987).
41. B. W. Neuman, B. D. Adair, M. Yeager, M. J. Buchmeier, Purification and electron cryomicroscopy of coronavirus particles. *Methods Mol. Biol.* **454**, 129-136 (2008).
42. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191-198 (2017).
43. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36-51 (2005).
44. X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584-590 (2013).
45. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
46. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12 (2016).
47. B. Turonova, F. K. M. Schur, W. Wan, J. A. G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A. *J. Struct. Biol.* **199**, 187-195 (2017).
48. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol* **116**, 71-76 (1996).
49. D. Castano-Diez, M. Kudryashev, H. Stahlberg, Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *J. Struct. Biol.* **197**, 135-144 (2016).
50. D. Castano-Diez, M. Kudryashev, M. Arheit, H. Stahlberg, 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 (2012).
51. S. Li *et al.*, A Molecular-Level Account of the Antigenic Hantaviral Surface. *Cell reports* **15**, 959-967 (2016).
52. T. A. M. Bharat, S. H. W. Scheres, Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 9-20 (2016).
53. M. Chen *et al.*, Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983-985 (2017).
54. J. T. Huiskonen *et al.*, Averaging of Viral Envelope Glycoprotein Spikes from Electron Cryotomography Reconstructions using Jsubtomo. *JOVE-J. VIS. EXP.*, e51714 (2014).
55. E. F. Pettersen *et al.*, UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
56. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14-25 (2018).

57. C. K. Chang *et al.*, Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J. Virol.* **83**, 2255-2264 (2009).
58. G. D. Pintilie, J. Zhang, T. D. Goddard, W. Chiu, D. C. Gossard, Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427-438 (2010).

Acknowledgments: We thank Tsinghua University for providing a Start-up fund, the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for the cryo-EM facility and the computational facility support, and Dr. F. Yang, J. Wen and D. Li for technical support. We thank Dr. H. Deng and C. Zhao in the Proteinomics Facility at Technology Center for Protein Sciences, Tsinghua University, for protein MS analysis. We thank the computational facility support on the cluster of Bio-Computing Platform (Tsinghua University Branch of China National Center for Protein Sciences Beijing). We are in debt to Dr. N. Yan, Dr. H. Wang, Dr. X. Wang, Dr. L. Zhang, Dr. Q. Ding and Dr. X. Li for providing critical advices. **Funding:** This work was supported in part by funds from Major Project of Zhejiang Provincial Science and Technology Department #2020C03123-1, and National Science and Technology Major Project for the Control and Prevention of Major Infectious Diseases in China (#2018ZX10711001, #2018ZX10102001). **Author contributions:** S.L. conceived and supervised the project. H.Y., N.W., T.W., Z.W., L.C., D.S. and X.L. propagated, fixed and verified the fixation of the virus sample. Y.S. and S.L. isolated the viruses, performed biochemical analysis and prepared the cryo-sample for cryo-EM. Y.S., Y.C., S.L., J.Z. and J.L. collected the EM data. Y.S., Y.C., J.Z., J.X., C.S., Z.Z., and S.L. processed the EM data. S.L. and J.X. performed the subtomogram averaging and classification. Y.C., J.X., L.S. and Y.S. analyzed the structures. C.S., and J.X. performed the statistical analysis. M.C., Y.S. and Z.Z. analyzed the glycan data. S.L., Y.Shi., M.C. and L.L. wrote the manuscript. All authors critically revised the manuscript. **Competing interests:** The authors declare no competing interests. Data and materials availability: Cryo-EM maps of the reported structure have been deposited in the Electron Microscopy Data Bank under accession codes EMD-XXXX, EMD-XXXX, EMD-XXXX and EMD-XXXX.

