




# Cryo-EM structures of HKU2 and SADS-CoV spike glycoproteins provide insights into coronavirus evolution

Jinfang Yu <sup>1</sup>, Shuyuan Qiao <sup>1</sup>, Runyu Guo<sup>1</sup> & Xinquan Wang <sup>1</sup>✉

Porcine coronavirus SADS-CoV has been identified from suckling piglets with severe diarrhea in southern China in 2017. The SADS-CoV genome shares ~95% identity to that of bat  $\alpha$ -coronavirus HKU2, suggesting that SADS-CoV may have emerged from a natural reservoir in bats. Here we report the cryo-EM structures of HKU2 and SADS-CoV spike (S) glycoprotein trimers at 2.38 Å and 2.83 Å resolution, respectively. We systematically compare the domains of HKU2 spike with those of  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -coronavirus spikes, showing that the S1 subunit N- and C-terminal domains of HKU2/SADS-CoV are ancestral domains in the evolution of coronavirus spike proteins. The connecting region after the fusion peptide in the S2 subunit of HKU2/SADS-CoV adopts a unique conformation. These results structurally demonstrate a close evolutionary relationship between HKU2/SADS-CoV and  $\beta$ -coronavirus spikes and provide insights into the evolution and cross-species transmission of coronaviruses.

<sup>1</sup>The Ministry of Education Key Laboratory of Protein Science, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structure, Collaborative Innovation Center for Biotherapy, School of Life Sciences, Tsinghua University, 100084 Beijing, China.

✉email: [xinquanwang@tsinghua.edu.cn](mailto:xinquanwang@tsinghua.edu.cn)

Coronaviruses, categorized into the order *Nidovirales*, family *Coronaviridae*, and subfamily *Coronavirinae*, are a large group of viral pathogens with a wide host range<sup>1</sup>. Their infections in humans, other mammals, and birds can cause respiratory, hepatic, enteric and neurological diseases with varying severity<sup>2</sup>. The zoonotic transmission of novel coronaviruses into humans poses severe threats to the global health, evidenced by the severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and the SARS-CoV-2 that is causing the ongoing pandemic COVID-19 in the world<sup>1,3–6</sup>. In the meantime, coronaviruses infecting domestic animals also bring substantial economic losses<sup>7</sup>. For example, the swine acute diarrhea syndrome coronavirus (SADS-CoV) (also known as SeACoV and PEAV) isolated in 2017 caused outbreaks of severe watery diarrhea of suckling piglets with a mortality up to 90% in several commercial pig farms in Guangdong Province of China<sup>8–13</sup>. SADS-CoV is an  $\alpha$ -coronavirus and other representative members in the  $\alpha$ -genus are porcine epidemic diarrhea virus (PEDV), porcine transmissible gastroenteritis coronavirus (TGEV), porcine respiratory coronavirus (PRCV), feline infectious peritonitis virus (FIPV), and human NL63 and 229E coronaviruses (HCoV-NL63 and HCoV-229E)<sup>1</sup>. Representative members in other three genera include mouse hepatitis coronavirus (MHV), bovine coronavirus (BCoV), SARS-CoV, SARS-CoV-2, MERS-CoV, HCoV-OC43, and HCoV-HKU1 in the  $\beta$ -genus, avian infectious bronchitis virus (IBV) in the  $\gamma$ -genus, and porcine deltacoronavirus (PdCoV) in the  $\delta$ -genus<sup>1</sup>.

Cross-species transmission promoted by genetic recombination and/or mutations underlies the host range expansion of coronaviruses<sup>3,14–16</sup>. Bats are the natural reservoir of more than 30 different  $\alpha$ - and  $\beta$ -coronaviruses that have great potential for interspecies transmission by recombination and/or mutation<sup>15,17–19</sup>. Data on genetic evolution, receptor binding, and pathogenesis have demonstrated that human SARS-CoV and MERS-CoV most likely originate from bats<sup>1</sup>. Palm civets and dromedary camels are the intermediate hosts of SARS-CoV and MERS-CoV from bats to humans, respectively<sup>1</sup>. The newly identified porcine SADS-CoV isolates are also found to share ~95% sequence identity with *Rhinolophus* bat coronavirus HKU2, and this further stressed the severe results of coronavirus spillover from bats to domestic animals<sup>8–13</sup>. However, the molecular mechanisms underlying the transmission of SADS-CoV from bats to pigs are still unknown and need to be further explored. Recently it was shown that SADS-CoV is able to infect cells from a broad range of species including mouse, chicken, pig, monkey, and human, indicating a high potential of the SADS-CoV for interspecies transmission<sup>20</sup>.

The spike glycoprotein of coronaviruses mediates viral entry by binding host receptor with the S1 subunit and fusing viral and cellular membranes with the S2 subunit, thereby determining viral host range and tissue tropism<sup>21,22</sup>. As a class I viral fusion protein, the spike exists on the envelope of virion as a homotrimer and each monomer contains more than 1000 amino acid residues that can be cleaved into S1 and S2 subunits<sup>21</sup>. For most coronaviruses, the N-terminal domain (NTD) of the S1 subunit recognizes cell-surface carbohydrates, while the C-terminal domain (CTD) specifically binds to cellular protein receptors<sup>21–23</sup>. SARS-CoV, SARS-CoV-2, and HCoV-NL63 utilize CTD to bind human receptor ACE2<sup>24–30</sup>; MERS-CoV utilizes CTD to bind human receptor DPP4<sup>31,32</sup>; and TGEV, PRCV, and 229E utilize CTD to bind receptor APN<sup>33–35</sup>; HCoV-OC43 utilizes NTD to recognize glycans<sup>36</sup>; and one exception is MHV, which utilizes the NTD to bind mouse receptor CEACAM1a<sup>37,38</sup>. Therefore, the S1 subunit, especially its NTD and CTD, is the most variable region of the spike, and is responsible for different tropisms of

coronaviruses. In comparison, the S2 subunit containing the fusion peptide (FP) and heptad repeats (HR1 and HR2) for membrane fusion are more conserved in both sequence and structure<sup>21,22</sup>. For the SADS-CoV, receptor analysis indicated that none of the known coronavirus protein receptors including ACE2, DPP4, and APN are essential for cell entry<sup>10,20</sup>. There are also no reports regarding the recognition of glycans by the NTD of SADS-CoV.

Structural studies of the spike and its binding with glycans and protein receptors have provided important insights into the origin, evolution, and interspecies transmission of coronaviruses. Cryo-EM structures of spike trimer from all four coronavirus genera have been reported: the  $\alpha$ -coronavirus spike structures are determined for HCoV-NL63<sup>39</sup>, HCoV-229E<sup>34</sup>, PEDV<sup>40</sup>, and FIPV<sup>41</sup>; the  $\beta$ -coronavirus spike structures are determined for MHV<sup>38,42,43</sup>, HCoV-HKU1<sup>44</sup>, HCoV-OC43<sup>36</sup>, SARS-CoV<sup>24,45–47</sup>, MERS-CoV<sup>45,48,49</sup>, and SARS-CoV-2<sup>50,51</sup>; and the  $\gamma$ -coronavirus spike structure is determined for IBV<sup>52</sup> and the  $\delta$ -coronavirus spike structure is determined for PdCoV<sup>53,54</sup>. The cryo-EM structures of bat coronavirus spike trimers have not been reported, and only crystal structures of the CTD from HKU4<sup>55</sup>, HKU5<sup>56</sup>, and HKU9<sup>57</sup> were determined.

