ORIGINAL ARTICLE

Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding

Miao Gui[1, *], Wenfei Song[2, *], Haixia Zhou[2], Jingwei Xu[1], Silian Chen[1], Ye Xiang[1], Xinquan Wang[2, 3]

[1]Center for Global Health and Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing Advanced Innovation Center for Structural Biology, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China; [2]The Ministry of Education Key Laboratory of Protein Science, Beijing Advanced Innovation Center for Structural Biology, Collaborative Innovation Center for Biotherapy, School of Life Sciences, Tsinghua University, Beijing 100084, China; [3]Collaborative Innovation Center for Biotherapy, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Chengdu, Sichuan 610041, China

The global outbreak of SARS in 2002-2003 was caused by the infection of a new human coronavirus SARS-CoV. The infection of SARS-CoV is mediated mainly through the viral surface glycoproteins, which consist of S1 and S2 subunits and form trimer spikes on the envelope of the virions. Here we report the ectodomain structures of the SARS-CoV surface spike trimer in different conformational states determined by single-particle cryo-electron microscopy. The conformation 1 determined at 4.3 Å resolution is three-fold symmetric and has all the three receptor-binding C-terminal domain 1 (CTD1s) of the S1 subunits in "down" positions. The binding of the "down" CTD1s to the SARS-CoV receptor ACE2 is not possible due to steric clashes, suggesting that the conformation 1 represents a receptor-binding inactive state. Conformations 2-4 determined at 7.3, 5.7 and 6.8 Å resolutions are all asymmetric, in which one RBD rotates away from the "down" position by different angles to an "up" position. The "up" CTD1 exposes the receptor-binding site for ACE2 engagement, suggesting that the conformations 2-4 represent a receptor-binding active state. This conformational change is also required for the binding of SARS-CoV neutralizing antibodies targeting the CTD1. This phenomenon could be extended to other betacoronaviruses utilizing CTD1 of the S1 subunit for receptor binding, which provides new insights into the intermediate states of coronavirus pre-fusion spike trimer during infection.

Keywords: SARS-CoV; spike glycoprotein; receptor-binding active state; cryo-EM

Introduction

Coronaviruses are a large group of highly diverse, enveloped, positive-sense, single-stranded RNA viruses that infect many mammalian and avian species. Currently, six coronavirus strains that are able to infect humans have been identified. Among them, alphacoronaviruses HCoV-229E and HCoV-NL63 and lineage A betacoronaviruses HCoV-OC43 and HCoV-HKU1 usually cause mild and self-limiting upper respiratory tract infection [1]. In 2002, the severe acute respiratory syndrome coronavirus (SARS-CoV), a lineage B betacoronavirus, was identified and infected more than 8 000 persons including nearly 800 related deaths worldwide in the 2002-2003 SARS pandemic [2-4]. Ten years later, another highly pathogenic lineage C betacoronavirus named the Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in Saudi Arabia in 2012 [5, 6]. Since its discovery, the MERS-CoV has infected 1 800 persons including 640 related deaths according to the WHO data in August, 2016. These two deadly coronaviruses have

---

been extensively studied in epidemiology, virology, clinical features and other aspects [7-10]. However, there are still no approved antiviral drugs and vaccines to treat and prevent the infections of SARS-CoV and MERS-CoV.

The spike (S) glycoprotein on the coronavirus envelope is responsible for host cell attachment, receptor binding, and for mediating host cell membrane and viral membrane fusion during infection. It is synthesized as a precursor single polypeptide chain of ~1 300 amino acids and then cleaved by host furin-like proteases into an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit [7, 11]. The S1 subunit contains domains for host cell attachment by recognizing cell surface sugar molecules and binding to specific cellular receptors [12, 13]. Therefore, the S1 subunit, especially its receptor-binding domain (RBD) is critical in determining cell tropism, host range and zoonotic transmission of coronaviruses [14, 15]. The S2 subunit contains a hydrophobic fusion loop and two heptad repeat regions (HR1 and HR2), which suggest a coiled helix structure of the S2 subunit [7]. Previous studies suggested that three S monomers assemble to form homo-trimer spikes anchoring on the outmost viral envelope [16]. Binding of RBD to cellular receptors triggers conformational changes in the S1 and S2 subunits, leading to the exposure of the fusion loop and its insertion into target cell membrane [17]. The HR1 and HR2 regions in the S glycoprotein trimer then form a six-helix bundle fusion core that bridges the viral and host cell membranes into close apposition to facilitate fusion [17]. For the highly pathogenic SARS-CoV and MERS-CoV, the RBD in the S1 subunit and the post-fusion core in the S2 subunit have been structurally and functionally studied as separate domains [18-23]. A previous study of the SARS-CoV virions by single-particle cryo-electron microscopy (cryo-EM) reported the structure of the S glycoprotein trimers on the virion at a low resolution of 16.0 Å [16]. Recently the pre-fusion structures of mouse hepatitis virus (MHV) and human coronaviruses HKU1 and HCoV-NL63 S glycoprotein trimers were determined by cryo-EM at 4.0, 4.0 and 3.4 Å resolutions, respectively [24-26]. However, high-resolution structures of the highly pathogenic SARS-CoV and MERS-CoV S glycoprotein trimers are still missing. In addition, intermediate states of the coronavirus S glycoprotein trimer are also required for a better understanding of the molecular mechanisms underlying receptor binding and membrane fusion.

We report here the cryo-EM structure determination of the SARS-CoV S glycoprotein trimer in four different conformations. Structural analyses revealed that these conformations are different in the position of one C-terminal domain 1 (CTD1), which functions as the RBD of the S glycoprotein trimer. Structural comparisons further indicated that a "down" to "up" positional change of the CTD1 switches the S glycoprotein trimer from receptor-binding inactive to active state, which is a prerequisite for the binding of SARS-CoV receptor ACE2 and for the neutralization by monoclonal antibodies.

Results

Structure determination

By using the Bac-to-Bac insect cell system, we expressed and purified a mutant SARS-CoV S glycoprotein ectodomain, in which the residue Arg667 at the S1/S2 cleavage site was mutated to alanine to enhance sample homogeneity and a strep tag was added to the C-terminus to facilitate purification [27] (Figure 1A). The purified SARS-CoV S glycoprotein was subjected to cryo-EM structural analysis using an FEI Titan Krios electron microscope equipped with a Gatan K2 Summit direct electron counting camera (Supplementary information, Figure S1A). Projected secondary structure features were clearly visible in the 2D classification analysis of the boxed particles (Supplementary information, Figure S1B). After 3D classification and refinement, four major different conformational states (conformations 1-4) were determined (Supplementary information, Figure S2A). The SARS-CoV spike has an overall mushroom-like shape. One of the conformation states is closely three-fold symmetric, whereas the other three conformation states show significant asymmetric features in the mushroom head region. Further calculation and processing were performed with C3 symmetry imposed for the symmetric conformation 1 and without any symmetry imposed for the asymmetric conformations 2-4 (Supplementary information, Figures S2 and S3).