List of Supplementary Materials:

Materials and Methods

Figures S1-S5

Tables S1

References (*41-58*)

Figures

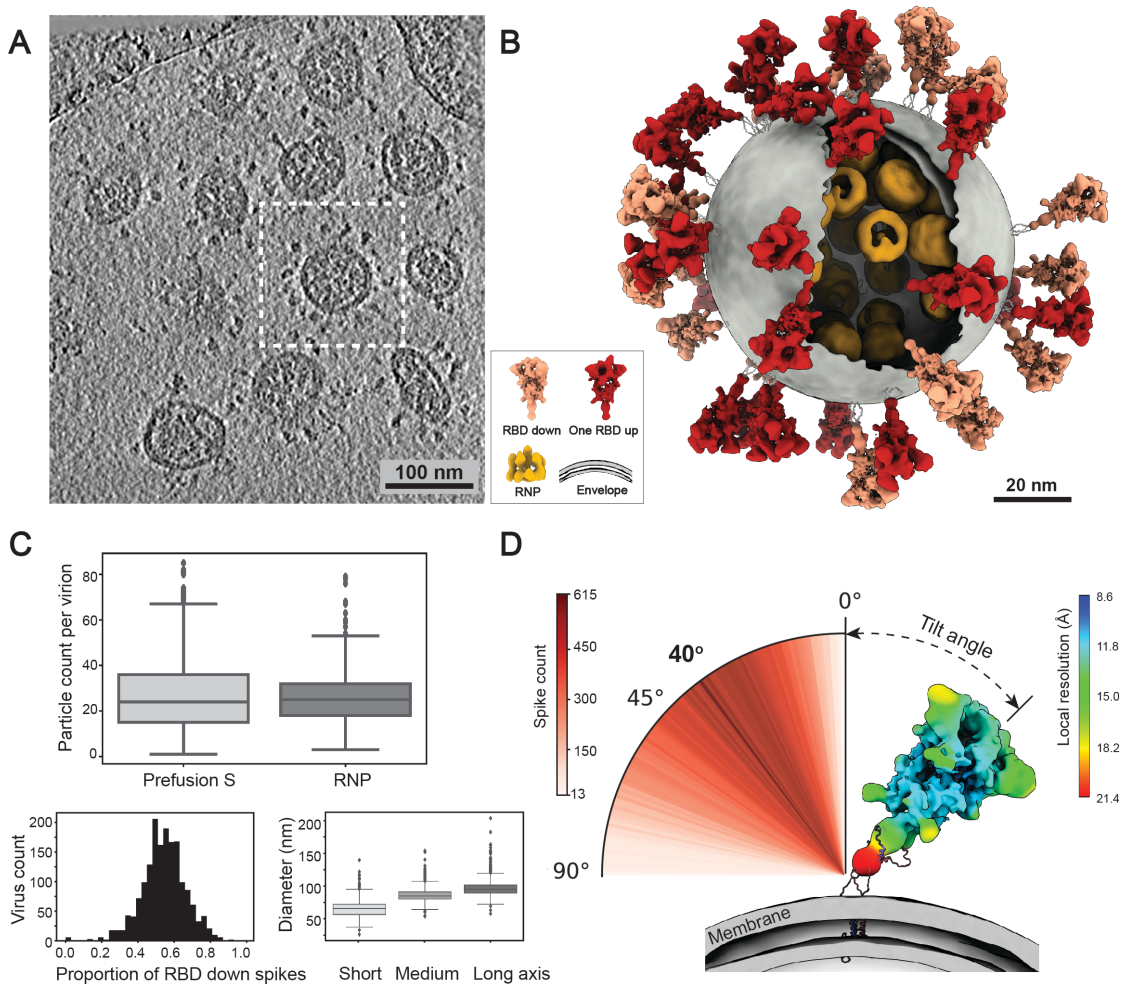

Fig. 1. Molecular architecture of the SARS-CoV-2 virus. (A) A representative tomogram slice (5 Å thick) showing pleomorphic SARS-CoV-2 virions. (B) A representative virus (boxed in A) is reconstructed by projecting the prefusion spikes in RBD down conformation (salmon) and one RBD up conformation (red), lipid envelope (grey), and RNPs (yellow) onto their refined coordinates. RNPs away from the envelope are hidden for clarity. The unsolved stem regions of the spikes are sketched from a predicted model (https://zhanglab.ccmb.med.umich.edu/COVID-19). (C) Summary of structural observations. *Upper*: number of prefusion S and RNPs per virus; *Lower left*: ratio of the S proteins between RBD down and one RBD up conformations; *Lower right*: statistics of the dimension of SARS-CoV-2 viral envelopes. (D) Distribution of the spike tilt angle reveals a prevailing tilt of 40° relative to the normal axis of the envelope. Shown here is a representative RBD down spike in authentic orientation to the envelope (grey). The spike is colored by local resolution, and the predicted model of the stem is fitted for illustration purpose.