The spikes of SADS-CoV (1130 amino acid residues) and HKU2 (1128 amino acid residues) are among the shortest coronavirus spike glycoproteins<sup>58</sup> and their amino acid identities to other known coronavirus spikes are lower than 28%, indicating the spikes of HKU2 and SADS-CoV are unique<sup>8–13,59</sup>. In this study, we report the cryo-EM structures of the SADS-CoV and HKU2 spike trimers at 2.83 and 2.38 Å resolution, respectively. The HKU2 spike trimer structure is the first one from bat coronavirus. We analyze the HKU2 and SADS-CoV trimer structures and also compare the NTD, CTD, SD1, and SD2 domains of the S1 subunit and the S2 subunit of HKU2 with other spikes from  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -coronaviruses. Our results strongly support that HKU2 and SADS-CoV preserve primitive structural features in their spikes that have a close evolutionary relationship with  $\beta$ -coronavirus spikes and provide insights into the evolution and cross-species transmission of coronaviruses.

## Results

**Protein expression and structure determination.** The cDNAs encoding HKU2 spike (YP\_001552236) and SADS-CoV spike (AVM41569.1) were synthesized with codons optimized for insect cell expression. HKU2 ectodomains (residues 1–1066) and SADS-CoV ectodomains (residues 1–1068) were separately cloned into pFastBac-Dual vector (Invitrogen) with C-terminal foldon tag and Strep tag. After expression in Hi5 insect cells and purification to homogeneity, the cryo-EM images on these two spike ectodomains were recorded using FEI Titan Krios microscope operating at 300 kV with a Gatan K2 Summit direct electron detector (Supplementary Fig. 1). About 1,400,000 particles for HKU2 spike and 900,000 particles for SADS-CoV spike were subjected to 2D classification, and a total of 421,490 particles of HKU2 spike and 152,334 particles of SADS-CoV spike were selected and subjected to 3D refinement with C3 symmetry to generate density maps (Supplementary Fig. 2). The overall density maps were solved to 2.38 Å for HKU2 spike and 2.83 Å for SADS-CoV spike (gold-standard Fourier shell correlation (FSC) = 0.143) (Supplementary Figs. 1–3). The atomic-resolution density map enabled us to build nearly all residues of HKU2 spike ectodomains (residues 17–995) except for a few breaks (residues 129–141 and 203–205), as well as 48 N-linked glycans and 71 water molecules (Supplementary Figs. 3a, 4a, and 5a). The final refined model of SADS-CoV spike contains residues 19–998 with some short breaks (residues 82–83, 134–143, and 488–490) and

**Table 1 Cryo-EM data collection, refinement, and validation statistics.**

	<b>HKU2 (EMDB-30037) (PDB 6M15)</b>	<b>SADS-CoV (EMDB-30038) (PDB 6M16)</b>
Data collection and processing		
Magnification	×130,000	×130,000
Voltage (kV)	300	300
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	49.784	49.784
Defocus range (μm)	-1 to -3	-1 to -3
Pixel size (Å)	1.061	1.061
Symmetry imposed	C3	C3
Initial particle images (no.)	-1,400,000	-900,000
Final particle images (no.)	421,490	152,334
Map resolution (Å)	2.38	2.83
FSC threshold	0.143	0.143
Map resolution range (Å)	2.38-4	2.83-6
Refinement		
Initial model used (PDB code)	5SZ5	5SZ5
Model resolution (Å)	2.38	2.83
FSC threshold	0.143	0.143
Model resolution range (Å)	2.38	2.83
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-87.05	-101.51
Model composition		
Nonhydrogen atoms	23,471	23,433
Protein residues	2895	2895
Ligands	72	60
<i>B</i> factors (Å <sup>2</sup> )		
Protein	25.63	55.82
Ligand	50.21	65.92
R.m.s. deviations		
Bond lengths (Å)	0.004	0.008
Bond angles (°)	0.624	0.763
Validation		
MolProbity score	1.69	2.39
Clashscore	3.88	7.99
Poor rotamers (%)	3.46	3.97
Ramachandran plot		
Favored (%)	97.5	91.68
Allowed (%)	2.5	8.22
Disallowed (%)	0	0.1

45 N-linked glycans (Supplementary Figs. 3b, 4b, and 5b). Data collection and refinement statistics for these two structures are listed in Table 1.

**Overall structures of HKU2 and SADS-CoV spikes.** The overall structures of HKU2 and SADS-CoV spikes we determined here resemble the previously reported prefusion structures of coronaviruses spikes. Both spike trimers have a mushroom-like shape (~150 Å in height and ~115 Å in width) (Fig. 1a), consisting of a cap mainly formed by β-sheets of the S1 subunit, a central stalk mainly formed by α-helices of the S2 subunit, and a root formed by twisted β-sheets and loops of the S2 subunit (Fig. 1a). In each trimer there is a C3 axis along the central stalk (Fig. 1a). The amino acid identity between HKU2 and SADS-CoV spikes is 86%, and these two spike structures are quite similar with the root mean square deviation (r.m.s.d.) being 0.53 Å for 962 aligned Ca atoms of the monomer and 0.56 Å for 2886 aligned Ca atoms of the trimer. Due to the high structural similarity, we will use the HKU2 structure to present the features