The resolution of the final C3 symmetric conformation 1 calculated with 34 152 particles was 4.3 Å (Supplementary information, Figure S1). The atomic model of the SARS-CoV S was built based on the C3 symmetric density map (Supplementary information, Table S1A).

SARS-CoV S glycoprotein trimer conformation 1

In the SARS-CoV S glycoprotein, the β-strand-rich S1 subunit is composed of an N-terminal domain (NTD, residues 14-294) and three C-terminal domains (CTD1, residues 320-516; CTD2, residues 517-578 and CTD3, residues 579-663; Figure 1A and 1B). The CTD1 functions as the RBD of SARS-CoV S glycoprotein, which specifically binds to the cellular receptor ACE2 [18, 28]. The CTD1 is immediately followed by the CTD2 and CTD3, and the NTD is connected to the CTD1 through a long linker (residues 295-319; Figure 1A and 1B). The
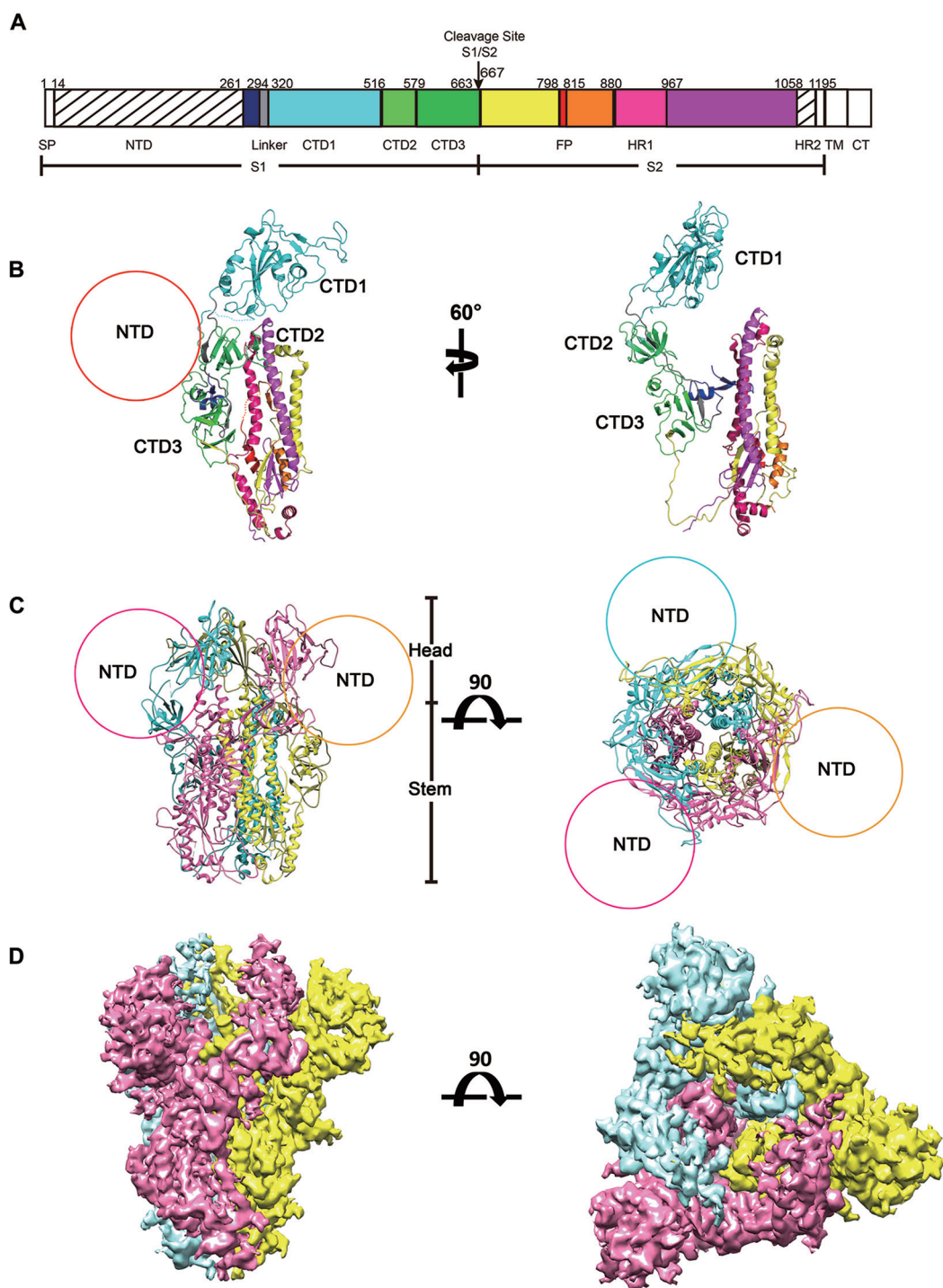
Figure 1 Overall structure of the SARS-CoV S glycoprotein. **(A)** A schematic diagram showing the domain organization of the SARS-CoV S glycoprotein. SP: signal peptide; NTD: N-terminal domain; CTD1: C-terminal domain 1, cyan; Linker: the linker connecting NTD and CTD1, grey; CTD2: C-terminal domain 2, light green; CTD3: C-terminal domain 3, dark green; FP: fusion peptide, red; HR1: heptad repeat 1, pink; HR2: heptad repeat 2; TM: transmembrane domain; CT: cytoplasmic tail. The TM and CT regions are not included in the expression construct. The NTD and HR2 regions that are not resolved in the reconstruction are represented with diagonal stripes. **(B)** Ribbon diagrams showing the structure of one SARS-CoV S glyco-protein monomer with domains colored as the same as in **A**. **(C)** Ribbon diagrams showing the structure of the SARS-CoV S glycoprotein trimer. The three protomers are colored pink, yellow and cyan, respectively. The circles indicate the locations of the unmodeled NTD regions. **(D)** Surface shadowed diagrams showing the 4.3 Å resolution 3D density map of the SARS-CoV S trimer. The protomers are colored the same as in **C**.