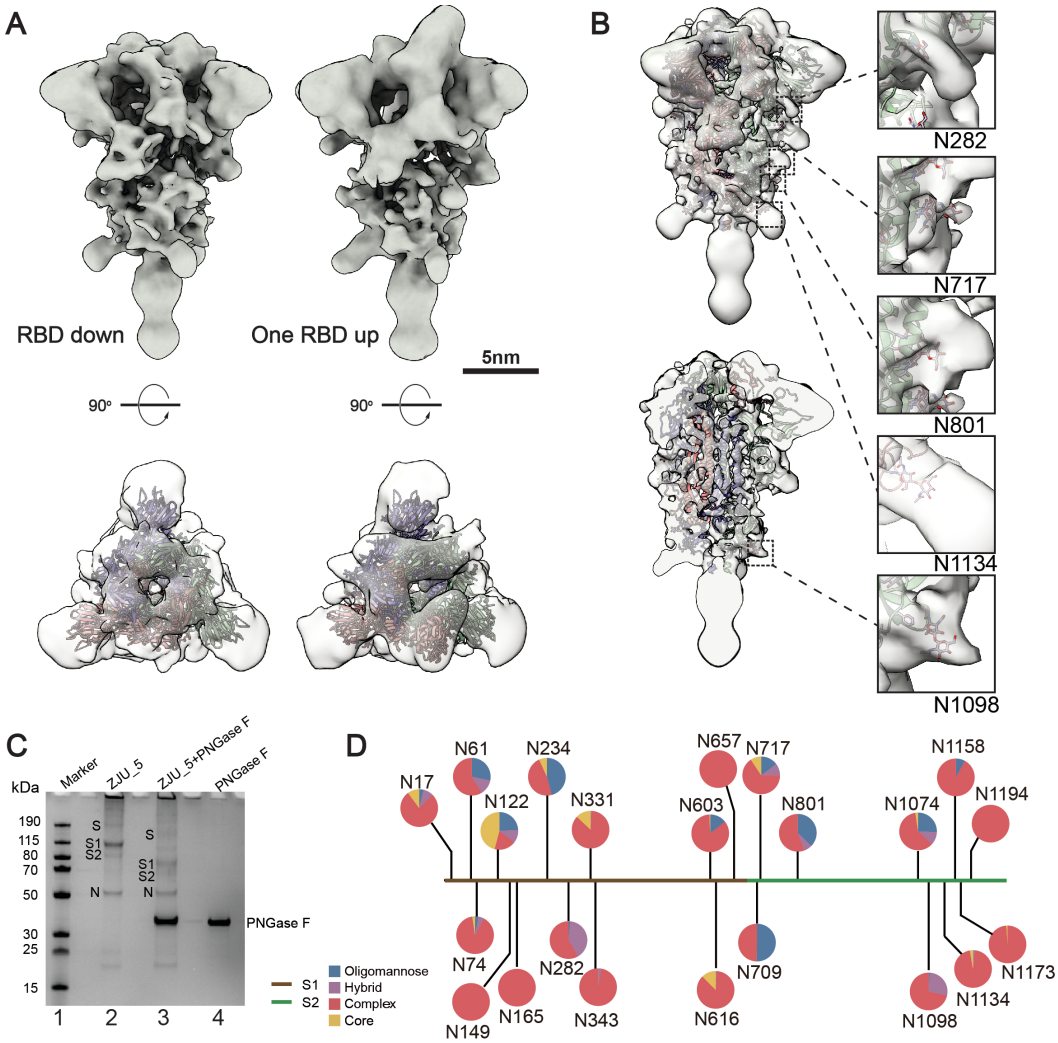

Fig. 2. Native structure of the S protein in the prefusion conformation. (A) S in RBD down (resolution 9.5 Å) and one RBD up (resolution 10.9 Å) conformations (threshold: 1.5). Sideview (upper) and topview (lower) of the maps respectively fitted with PDB entries 6VXX and 6VYB are shown. (B) The RBD down S maps fitted with PDB entry 6VXX. Densities of five glycans are highlighted in insets. (C) Compositional analysis of the viral sample. The treated viruses, designated ZJU_5, were resolved by SDS-PAGE and visualized by Coomassie blue staining. Lane (1): protein ladder; (2): purified ZJU_5; (3): 2 treated with PNGase F and (4): PNGase F as control. After PNGase treatment, the molecular weight of the S1 subunit is reduced by ~30 KDa, and S2 by ~20 KDa in weight. (D) Compositional analysis of the surface glycans. The identity and proportion of 22 N-linked glycans from the native S glycans were analyzed by mass spectrometry and presented in pie charts.