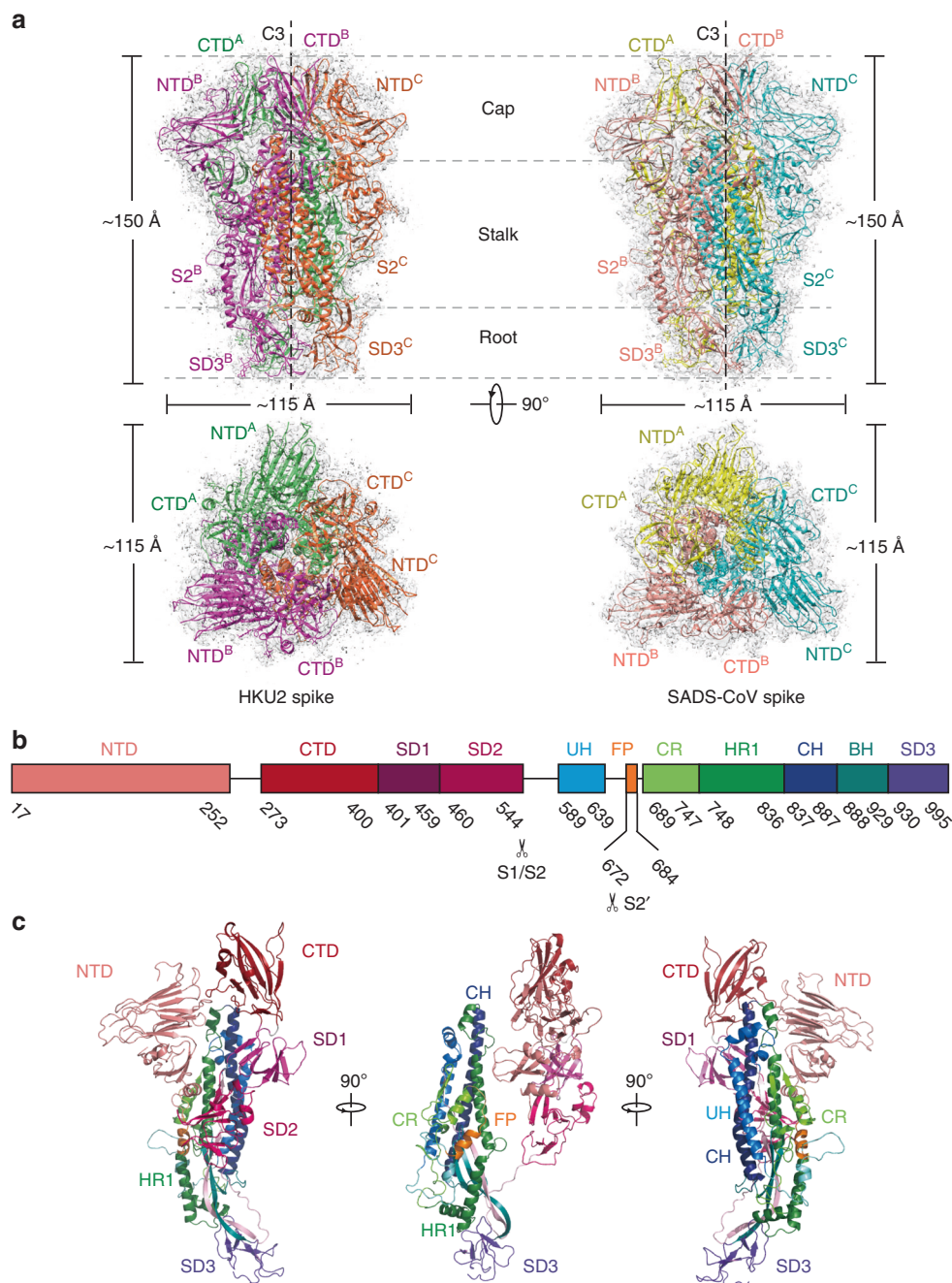
of both spikes in the subsequent description, whereas significant differences between them will be pointed out only when necessary.

The S1 subunit of the HKU2 spike comprises two major domains, NTD and CTD, which are followed by two subdomains 1 and 2 (SD1 and SD2) connecting them to the S2 subunit (Fig. 1b, c). The S1 subunits from three monomers form the cap of the spike, in which the three CTDs in the inner part are at the apex sitting on top of the central stalk and the three NTDs are located outside the CTDs surrounding the central stalk (Fig. 1a). The NTD, CTD, SD1, and SD2 of the S1 subunit are all mainly composed of β strands (Fig. 1c). In contrast, the upstream helix (UH), FP, connecting region (CR), heptad repeat 1 (HR1), and central helix (CH) of the S2 subunit are mainly composed of helices, whereas the β-hairpin (BH) and subdomain 3 (SD3) at the bottom part of the S2 subunit mainly consist of β strands and loops (Fig. 1c). Moreover, the residues after the SD3, which contain the heptad repeat 2 (HR2), are not resolved in the HKU2 and SADS-CoV spike structures, as well as in all other reported coronavirus spike structures in the prefusion state.

The SD1 and SD2 of the S1 subunit and the S2 subunit are highly similar in amino acid sequence (85, 84, and 95% identities) and structure (Ca r.m.s.d. less than 0.5 Å) between HKU2 and SADS-CoV spikes (Supplementary Fig. 6). The NTD has the lowest sequence identity of 70% among all domains. Structural superimposition also gave a higher Ca r.m.s.d. value of 1.2 Å between these two NTDs. The core β-sheet structure is structurally conserved and conformational variations reside in the loops (Supplementary Fig. 6a). The CTD sequence identity between HKU2 and SADS-CoV spikes is 82%, and the Ca r.m.s.d. between these two CTDs is 1.1 Å, also indicating more structural variations in the CTD in comparison with the highly conserved SD1, SD2, and S2 subunit. The NTD and CTD of the S1 subunit are commonly utilized by coronaviruses for binding cell-surface carbohydrates or protein receptors for cell attachment<sup>23</sup>. Therefore, the sequence and/or structural variations indicate that HKU2 and SADS-CoV would also bind different host receptors by NTD and/or CTD of the S1 subunit, although their receptors in bat and pig are unknown and the receptor-binding sites on spike have not been defined.

**NTD structure and comparisons.** The NTD of HKU2 has three layers of antiparallel β-sheet with the top one consisting of six strands, the middle one consisting of five strands, and the bottom one consisting of three strands. Below the bottom sheet is a short α-helix (Fig. 2a). The top and middle β-sheets form a galectin-like β-sandwich fold, which is inserted between two strands of the bottom sheet (Fig. 2a). To supplement, three disulfide bonds are detected in the HKU2 NTD structure: C<sup>17</sup>-C<sup>56</sup> connecting the N-terminus of the NTD to its upper loop, C<sup>124</sup>-C<sup>149</sup> connecting β6 and β7 strands in the top sheet, and C<sup>234</sup>-C<sup>244</sup> connecting the bottom helix to the bottom sheet (Fig. 2a).

Although the NTDs of all coronaviruses adopt a similar overall architecture, the NTD of HKU2 has the highest structural similarity with the NTD1 (named domain 0 in previous reports) of α-coronavirus HCoV-NL63 with an r.m.s.d. of 2.7 Å for 186 aligned Ca atoms (Fig. 2b and Supplementary Fig. 8d). The NTD1 of α-coronavirus PEDV is not completely modeled in the spike trimer structure; however, the partial model still fits well with the NTD of HKU2 with an r.m.s.d. of 2.3 Å for 73 aligned Ca atoms. Both HCoV-NL63 and PEDV have a second NTD (NTD2, also named domain A in previous reports), and the NTD of HKU2 is structurally less similar to the NTD2 with an r.m.s.d. of 4.3 Å against HCoV-NL63 NTD2 and of 4.1 Å against PEDV NTD2 (Fig. 2b and Supplementary Fig. 8d). Recent structural