NTDs of MHV and HKU1 S glycoproteins are structurally similar and both adopt a galectin-like β-sandwich fold. The model to map correlation values are 0.86 for the MHV NTD and 0.87 for the HKU1 NTD (Supplementary information, Figure S4A and S4B). Fitting of the MHV or HKU1 NTD structure into the corresponding SARS-CoV S glycoprotein density map gave relatively low resolution to map correlation values (0.55 for MHV NTD and 0.52 for HKU1 NTD; Supplementary information, Figure S4C and S4D), indicating that the SARS-CoV NTD has different local conformations, although it may still adopt the galectin-like β-sandwich fold. The relatively low resolution in this region also did not allow us to perform *ab initio* model building of the NTD, therefore only two strands and a short α-helix were built and residues 14-260 were not included in the atomic model (Figure 1A and 1B). The α-helix-rich S2 subunit begins after the S1/S2 cleavage site at residue 667 (Figure 1A). The atomic model of the S2 subunit includes the functionally important fusion peptide (residues 798-815) and HR1 (residues 880-967; Figure 1A and 1B). The C-terminal HR2 (residues 1 154-1 183) was not built due to relatively poor density in this region (Figure 1A and 1B).

In the conformation 1 with three-fold symmetry, the three S glycoprotein monomers intertwine around each other to form a closely packed mushroom-shaped homotrimer (Figure 1C and 1D). The triangular head of the trimer spike is composed of the NTDs and CTD1s of three S1 subunits. Three CTD1s locate in the center of the triangular head and are arranged around the 3-fold symmetry axis (Figure 1C and 1D). Three NTDs locate at the outside of the triangular head and each NTD interacts with one CTD1 from the neighboring S1 subunit (Figure 1C and 1D). The stem of the trimer spike consists of a core helix bundle formed by the long helices of three S2 subunits, and the core helix bundle is further surrounded by CTD2s and CTD3s of three S1 subunits (Figure 1C and 1D). The three CTD1s in the head all lay on and cover the top of the S2 subunits (Figures 1C, 1D and 2A).

SARS-CoV S glycoprotein trimer conformations 2-4

We observed three other conformations 2-4 showing asymmetric features of the CTD1 in the triangular head (Figure 2B-2D). In the symmetric conformation 1, the CTD1s in the head are all in a "down" position, covering the S2 subunits in the stem (Figure 2A). The angle between the long axis of the CTD1 and its projection on the horizontal plane perpendicular to the 3-fold axis is ~19 degree (Figure 2A). In the conformations 2-4, two CTD1s still adopt the same "down" conformation as in the conformation 1, whereas one CTD1 rotates outward to an "up" position and no longer covers the S2 subunit

(Figure 2B-2D). The angles between the long axes of the "up" CTD1s and their projections on the horizontal plane have increased to 52, 64 and 70 degrees, respectively (Figure 2B-2D). Of note, the NTD and CTD2 around the rotated CTD1 do not have significant conformational changes (Supplementary information, Figure S5A). Therefore, the rotation of the CTD1 is a hinge motion around the loops connecting NTD to CTD1 and CTD1 to CTD2 that are in close proximity (Supplementary information, Figure S5B and S5C), suggesting that these two loops would play key roles in the conformational switch between the "down" and "up" positions of CTD1. Although the position of the CTD2 is not affected within the monomer where the CTD1 adopts an "up" conformation, the CTD2 may become more liable to undergo conformational changes to expose the fusion loop underneath.

Receptor-binding inactive and active states of the S glycoprotein trimer

As the RBD of SARS-CoV S glycoprotein trimer, the CTD1 specifically binds to the cellular receptor ACE2, which is a prerequisite for the host cell attachment of the virion and the subsequent membrane fusion [18, 28]. We superimposed the previously determined CTD1-ACE2 complex crystal structure onto one CTD1 of the S glycoprotein trimer (Figure 3A). For conformation 1 in which all CTD1s are in the "down" positions, numerous steric clashes were observed between ACE2 and the neighboring CTD1, and the volume of the steric clashes reaches 10 696 $Å^3$ (Figure 3B and 3C). Therefore, the SARS-CoV S glycoprotein trimer with all its CTD1s in the "down" positions would not be able to bind the cellular receptor ACE2, suggesting that the three-fold symmetric conformation 1 represents a receptor-binding inactive state of the spike. In contrast, similar structural superimpositions showed that ACE2 binds the "up" CTD1 well without steric clashes with other regions of the S glycoprotein trimer (Figure 3D-3F, Supplementary information, Figure S6A and S6B), suggesting that spikes with one CTD1 in the "up" position are able to bind receptor ACE2 and asymmetric conformations 2-4 represent a receptor-binding activate state of the spike. Notably, the "up" position of the CTD1 also exposes one of the covered S2 subunits, leaving the space for the large-scale conformational change of the S2 subunit to expose and insert the fusion peptide into the target cell membrane.

The "up" conformation of the CTD1 is also required for the binding of neutralizing antibodies