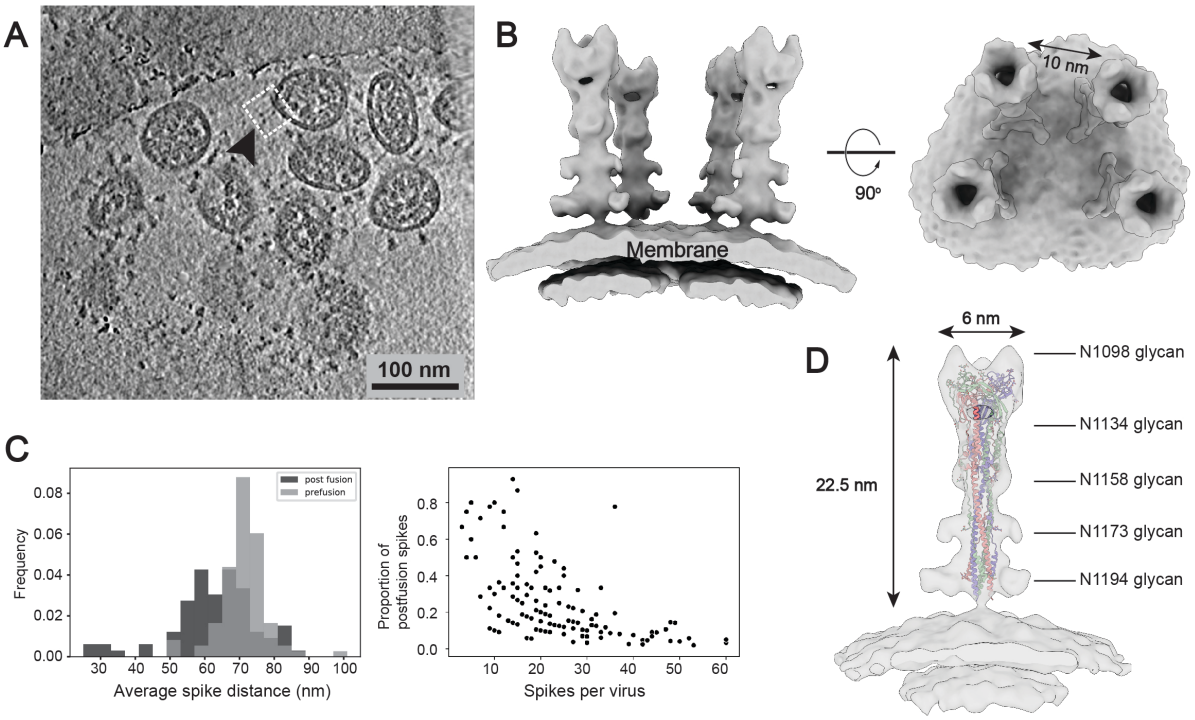

Fig. 3. Native structure of S2 in the postfusion conformation. (A) A representative tomogram slice (5 nm thick) showing a cluster of spikes in postfusion conformation on a SARS-CoV-2 virion. (B) 3D reconstructions of S2 in postfusion state. The boxed region in (A) is reconstructed by projecting the solved postfusion spike structure onto their refined coordinates. Resolution: 15.3 Å, threshold: 1. (C) Distribution of the postfusion spikes. *Left:* Statistics indicates the postfusion spikes are closer to each other on the viral surface compared to the prefusion spikes; *Right:* viruses possessing less spikes in total tend to have more postfusion spikes. (D) The postfusion spike structure displaying densities of N1098, N1134, N 1158, N1173 and N1194.