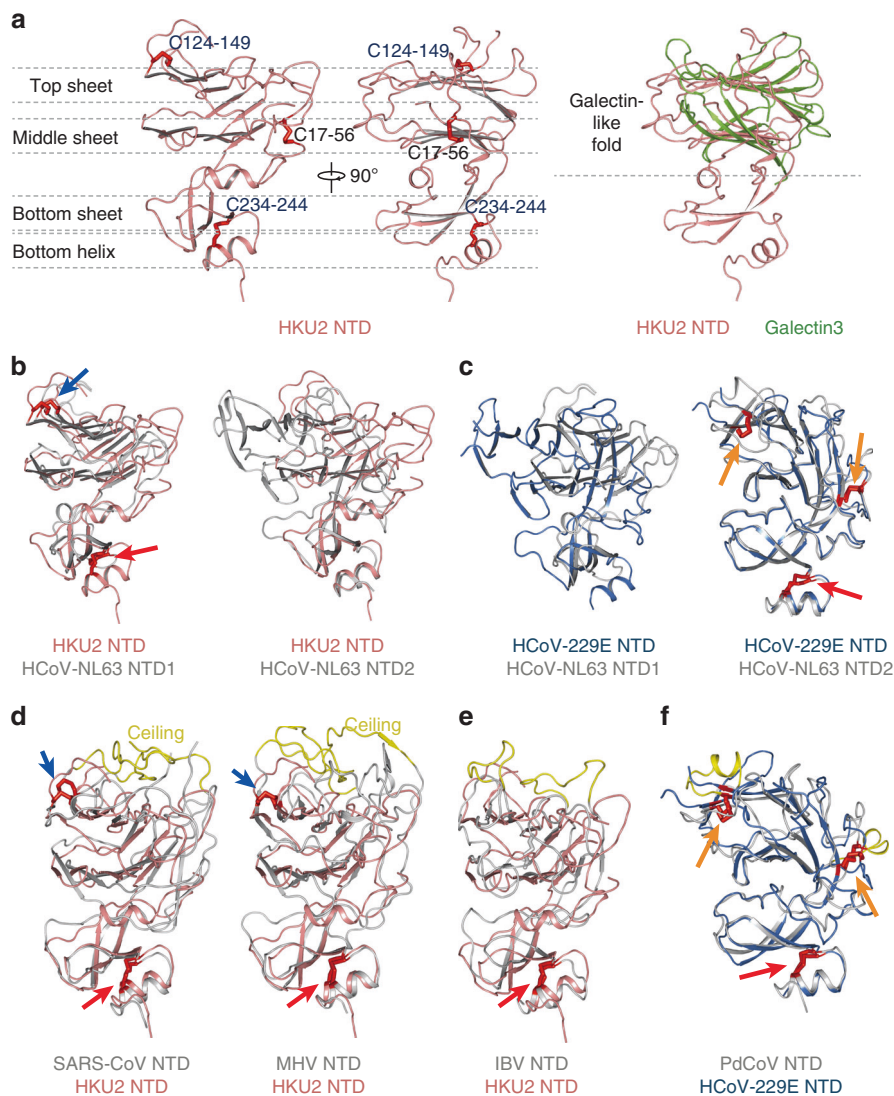


**Fig. 1 Overall structures of HKU2 and SADS-CoV spike glycoproteins.** **a** Overall structures of HKU2 and SADS-CoV spike glycoproteins shown in side view (upper panel) and top view (lower panel). Three monomers of HKU2 spike are colored magenta, green, and orange, respectively; three monomers of SADS-CoV spike are colored pink, yellow, and cyan, respectively. The cryo-EM maps are shown as semitransparent surface and contoured at 2.6 RMS and 3 RMS for HKU2 and SADS-CoV spikes, respectively. The trigonal axes are shown as black dashed lines. Visible segments of each monomer are labeled accordingly. The cap, stalk, and root parts are partitioned by gray dashed lines. **b** Segmentation of HKU2 monomer. The segments of HKU2 are shown as boxes with the width related to the length of amino acid sequence. The start and end amino acids of each segment are labeled. The position of S1/S2, and S2' cleavage sites are indicated. NTD N-terminal domain, CTD C-terminal domain, SD1 subdomain 1, SD2 subdomain 2, UH upstream helix, FP fusion peptide, CR connecting region, HR1 heptad repeat 1, CH central helix, BH  $\beta$ -hairpin, SD3 subdomain 3. **c** Overall structure of HKU2 monomer. Side views of HKU2 monomer shown in three directions. The segments are colored the same as in **b**.

determination showed that another  $\alpha$ -coronavirus, HCoV-229E, also has one NTD, which is more structurally similar to the NTD2 (Ca r.m.s.d. of 2.0 Å) than the NTD1 (Ca r.m.s.d. of 3.7 Å) of HCoV-NL63 (Fig. 2c). All above comparisons indicate that there are two subtypes of NTD in the  $\alpha$ -coronaviruses: the subtype I is represented by the NTDs of HKU2, SADS-CoV, the NTD1s of HCoV-NL63 and PEDV, and the subtype II is

represented by the NTD of HCoV-229E, the NTD2s of HCoV-NL63 and PEDV (Fig. 2b, c). Although sharing an overall architecture, these two NTD subtypes have a structural difference in the galectin-like  $\beta$ -sandwich fold containing the top and middle sheets stacked together through hydrophobic interactions. These two  $\beta$ -sheets are well aligned in the galectin-like domain of subtype I, whereas there is an alignment shift in the galectin-like





**Fig. 2 Structure of HKU2 NTD and comparisons.** **a** Structure of HKU2 NTD. Side views of HKU2 NTD are shown in two orthogonal directions. Disulfide bonds are shown as red sticks. Conserved disulfide bonds are labeled blue; other disulfide bonds are labeled black. Top sheet, middle sheet, bottom sheet, and bottom helix are partitioned by gray dashed lines. Comparison of HKU2 NTD and galectin3 are shown in the right panel. HKU2 NTD is colored salmon; galectin3 is colored green. PDB code: galectin3, 1A3K. **b** Subtype I NTD of  $\alpha$ -coronavirus. Structural alignments of HKU2 NTD with HCoV-NL63 NTD1, and with HCoV-NL63 NTD2 are shown. HKU2 NTD is colored salmon. HCoV-NL63 NTD1 and NTD2 are colored gray. Disulfide bonds are shown as red sticks. Disulfide bonds conserved in both types of NTDs are indicated by red arrows; disulfide bonds conserved in subtype I NTD are indicated by blue arrows; PDB code: HCoV-NL63, 5S2S. **c** Subtype II NTD of  $\alpha$ -coronavirus. Structural alignments of HCoV-229E NTD with HCoV-NL63 NTD1, and with HCoV-NL63 NTD2 are shown. HCoV-229E NTD is colored marine. HCoV-NL63 NTD1 and NTD2 are colored gray. Disulfide bonds conserved in both types of NTDs are indicated by red arrows; disulfide bonds conserved in subtype II NTD are indicated by orange arrows. PDB codes: HCoV-229E, 6U7H. **d**  $\beta$ -coronavirus NTDs resemble subtype I. Structural alignments of HKU2 NTD with SARS-CoV NTD, and with MHV NTD are shown. Disulfide bonds are shown and labeled the same as in **b**. Ceilings in  $\beta$ -coronavirus NTDs are shown in yellow. PDB codes: SARS, 5XLR; MHV, 3JCL. **e**  $\gamma$ -coronavirus IBV NTD resembles subtype I. Structural alignment of HKU2 NTD with IBV NTD is shown. Disulfide bonds are shown and labeled the same as in **b**. The additional loops in IBV NTD is shown in yellow. PDB code: IBV, 6CVO. **f**  $\delta$ -coronavirus PdCoV NTD resembles subtype II. Structural alignment of HCoV-229E NTD with PdCoV NTD is shown. Disulfide bonds are shown and labeled the same as in **c**. The additional helices in PdCoV NTD are shown in yellow. PDB codes: HCoV-229E, 6U7H; PdCoV, 6B7N.