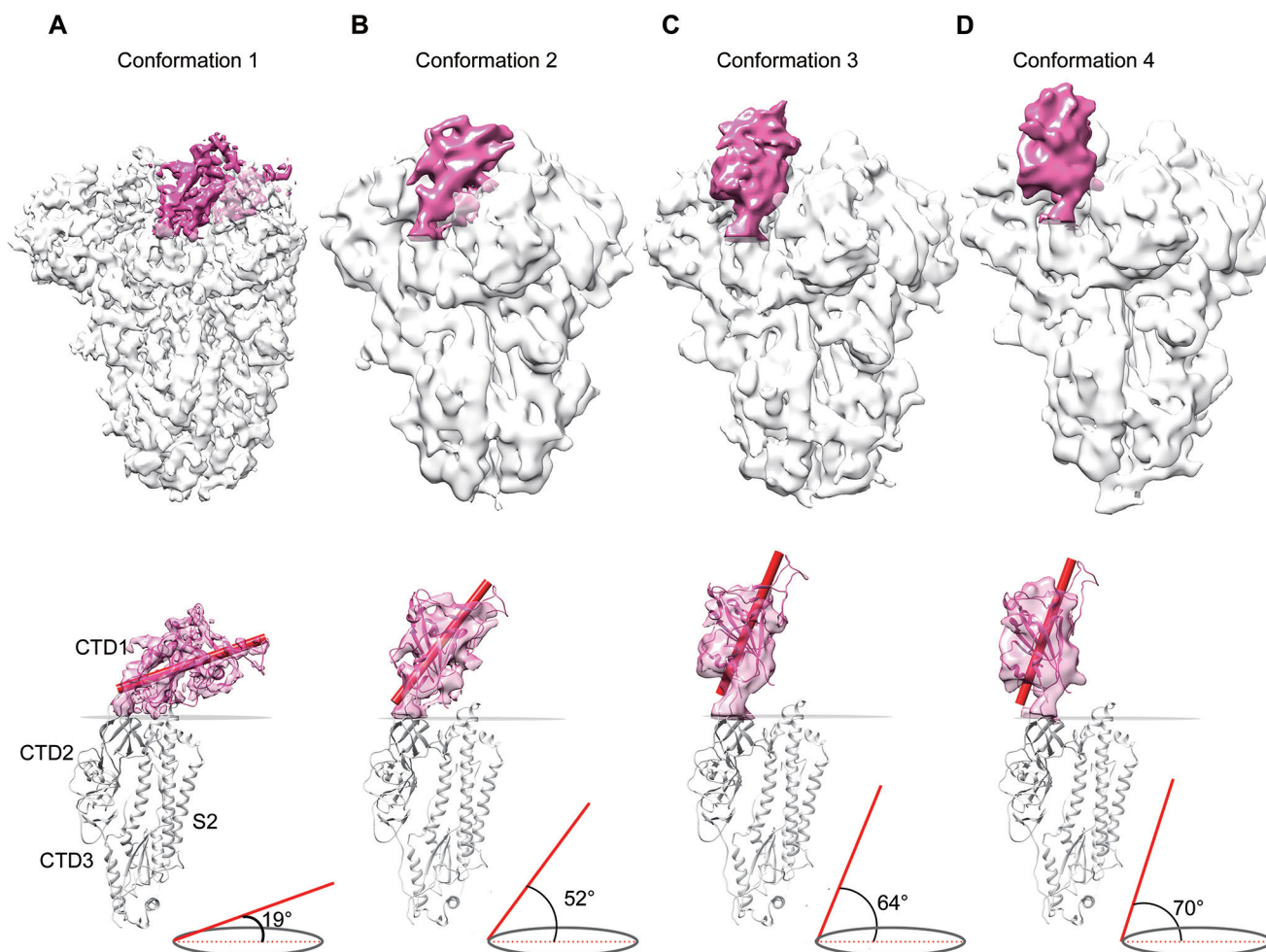Several neutralizing antibodies targeting the CTD1 (m339, 80R and F26G19) have shown potent inhibition

Figure 2 Four different conformations of the SARS-CoV S glycoprotein trimer. Top: surface shadowed diagrams showing the four different conformations (conformations 1-4) of the S trimer. The CTD1s are colored pink. Bottom: ribbon diagrams showing S monomers with the semi-transparent CTD1 densities colored pink. The tilt angles of the CTD1s are defined by the angle between the long axis of the CTD1 (red cylinder) and its projection on the horizontal plane (grey ellipse). **(A)** Three-fold symmetric conformation 1 with all the three CTD1s in the "down" conformations. **(B-D)** Asymmetric conformations 2-4 with one CTD1 in the "up" conformation.

activity against the cell infection of pseudo-typed or live SARS-CoV, and the antibody binding epitopes have been elucidated by crystal structure determination of the antibody-CTD1 complexes [29-31]. These antibodies binds to the receptor-binding site on the CTD1, thereby directly inhibiting the engagement of ACE2 receptor. Structural superimpositions showed that these antibodies all have steric clashes with other regions of the S glycoprotein trimer in the conformation 1 state (Figure 4A-4C). In contrast, the S glycoprotein trimer conformations 2-4 would allow the antibodies to bind the "up" CTD1 and inhibit its interaction with the ACE2 receptor (Figure 4D-4F). These results indicated that the receptor-binding inactive to active state transition of the S glycoprotein

trimer is required not only for successful infections of the SARS-CoV virions, by also for efficient neutralization by antibodies targeting the CTD1.

Discussion

As the known largest class I viral fusion protein, the coronavirus S glycoprotein trimer recognizes a variety of host cell receptors through the NTD or CTD1 of the S1 subunit, and subsequently mediates the viral and cell membrane fusion through the fusion peptide and two heptad repeats of the S2 subunit. Since the discoveries of highly pathogenic SARS-CoV in 2002 and MERS-CoV in 2012, most studies have been focused on the RBD
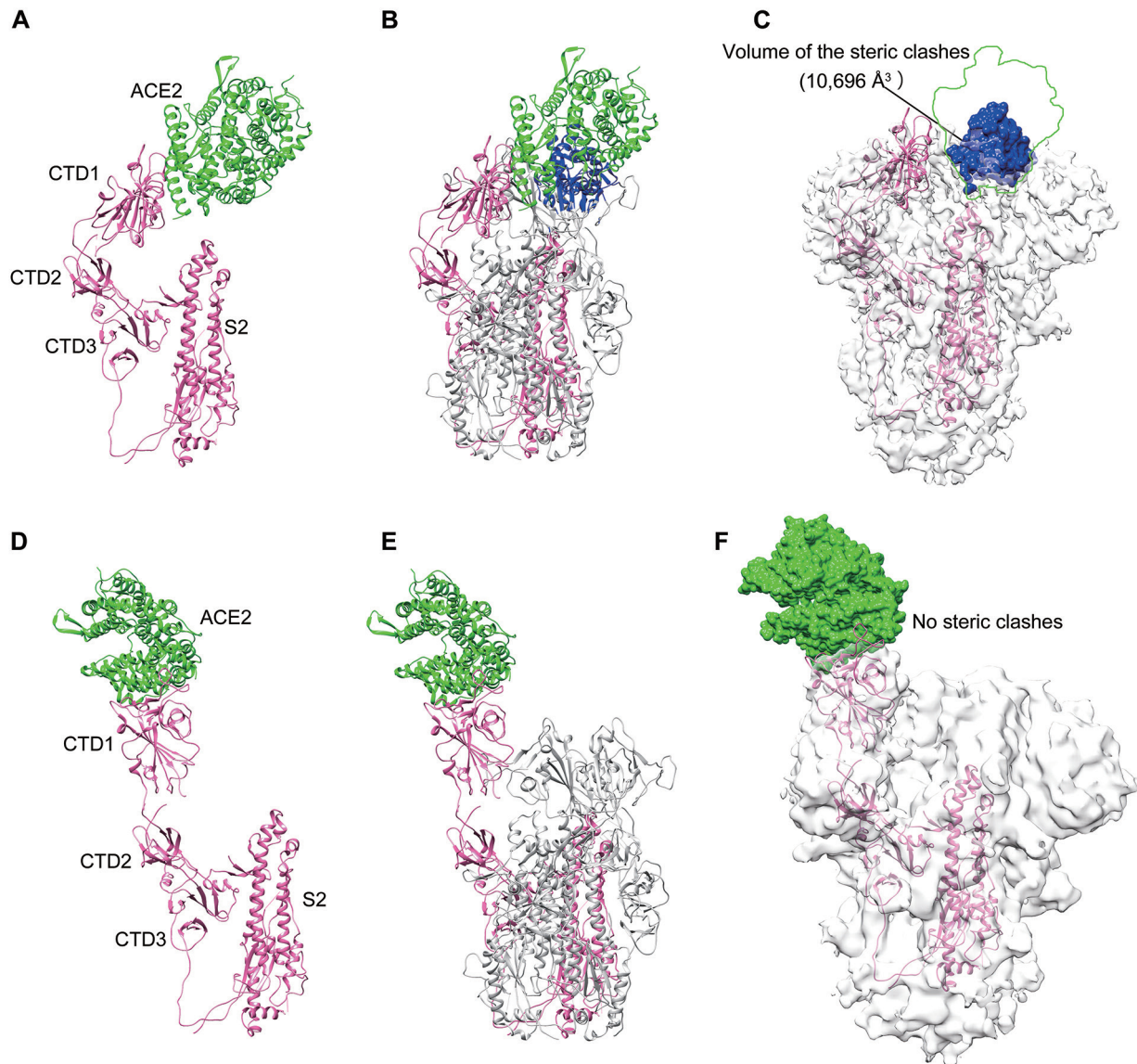
Figure 3 Models of the SARS-CoV S monomer and trimer bound with the receptor ACE2. **(A)** "Binding" of the receptor ACE2 (green) to one S monomer (pink) of the conformation 1 S trimer. The CTD1 is in the "down" conformation. **(B)** "Binding" of the receptor ACE2 (green) to the conformation 1 S trimer. Three CTD1s are all in the "down" conformations. The steric clashes between a neighboring CTD1 (grey) and ACE2 (green) are colored blue. **(C)** The same model as in **A**, with the S trimer density map presented. Only the boundary profile of the "bound" ACE2 is shown (green lines) for a better view of the clashes (volume: 10 696 Å^3). **(D)** "Binding" of the receptor ACE2 (green) with the conformation 3 S monomer (pink) of which the CTD1 is in the "up" conformation. **(E)** "Binding" of the receptor ACE2 (green) to the "up" CTD1 (pink) of the conformation 3 S trimer showing no steric clashes with any neighboring "down" CTD1 (grey). **(F)** The same model as in **D** with the S trimer 3D density map presented. All models are generated by superimposing the CTD1-ACE2 complex crystal structure onto the CTD1 of the corresponding SARS-CoV S monomer or trimer. The NTD models are not shown.