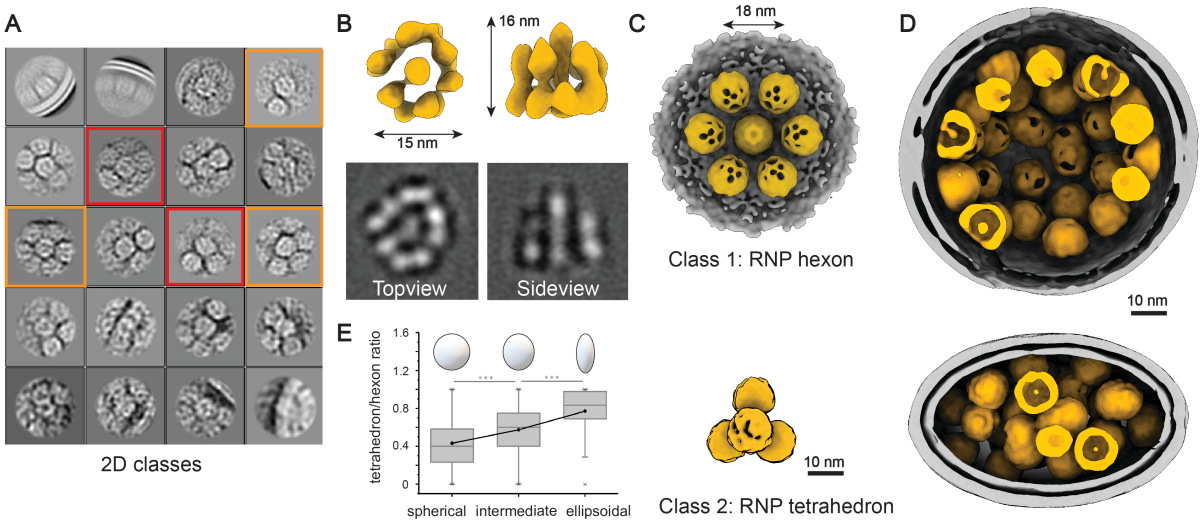

Fig. 4. Native assembly of the ribonucleoproteins (RNPs). (A) 2D class averages of RNPs reveal two distinctive types of the RNP ultrastructure: hexameric (boxed in orange) and tetrahedral (boxed in red) assembly. (B) 3D reconstruction and 2D slices of the RNP. Resolution: 13.1 Å, threshold: 1. (C) Ultrastructure of the RNP hexon and tetrahedron assemblies. Seven RNPs are packed against the viral envelope (grey), forming an "eggs-in-a-nest" shaped hexagonal assembly (upper). Four RNPs are packed as a membrane-free tetrahedral assembly (lower), most of which were found inside the virus away from the envelope. Structural features of the RNPs on the assembly are smeared due to the symmetry mis-match between individual RNPs and the assembly. (D) Representative projection of the RNP hexons assembling into a spherical virus, and tetrahedrons into an ellipsoidal virus. (E) Statistics of the ratio of tetrahedron/hexon assembly reveals that the spherical and ellipsoidal virions are more likely to be packed with hexons and tetrahedrons, respectively. The ratio is estimated by sorting 382 virions that have over 5 RNP assemblies by their ratio of long/short diameter and counting their ratio of tetrahedron to hexon RNPs.