domain of subtype II (Fig. 2b). The other notable difference is the distribution of signature disulfide bonds. A signature disulfide bond C<sup>124</sup>-C<sup>149</sup> (numbered in HKU2 and connecting  $\beta$ 6 and  $\beta$ 7 strands in the top sheet) is conserved in all subtype I NTDs (Fig. 2b), and the subtype II NTDs have two signature disulfide bonds: the first one C<sup>145</sup>-C<sup>168</sup> (numbered in HCoV-229E) connecting  $\beta$ 5 and  $\beta$ 6 strands and the second one C<sup>81</sup>-C<sup>105</sup> (numbered in HCoV-229E) connecting  $\beta$ 2 and neighbor loop (Fig. 2c). Besides, C<sup>234</sup>-C<sup>244</sup> (numbered in HKU2 and connecting

the bottom helix to the bottom sheet) is conserved in both subtype I and subtype II NTDs (Fig. 2b, c).

The NTDs of  $\beta$ -coronaviruses including BCoV, HCoV-HKU1, HCoV-OC43, MERS-CoV, SARS-CoV, and MHV resemble the subtype I, rather than the subtype II NTD in the topology and distribution of the disulfide bonds (Fig. 2d). These  $\beta$ -coronavirus NTDs have additional loops in the N-terminus, between  $\beta$ 1 and  $\beta$ 2 strands, and between  $\beta$ 6 and  $\beta$ 7 strands (numbered in HKU2 structure), forming an extensive ceiling-like structure on

top of the galectin-like fold (Fig. 2d). It has been found that the involvement of this ceiling-like structure has functional outcomes such as immune evasion or receptor binding<sup>52</sup>. The NTD of  $\gamma$ -coronavirus IBV also resemble the subtype I NTD in the topology, although its disulfide bond positions are not conserved as in the subtype I NTD (Fig. 2e). Of note, the NTD of  $\delta$ -coronavirus PdCoV resemble the subtype II NTD in the topology and distribution of the disulfide bonds (Fig. 2f). Both IBV and PdCoV NTDs also have additional insertions including loops and short helices in the galectin-like fold compared with the two subtypes of NTD in  $\alpha$ -coronaviruses (Fig. 2e, f).

**CTD structure and comparisons.** The CTD of HKU2 has a twisted five-stranded antiparallel  $\beta$ -sheet as the core with connecting loops between the strands (Fig. 3a). It contains four disulfide bonds: C<sup>277</sup>–C<sup>300</sup> and C<sup>285</sup>–C<sup>290</sup> at the N-terminus, C<sup>341</sup>–C<sup>397</sup> at the C-terminus, and the last one C<sup>331</sup>–C<sup>369</sup> connecting the  $\beta$ 2 and  $\beta$ 5 strands in the core  $\beta$ -sheet (Fig. 3a). Interestingly, the CTD core of HKU2 is of high structural similarity with the conserved CTD core of  $\beta$ -coronaviruses and the disulfide bonds in the CTD of HKU2 except for C<sup>285</sup>–C<sup>290</sup> are also detected in all  $\beta$ -coronavirus CTDs (Fig. 3b). These CTDs have the core of one twisted  $\beta$ -sheet and here we name them as one-layer CTD subtype (Fig. 3a, b). The  $\beta$ -coronavirus CTDs always have an insertion consisting of loops and/or strands between the  $\beta$ 5 and  $\beta$ 6 strands of the core (Fig. 3b). SARS-CoV<sup>26</sup>, SARS-CoV-2<sup>27–30</sup>, MERS-CoV<sup>31,32</sup>, HKU4<sup>55</sup>, and HKU5<sup>56</sup> have receptor-binding motif (RBM) in this insertion region responsible for binding their respective protein receptors. In the CTD of HKU2, there is only one short loop between the  $\beta$ 5 and  $\beta$ 6 strands of the core twisted  $\beta$ -sheet (Fig. 3a).

Although as members in the  $\alpha$ -genus, HKU2, and SADS-CoV CTD structures are significantly different from those of other  $\alpha$ -coronaviruses HCoV-NL63, HCoV-229E, PEDV, TGEV, and PRCV that contain two layers of  $\beta$ -sheets (Fig. 3c). And we named these CTDs as two-layer CTD subtype. All available two-layer CTD structures can be well aligned with Ca r.m.s.d. in the range of 1.0–3.4 Å. These two-layer CTDs contain two highly conserved disulfide bonds: C<sup>540</sup>–C<sup>586</sup> and C<sup>569</sup>–C<sup>596</sup> (numbered in PEDV CTD) (Fig. 3c). The C<sup>569</sup>–C<sup>596</sup> is conserved among all coronaviruses, whereas the C<sup>540</sup>–C<sup>586</sup> is conserved in all  $\alpha$ -coronaviruses (except for HKU2 and SADS) and  $\delta$ -coronavirus PdCoV (Fig. 3c).

The CTD of  $\delta$ -coronavirus PdCoV has a core of two  $\beta$ -sheets, belonging to the two-layer CTD subtype (Fig. 3c). As for the  $\gamma$ -coronavirus IBV, the core of its CTD is also similar to the typical two-layer CTD (Fig. 3c). However, several  $\beta$  strands are replaced by loops and the disulfide bonds are in different positions from the two-layer CTD (Fig. 3c). IBV CTD also has an extra region of loops, reminiscent of the extra domain in the CTDs of  $\beta$ -coronaviruses (Fig. 3c).