(NTD or CTD1) of the S1 subunit because it binds to the host receptors and is also the main target of neutralizing antibodies during infection. Our knowledge about the structures of complete SARS-CoV and MERS-CoV spike trimers in pre-fusion and post-fusion states are still very limited.

In the present study, we determined the structure of SARS-CoV S glycoprotein trimer in four different conformational states. Recently reported cryo-EM structures of the MHV and HKU1 S glycoprotein trimers are simi-
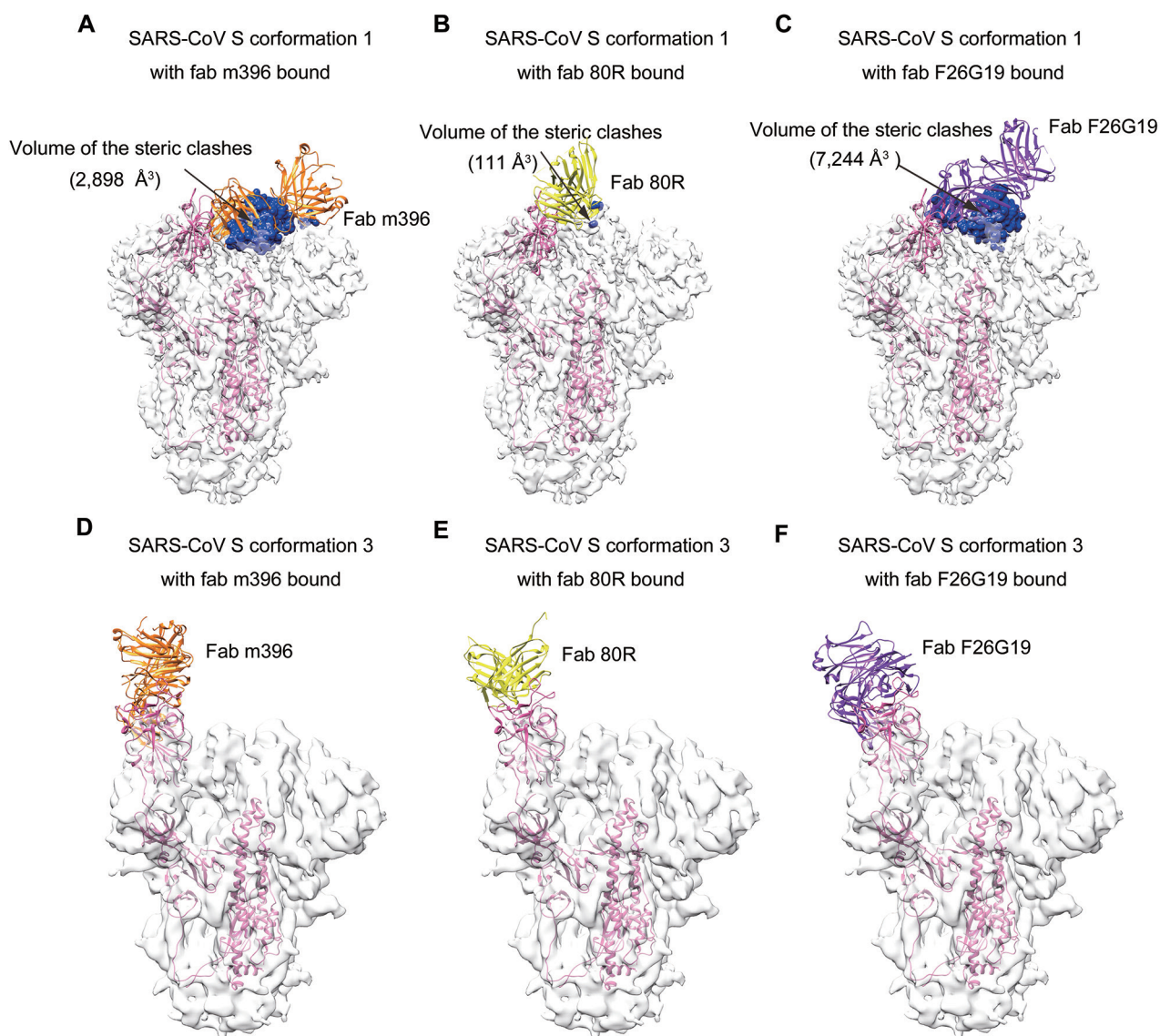
A SARS-CoV S corformation 1
with fab m396 bound