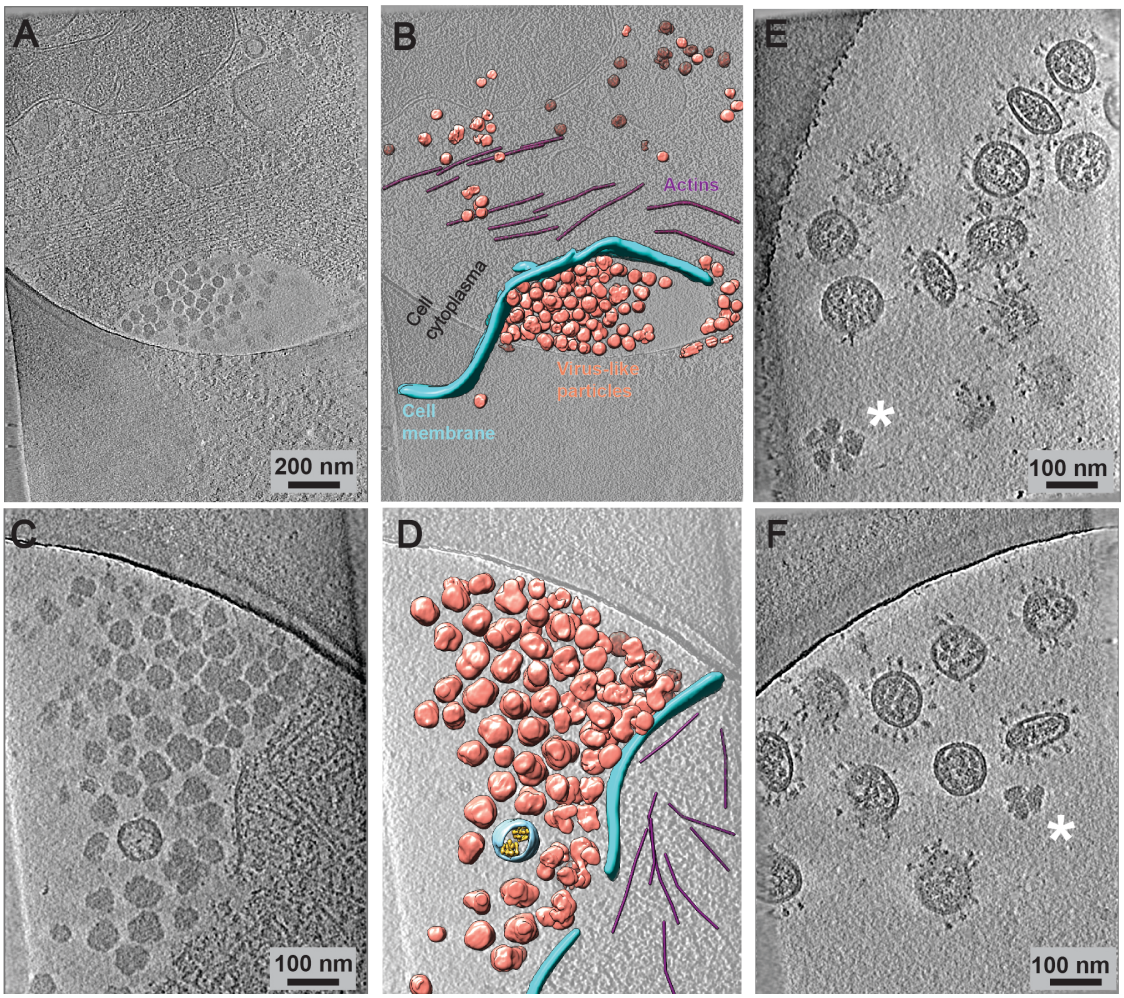

Fig. 5. Dense virus-like particles in the vicinity of the infected Vero cells. Vero cells were infected with SARS-CoV-2 followed by chemical fixation four days post-infection. (A, C) Cryo-ET captured a significant number of electron dense particles being released from the cells. Shown here are representative tomograms, which are segmented for clarification in (B and D). (salmon: dense particles, blue: cell membrane, purple: actin, yellow: viral RNP). (E, F) In our virus sample purified from the supernatant of these infected cells, these virus-like particles are often seen (*).