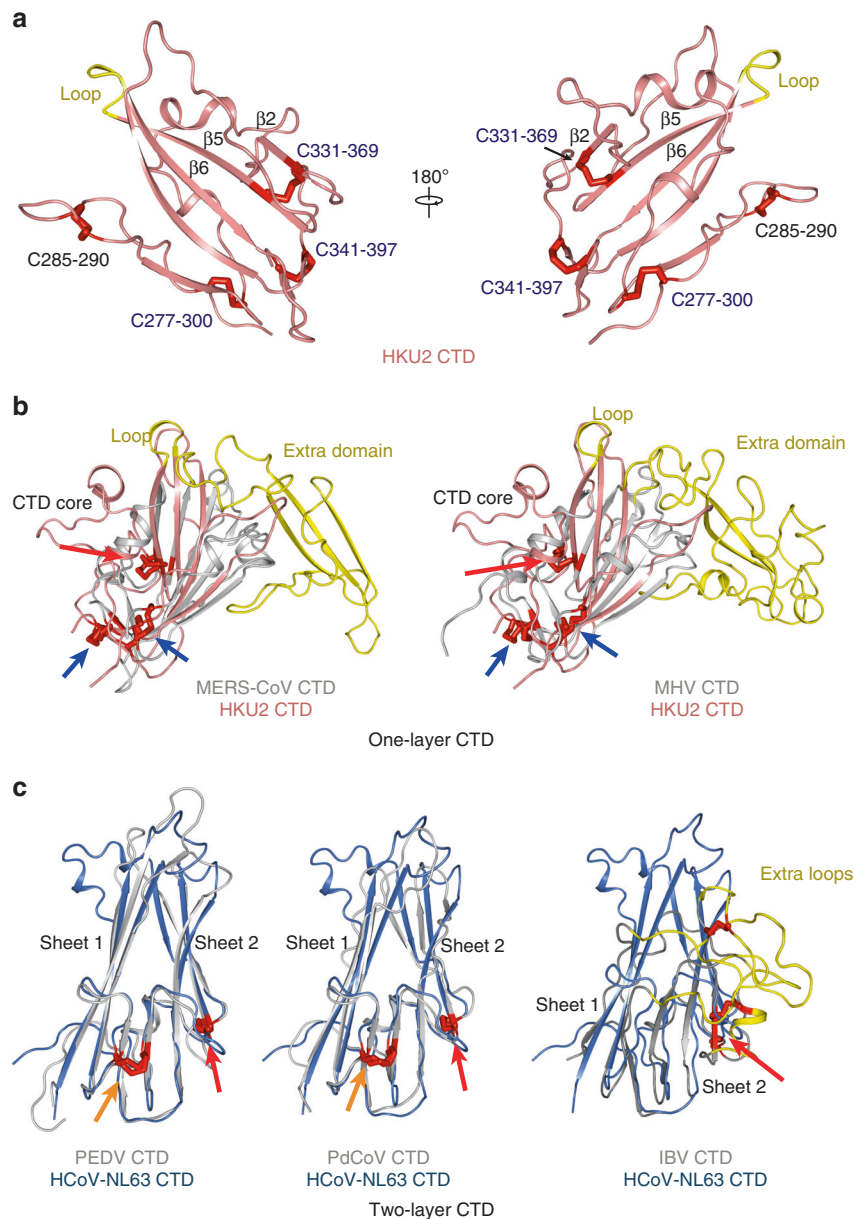
**SD1 and SD2 structures and comparisons.** The SD1 and SD2 are two subdomains following the CTD in the S1 subunit, linking the CTD to the S2 subunit. The HKU2 SD1 is a partial  $\beta$  barrel consisting of five  $\beta$  strands and a disulfide bond (C<sup>409</sup>–C<sup>458</sup>) connecting its C-terminus to the  $\beta$ 1 strand (Fig. 4a). This five-stranded  $\beta$  barrel and the linking disulfide bond are conserved among all four genera of coronavirus (Fig. 4a). The HKU2 SD2 has a structure of two layers of  $\beta$ -sheet with an additional short  $\alpha$ -helix over the top sheet (Fig. 4b). The additional  $\alpha$ -helix and the top sheet is linked by a disulfide bond (C<sup>482</sup>–C<sup>509</sup>), and another disulfide bond (C<sup>524</sup>–C<sup>533</sup>) links the C-terminal loop to the bottom sheet (Fig. 4b). The two-layer core structure and the second disulfide bond are conserved among all genera of

coronavirus; however, the additional  $\alpha$ -helix and the first linking disulfide bond is a distinct feature of  $\beta$ -coronaviruses plus  $\alpha$ -coronavirus HKU2 and SADS-CoV (Fig. 4b). This additional helix appears to be an insertion between the  $\beta$ 2 and  $\beta$ 3 strands of the SD2, and is retained during evolution of  $\beta$ -coronaviruses.

**Quaternary packing of the NTD and CTD in the spike.** It has been observed that coronaviruses have two types of quaternary packing mode of the S1 subunits in the trimer: intrasubunit packing and cross-subunit packing<sup>52</sup>. Actually, this is mainly caused by different positioning and interaction between NTD and CTD in the spike monomer. The HKU2 S1 subunit, similar to those in  $\alpha$ -coronaviruses HCoV-NL63, HCoV-229E, and PEDV and  $\delta$ -coronavirus PdCoV, has an inward CTD that contacts with the NTD (Fig. 5a). The three structural NTD–CTD modules in the cap region of these spikes are composed of NTD and CTD from the same monomer, forming the intrasubunit packing in the spike trimer (Fig. 5a). The S1 subunits of other coronaviruses in the  $\beta$ - and  $\gamma$ -genera including MHV, SARS-CoV, MERS-CoV, HCoV-OC43, and IBV have an outward CTD that is far away from the NTD (Fig. 5b). Therefore, the three structural NTD–CTD modules in the cap region of these spikes have the NTD from one monomer and the CTD from the adjacent monomer, forming the cross-subunit packing in the spike trimer (Fig. 5b). Interestingly, we found that the outward CTDs always have an insertion in the core structure, such as  $\beta$ -coronavirus CTDs and  $\gamma$ -coronavirus IBV CTD (Fig. 5b). In contrast, all inward CTDs only have the one-layer or two-layer core structure without obvious inserted region.

**Conserved S2 subunit and a distinct CR.** Sequence analysis suggested that the S1/S2 protease cleavage site at the boundary between the S1 and S2 subunits is R544–M545 in HKU2 spike and R546–M547 in SADS-CoV spike<sup>8,11,59</sup>. Compared with the S1 subunit, the topology and structure of S2 subunit are highly conserved in all coronavirus spikes. The HKU2 S2 subunit contains a 3-helix UH (residues 589–639), an FP helix (residues 672–684), a CR (residues 689–747), a 4-helix HR1 (residues 748–836), a CH (residues 837–887), a twisted BH (residues 888–929), and a  $\beta$ -sandwich like SD3 (residues 930–995) (Figs. 1b and 6a). Like in other coronavirus spikes in the prefusion state, the model of HR2 after SD3 was not built in the structure due to poor density. Five disulfide bonds in S2 are detected. Two of them (C<sup>590</sup>–C<sup>612</sup> and C<sup>595</sup>–C<sup>601</sup>) stabilize the folded helices of UH, C<sup>696</sup>–C<sup>706</sup> bends the CR, C<sup>884</sup>–C<sup>895</sup> links the CH and the BH, and C<sup>934</sup>–C<sup>943</sup> is within the SD3 (Fig. 6a). The first four disulfide bonds are conserved in all coronaviruses, and the last one in the SD3 has different positions in different spikes. Specifically, it links the  $\beta$ 2 and  $\beta$ 3 strands of SD3 in the spikes of HKU2, SADS-CoV, and MERS-CoV (numbered in MERS-CoV), and in other coronavirus spikes it links the  $\beta$ 2 strand to the C-terminal loop of SD3 (numbered in MERS-CoV) (Supplementary Fig. 7).