B SARS-CoV S corformation 1
with fab 80R bound

C SARS-CoV S corformation 1
with fab F26G19 bound

Figure 4 Structural superimpositions showing the "binding" of neutralization antibodies to the SARS-CoV S trimers. **(A)** Structural superimposition of the CTD1-Fab m396 complex (PDB accession code: 2DD8) onto one CTD1 of the SARS-CoV S trimer in conformation 1, showing the "binding" of Fab m396 to the SARS-CoV S trimer. The EM densities of the S trimer are represented using shadowed surfaces in semi-transparent grey. **(B-C)** Similar structural superimpositions showing the "binding" of neutralization antibodies 80R (PDB accession code: 2GHW; **B**) and F26G19 (PDB accession code: 3BGF; **C**). The steric clashes with the "bound" Fab are colored blue and the corresponding volumes are shown in bracket. **(D-F)** Structural superimpositions of three CTD1-antibody complex structures with the SARS-CoV S trimer in conformation 3. No steric clashes between the "bound" Fab and the S trimer.

lar to the conformation 1 of the SARS-CoV S glycoprotein trimer in the receptor-binding inactive state, in which all three CTD1s in the S1 subunit are in "down" positions (Figure 2A). The SARS-CoV S glycoprotein trimer conformations 2-4 in the receptor-binding active state (Figure 2B-2D), in which one CTD1 is in an "up" position, were not observed in the MHV and HKU1 S gly-

coprotein trimers. The protein receptor of the HKU1 has not been identified and the NTD in the S1 subunit was shown to bind O-acetylated sialic acids on host cells [32]. The RBD of MHV S glycoprotein is the NTD of the S1 subunit, and structural alignment revealed that the MHV protein receptor CEACAM1a binds to the NTD in the S glycoprotein trimer without steric clashes (Supplementary

information, Figure S6C) [33]. Therefore, the MHV and HKU1 S glycoprotein trimer may not need the prerequisite receptor-binding inactive to active state transition for receptor binding. This intrinsic receptor-binding difference among these coronavirus S glycoprotein trimers may be one of the reasons why the previous studies were not able to reveal other conformations rather than conformation 1 in the MHV and HKU1 S glycoprotein trimers. Nevertheless, the conformational change of CTD1 may still be needed for the subsequent exposure and conformational changes of the S2 subunit for membrane fusion during the infection of MHV and HKU1.

The protein receptor-binding site in the CTD1 has also been confirmed for MERS-CoV and bat coronavirus HKU4, which are lineage C betacoronaviruses and have the same host receptor DPP4 [19, 20, 34-36]. Structural superimposition of DPP4 complexed with the CTD1 of MERS-CoV or HKU4 onto the SARS-CoV S glycoprotein trimer also showed that the "up" conformation of CTD1 is required for the binding of DPP4 by these two betacoronaviruses (Supplementary information, Figure S7). Therefore, we suggest that the receptor-binding inactive to active state transition observed in the SARS-CoV spike would also occur in the spikes of other betacoronaviruses that bind to their host receptors through the CTD1. In receptor-binding inactive state, all three CTD1s in the "down" positions would not allow efficient binding to receptors (Figure 5). The transition of

one CTD1 from "down" to "up" position would release the steric clashes and enable the binding of one receptor molecule (Figure 5). Whether the binding of the first receptor molecule is enough, or the binding of the second and third receptors is required to trigger conformational changes in the S1 and S2 subunit necessary for membrane fusion will be an interesting question for future studies.

Materials and Methods

Protein expression and purification

A human codon-optimized gene coding the SARS-CoV spike (S) glycoprotein ectodomain (NCBI Accession NP_828851.1) residues 1-1 195 with an R667A mutation to enhance sample homogeneity and a C-terminal strep tag for purification was cloned and inserted into pFastBac-Dual vector (Invitrogen). The construct was transformed into bacterial DH10Bac competent cell and the extracted Bacmid was then transfected into Sf9 cells using Cellfectin II Reagent (Invitrogen). The low-titer viruses were harvested and then amplified to generate high-titer virus stock, which was used to infect 2 L Sf9 cells at a density of 2×10^6 cells/ml. The supernatant of cell culture containing the secreted SARS-CoV S glycoprotein was harvested 60 h after infection, concentrated and buffer-exchanged to binding buffer (10 mM HEPES, pH 7.2, 500 mM NaCl). SARS-CoV S glycoprotein was captured by StrepTactin Sepharose High Performance (GE Healthcare) and eluted with 10 mM D-desthiobiotin in binding buffer. The eluted protein was then purified by gel filtration chromatography using the Superose 6 column (GE Healthcare) pre-equilibrated with HBS buffer (10
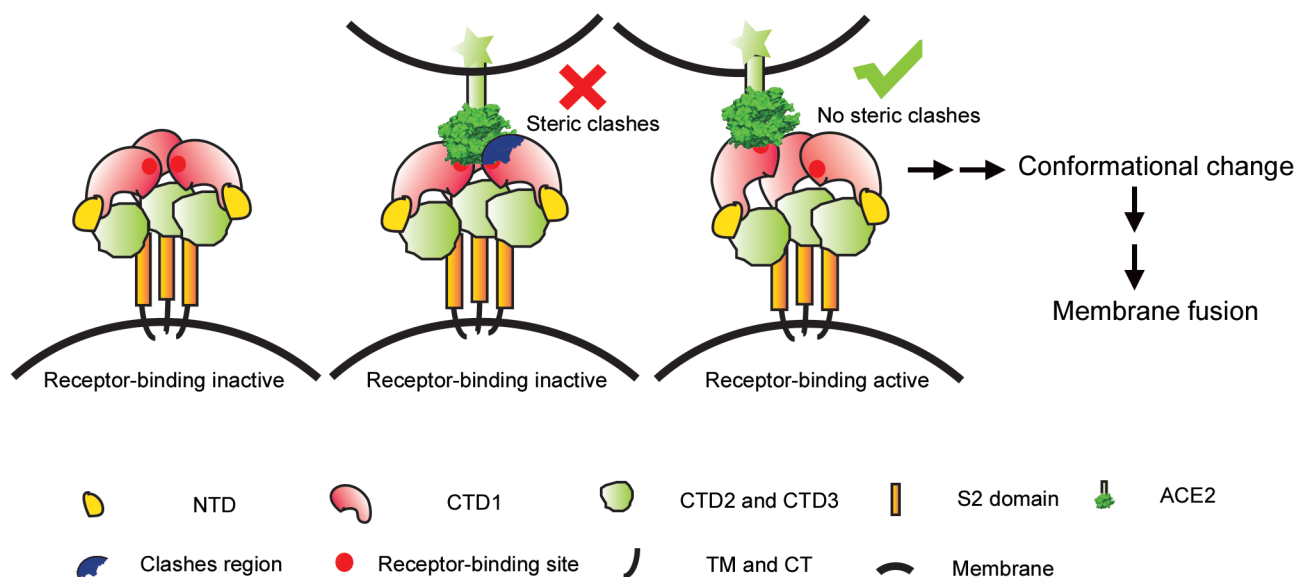

Figure 5 A cartoon model showing the transition of the S trimer spikes from receptor-binding inactive to active and subsequent fusogenic states. For betacoronavirus S glycoprotein utilizing the CTD1 as the receptor-binding domain, the state transition of the S trimer with the "down" to "up" conformational change of CTD1 would allow receptor binding and may initiate subsequent conformational changes in the S2 subunits to mediate membrane fusion.

mM HEPES, 150 mM NaCl). Fractions containing SARS-CoV S glycoprotein were pooled and concentrated for electron microscopy analysis.

Cryo-EM

Aliquots of 3 μl purified SARS-CoV S glycoprotein were applied to glow-discharged holey carbon grids (Quantifoil, Cu 400 mesh, R1.2/1.3) or grids with a layer of continuous ultrathin carbon film (with/without glow-discharge treatment) (Ted Pella, Inc.). The protein concentration was ~0.34 mg/ml for holey carbon grids and ~0.17 mg/ml for continuous carbon grids. The grids were blotted and then were plunged into liquid ethane using an FEI Vitrobot. The grids were checked in an FEI F20 microscope operating at 200 kV with a Gatan 895 4 k × 4 k CCD camera and a small data set was collected for generating the initial model. Data for final classification and refinement were collected on an FEI Titan Krios microscope operating at 300 kV using a K2 Summit camera (Gatan Inc.) in super-resolution mode with a nominal magnification of 22 500× (yielding a calibrated pixel size of 1.32 Å). Each image was fractionated into 32 movie frames with a total exposure time of 8 s and at a dose rate of ~8 counts per physical pixel per second (~4.7 electrons/Å^2/s). UCSFImage4 was used for all data collection [37].

Image processing

The image processing procedures were summarized in Supplementary information, Figure S2A. An initial map was generated using the program EMAN2 with the images collected on the F20 microscope [38]. A total of 3 309 movie stacks were collected using the K2 camera. All movie frames were aligned using the program motioncorr and the CTF parameters were determined by CTFFIND4 [39, 40]. A total of 4 015 particles were manually picked and 2D classifications were performed in RELION 1.3 [41]. Four representative class averaged images were selected as references for automatic particle picking with the whole data set using RELION 1.4. All 2D and 3D classifications and refinements were performed using RELION 1.4. Auto-picked particles were visually inspected and then were selected by several rounds of reference-free 2D classifications. During the 2D classification, particles from classes not showing clear secondary structure features were deleted, finally yielding 210 129 selected particles. These selected particles were subjected to an initial 3D refinement with C3 symmetry imposed, followed by two runs of 3D classifications. A total of 52 983 particles of the best class were selected and subjected to a 3D auto-refine. At the final stage of the refinement, 34 152 particles with a higher "loglikelicontribution" value were used. The resolution of the final C3 symmetric density map was 4.3 Å post processing in RELION.

The initial 210 129 selected particles were also subjected to 3D refinements and classifications without any symmetry imposed. The central parts of the different classes were similar but most classes showed obvious asymmetric features with one of three CTD1 domains protruding up. These asymmetric particles could be grouped into three major classes according to the tilt angle of the "up" CTD1. Particles from these classes were separately subjected to 3D auto-refine. The resolutions of the final density maps were 7.3, 5.7 and 6.8 Å for the three classes.

The handedness of the density maps was verified by docking the crystal structure of SARS-CoV S glycoprotein CTD1 into the maps. Reported resolutions are based on the gold-standard FSC 0.143 criterion. Local resolution variations were estimated using ResMap [42].

Model building and structure refinement

The crystal structure of the SARS-CoV S glycoprotein RBD (or CTD1, residues 324-502, PDB accession code: 3D0G) and the structures of HKU1 S glycoprotein (PDB accession code: 5i08), CTD2 (residues 596-673), CTD3 (residues 674-771) and S2 domain (residues 793-1147) were initially fitted into the SARS-CoV density map using UCSF Chimera [43]. Sequence alignment of the SARS-CoV S and the fitted structures was performed using DNAman (Lynnon Corporation, Quebec, Canada) and ClustalX [44]. The fitted model was rebuilt using RosettaCM with C3 symmetry imposed and the best output model was selected according to the energy and fitness of the model to the EM density map [45]. The model of CTD2, CTD3 and S2 region was refined using RosettaRelax and PHENIX real-space refinement [46, 47]. The crystal structure of the CTD1 domain was fitted into the EM density and was refined as a rigid body in real space by using PHENIX. Then the two models were merged, manually adjusted in COOT and refined in PHENIX again with reference model restrains, secondary structure restrains and geometry restrains [48]. Cross-validation of overfitting was performed following the procedures described before [49, 50]. Briefly, the atom coordinates of the model (including CTD1, CTD2, CTD3 and S2) were randomly displaced by 0.5 Å using PHENIX PDB tools. Then the displaced model was refined against one of the two half maps in reciprocal space by using PHENIX. FSC curves were calculated between the refined model and the corresponding half map that was used for the refinement (half1, FSC$_{work}$) and between the refined model and the other half map (half2, FSC$_{test}$) that was not used for the model refinement. No big gaps were observed between the work and test FSC curves (Supplementary information, Figure S1D), indicating that the model was not overfitted. Molprobity was used to evaluate the final refined model.

Models of the three asymmetric conformations were built by fitting the symmetric model of SARS-CoV S into the density maps using UCSF Chimera. The fitted models were then refined as rigid bodies in real space using PHENIX. For the subunit with the "up" CTD1, the subunit was split into two rigid bodies for refinement, including the "up" CTD1 and the rest part of the subunit.

Structural comparisons between different density maps and between density maps and models were performed using UCSF Chimera measure correlation, PHENIX *get_cc_mtz_mtz* and *get_cc_mtz_pdb*. The cross correlation values were listed in Supplementary information, Figure S4A-S4D and Table S1B.