All coronavirus spikes have the S2' protease site upstream from the FP in the S2 subunit, which is essential for proteolytic fusion activation of the spike. Receptor binding and cleavage at the S2' site promote large-scale conformational changes of the FP, CR, HR1, and HR2, allowing the insertion of FP into host cell membrane and the formation of six-helix bundle. The FP and CR, which are often not well and totally resolved in other coronavirus spike structures, can be clearly modeled in the HKU2 spike due to the atomic resolution of the map (Supplementary Fig. 5). The typical CR in the S2 subunit contains three helices and one short strand, with a disulfide bond bending the first and second helix to form a turn (Fig. 6b). In HKU2, the second helix is replaced by a



**Fig. 3 Structure of HKU2 CTD and comparisons.** **a** HKU2 CTD shown in two opposite directions. Strands mentioned in the main text are labeled. The loop replaced by extra domain in  $\beta$ -coronavirus CTDs is shown in yellow. Disulfide bonds are shown as red sticks. Conserved disulfide bonds are labeled blue; other disulfide bonds are labeled black. **b**  $\beta$ -coronavirus CTDs belong to one-layer CTD. Structural alignments of HKU2 CTD with MERS-CoV CTD, and with MHV CTD are shown. Extra domains in  $\beta$ -coronavirus CTDs are colored yellow. Disulfide bonds are shown as red sticks. Disulfide bonds conserved in both one-layer CTD and two-layer CTD are indicated by red arrows. Disulfide bonds only conserved in one-layer CTD are indicated by blue arrows. PDB codes: MHV, 3JCL; MERS-CoV, 6Q05. **c**  $\alpha$ -coronavirus (except for HKU2 and SADS-CoV),  $\gamma$ -coronavirus, and  $\delta$ -coronavirus CTDs belong to two-layer CTD. Structural alignments of HCoV-NL63 CTD with PEDV CTD, with PdCoV CTD, and with IBV CTD are shown. Two layers of  $\beta$ -sheets are labeled. Extra loops in IBV CTD are colored yellow. Disulfide bonds are shown as red sticks. Disulfide bonds conserved in both one-layer CTD and two-layer CTD are indicated by red arrows. Disulfide bonds only conserved in two-layer CTD are indicated by orange arrows. PDB codes: HCoV-NL63, 5SZS; PEDV, 6U7K; IBV, 6CV0; PdCoV, 6B7N.

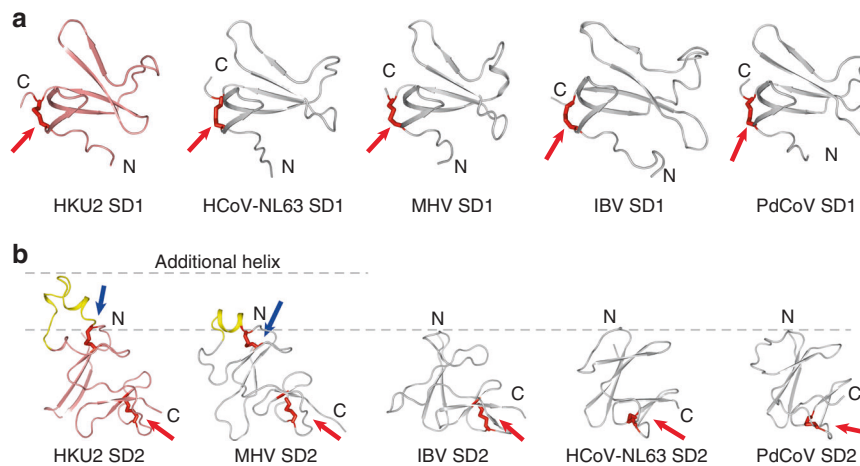
short strand (713–716) and the third helix is replaced by a loop (721–741), therefore there are two short strands and only one helix in HKU2 CR (Fig. 6c). The conserved disulfide bond C<sup>696</sup>–C<sup>706</sup> makes the first helix of CR in HKU2 spike turn upside down. The S2' cleavage site (between R671 and S672) is then covered by the reversed CR helix and loops, and R671 interacts with E723 in the loop and K697 and K698 in helix 1 (Fig. 6c). In other coronaviruses, taking the MHV S2 for example, the helix 1 does not cover the S2' site (between R869 and S870), and R869 only loosely interacts with T929 in helix 3 (Fig. 6b). After the

dissociation of the S1 subunit triggered by receptor binding, the exposure of the S2' site for cleavage is a prerequisite for the proteolytic activation of the coronavirus spike to mediate membrane fusion. The buried S2' site indicates that HKU2 spike, compared with other coronavirus spikes, would require more conformational changes around the S2' site for the exposure.

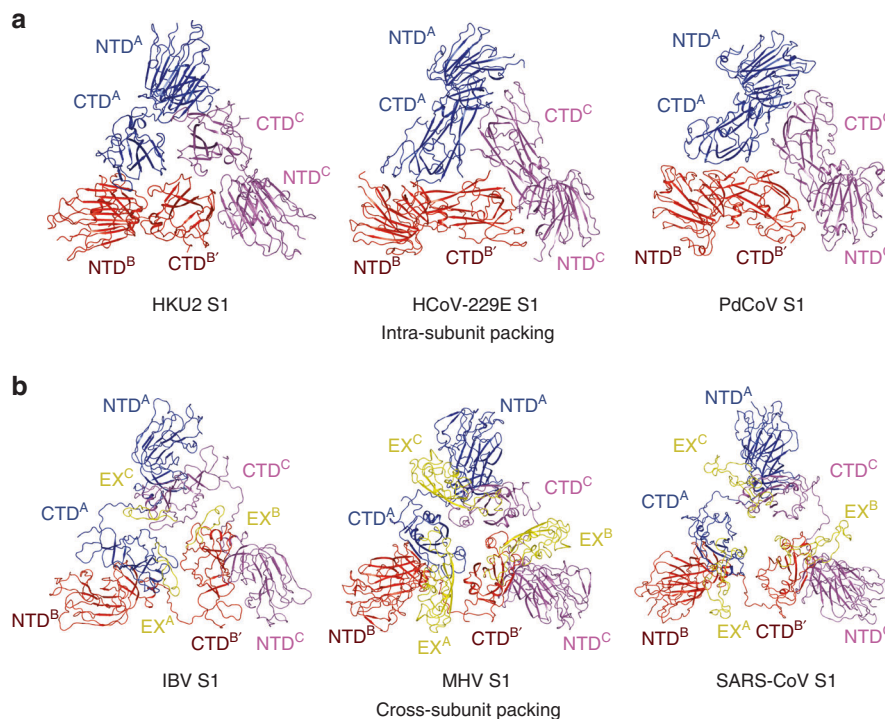
## Discussion

A porcine coronavirus SADS-CoV (also named SeACoV and PEAV in other reports) was recently identified from suckling