The SARS-CoV S RBD-receptor complex (PDB accession code: 3R4D) and SARS-CoV S RBD-Fab complexes (PDB accession code: 2GHW, 2DD8 and 3BGF) were superimposed onto the models with the SARS-CoV S RBD structure as reference using the "match" command in UCSF Chimera. To calculate the "up" angles of the CTD1s (Figure 2), the horizontal plane of the S perpendicular to the 3-fold axis and the long axis of the CTD1 were generated using the UCSF Chimera "define" command and then the angle between the axis and the plane was calculated using the UCSF Chimera "angle" command. The overlap volumes between the fitted receptor and the S spike were also calculated using UCSF Chimera. Two atoms were considered to be overlapped if their overlap score > 0. The overlap score is defined as the sum of

two VDW (van der Waals) radii minus the distance between them and minus an allowance (0.4 Å) for potentially hydrogen-bonded pairs. All the overlap atoms were used for the generation of the steric clash area. The volume of the steric clashes was calculated using UCSF Chimera "measure volume" command. All figures were generated with UCSF Chimera and Pymol (The PyMOL Molecular Graphics System, Version 1.7.0.5 Schrödinger, LLC.).

Accession numbers

The cryo-EM maps and related materials have been deposited to the EM Data Bank under accession codes EMD-6679, EMD-6680, EMD-6681 and EMD-6682. The atomic coordinate has been deposited to the Protein Data Bank under accession code 5WRG.

Acknowledgments

Author Contributions

MG, WS, YX and XW designed the experiments. WS designed and cloned the protein constructs. WS and HZ carried out protein expression, purification and biophysical characterization. MG, JX and SC performed cryo-EM sample preparation and data collection. MG processed the cryo-EM data and built the atomic model. MG, WS, YX and XW analyzed the data. MG, WS, YX and XW wrote the manuscript with input from all authors.

Competing Financial Interests

The authors declare no competing financial interests.

References

1   Su S, Wong G, Shi W, *et al.* Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 2016; **24**:490-502.

2   Ksiazek TG, Erdman D, Goldsmith CS, *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003; **348**:1953-1966.

3   Drosten C, Gunther S, Preiser W, *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 2003; **348**:1967-1976.

4   Peiris JS, Lai ST, Poon LL, *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 2003; **361**:1319-1325.

5   Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012; **367**:1814-1820.

6   de Groot RJ, Baker SC, Baric RS, *et al.* Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol* 2013; **87**:7790-7792.

7   Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nat Rev Microbiol* 2009; **7**:226-236.

8   de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016; **14**:523-534.

9   Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen KY. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin Microbiol Rev* 2015; **28**:465-522.

10 Vijay R, Perlman S. Middle East respiratory syndrome and severe acute respiratory syndrome. *Curr Opin Virol* 2016; **16**:70-76.

11 Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 2015; **202**:120-134.

12 Li F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* 2015; **89**:1954-1964.

13 Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 2016; **3**:237-261.

14 Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol* 2015; **23**:468-478.

15 Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 2010; **84**:3134-3146.

16 Beniac DR, Andonov A, Grudeski E, Booth TF. Architecture of the SARS coronavirus prefusion spike. *Nat Struct Mol Biol* 2006; **13**:751-752.

17 Hofmann H, Pohlmann S. Cellular entry of the SARS coronavirus. *Trends Microbiol* 2004; **12**:466-472.

18 Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005; **309**:1864-1868.

19 Lu G, Hu Y, Wang Q, *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 2013; **500**:227-231.

20 Wang N, Shi X, Jiang L, *et al.* Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 2013; **23**:986-993.

21 Gao J, Lu G, Qi J, *et al.* Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. *J Virol* 2013; **87**:13134-13140.

22 Lu L, Liu Q, Zhu Y, *et al.* Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat Commun* 2014; **5**:3067.

23 Supekar VM, Bruckmann C, Ingallinella P, Bianchi E, Pessi A, Carfi A. Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc Natl Acad Sci USA* 2004; **101**:17958-17963.

24 Walls AC, Tortorici MA, Bosch BJ, *et al.* Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 2016; **531**:114-117.

25 Kirchdoerfer RN, Cottrell CA, Wang N, *et al.* Pre-fusion structure of a human coronavirus spike protein. *Nature* 2016; **531**:118-121.

26 Walls AC, Tortorici MA, Frenz B, *et al.* Glycan shield and epitope masking of a coronavirus spike protein observed by

cryo-electron microscopy. *Nat Struct Mol Biol* 2016; **23**:899-905.

27 Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci USA* 2009; **106**:5871-5876.

28 Li W, Moore MJ, Vasilieva N, *et al*. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003; **426**:450-454.

29 Hwang WC, Lin Y, Santelli E, *et al*. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *J Biol Chem* 2006; **281**:34610-34616.

30 Prabakaran P, Gan J, Feng Y, *et al*. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem* 2006; **281**:15829-15836.

31 Pak JE, Sharon C, Satkunarajah M, *et al*. Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. *J Mol Biol* 2009; **388**:815-823.

32 Huang X, Dong W, Milewska A, *et al*. Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. *J Virol* 2015; **89**:7202-7213.

33 Peng G, Sun D, Rajashankar KR, Qian Z, Holmes KV, Li F. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci USA* 2011; **108**:10696-10701.

34 Wang Q, Qi J, Yuan Y, *et al*. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 2014; **16**:328-337.

35 Raj VS, Mou H, Smits SL, *et al*. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013; **495**:251-254.

36 Yang Y, Du L, Liu C, *et al*. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci USA* 2014; **111**:12516-12521.

37 Li X, Zheng S, Agard DA, Cheng Y. Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. *J Struct Biol* 2015; **192**:174-178.

38 Tang G, Peng L, Baldwin PR, *et al*. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 2007; **157**:38-46.

39 Li X, Mooney P, Zheng S, *et al*. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 2013; **10**:584-590.

40 Rohou A, Grigorieff N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 2015; **192**:216-221.

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

42 Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 2014; **11**:63-65.

43 Pettersen EF, Goddard TD, Huang CC, *et al*. UCSF Chimera — a visualization system for exploratory research and analysis. *J Comput Chem* 2004; **25**:1605-1612.

44 Larkin MA, Blackshields G, Brown NP, *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**:2947-2948.

45 Song Y, DiMaio F, Wang RY, *et al*. High-resolution comparative modeling with RosettaCM. *Structure* 2013; **21**:1735-1742.

46 DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol* 2009; **392**:181-190.

47 Adams PD, Afonine PV, Bunkoczi G, *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.

48 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:486-501.

49 Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr* 2015; **71**:136-153.

50 Amunts A, Brown A, Bai XC, *et al*. Structure of the yeast mitochondrial large ribosomal subunit. *Science* 2014; **343**:1485-1489.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)