**Fig. 4 Structures of SD1 and SD2 and comparisons.** **a** Comparison of SD1 from four genera of coronaviruses. HKU2 SD1 is colored salmon; SD1 from other coronaviruses are colored gray. Disulfide bonds are shown as red sticks. Red arrows indicate the disulfide bonds conserved in all genera of coronaviruses. PDB codes: HCoV-NL63, 5SZS; MHV, 3JCL; IBV, 6CV0; PdCoV, 6B7N. **b** Comparison of SD2 from four genera of coronaviruses. HKU2 SD2 is colored salmon; SD2 from other coronaviruses are colored gray. Disulfide bonds are shown as red sticks. Red arrows indicate the disulfide bonds conserved in all genera of coronaviruses. Blue arrows indicate the disulfide bonds only found in HKU2 (and SADS-CoV) and  $\beta$ -CoVs. The additional helices of SD2 from HKU2 (and SADS-CoV) and  $\beta$ CoVs are colored yellow and partitioned by gray dashed lines. PDB codes: HCoV-NL63, 5SZS; MHV, 3JCL; IBV, 6CV0; PdCoV, 6B7N.

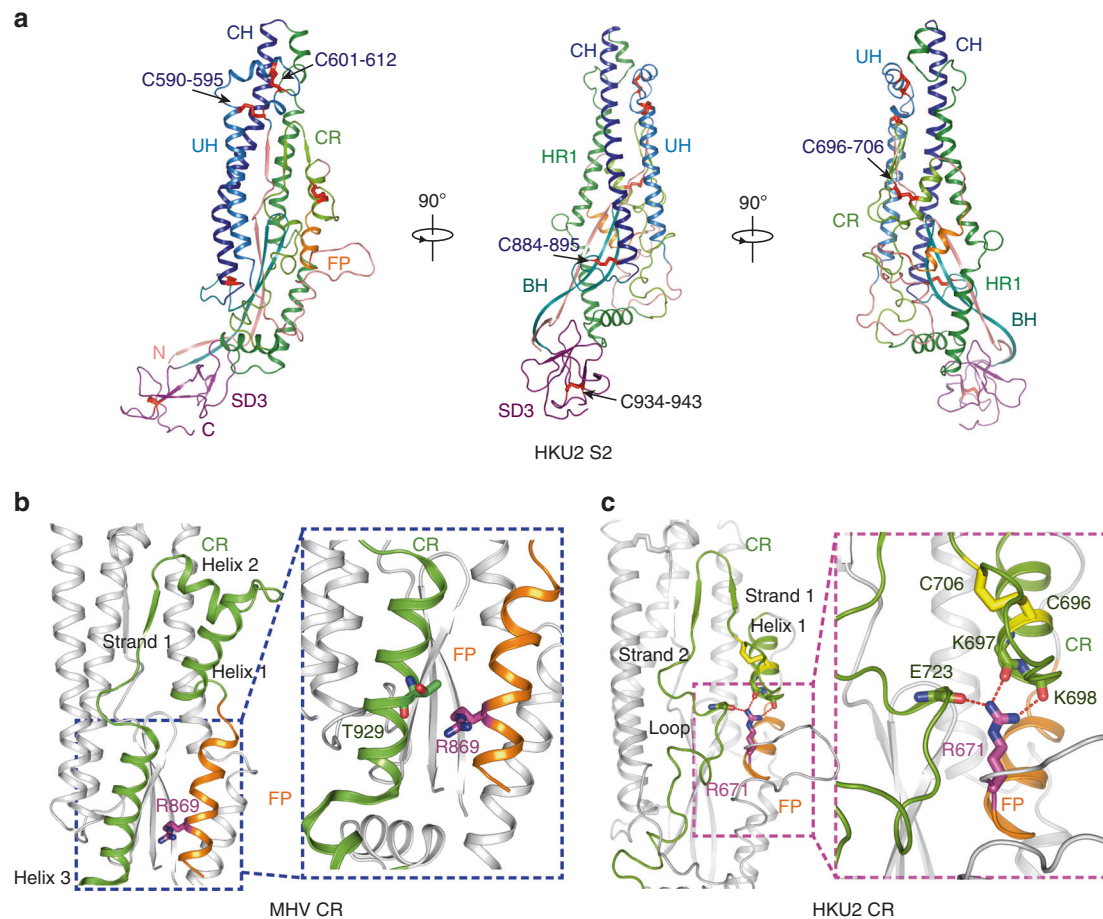


**Fig. 5 Quaternary packing of NTD and CTD.** **a**  $\alpha$ -coronavirus S1 and  $\delta$ -coronavirus S1 use intrasubunit packing pattern. NTD and CTD from the first monomer are colored blue, the second are colored red, and the third are colored magenta. PDB codes: HCoV-229E, 6U7H; PdCoV, 6B7N. **b**  $\beta$ -coronavirus S1 and  $\gamma$ -coronavirus S1 use cross-subunit packing pattern. NTD and CTD from the first monomer are colored blue, the second are colored red, and the third are colored magenta. The extra loop of IBV and the extra domains of  $\beta$ -CoVs are colored yellow and labeled as EX. PDB codes: MHV, 3JCL; IBV, 6CV0; SARS, 5XLR.

piglets with diarrhea in southern China, and its genome sequence was most identical (~95% identity) to that of *Rhinolophus* bat  $\alpha$ -coronavirus HKU2<sup>8–13</sup>. The SADS-CoV and HKU2 are phylogenetically located in a sublineage closely related to the proposed  $\alpha$ -coronavirus group-1b lineage at the complete genome level<sup>8–13</sup>. However, phylogenetic analysis based on the spike glycoprotein indicated that they are members of a separate lineage clustered

within  $\beta$ -coronavirus (Supplementary Fig. 8a), suggesting that HKU2 and SADS-CoV probably resulted from recombination of an  $\alpha$ -coronavirus with an unrecognized  $\beta$ -coronavirus S gene<sup>8–13</sup>. These results, together with the lower than 28% amino acid identities to other known coronavirus glycoproteins, strongly indicate that the spike glycoproteins of HKU2 and SADS-CoV are unique<sup>8–13</sup>. In this study, we determined the cryo-EM structures





**Fig. 6 Structure of HKU2 S2.** **a** Side views of HKU2 S2 shown in three directions. Seven segments of S2 are shown as different colors. UH upstream helix, FP fusion peptide, CR connecting region, HR1 heptad repeat 1, CH central helix, BH  $\beta$ -hairpin, SD3 subdomain 3. Disulfide bonds are shown as red sticks. Disulfide bonds conserved in all coronaviruses are labeled blue; the other disulfide bond is labeled black. **b** Conserved CR represented by MHV CR. CR and FP are colored the same as in **a**. Helices and strands in CR are labeled. R869 (S2' cleavage site) is shown as stick and colored magenta. The blue dashed box shows R869 does not interact tightly with MHV CR. PDB code: MHV, 3JCL. **c** Unique feature of HKU2 CR. CR and FP are colored the same as in **a**. Helices and strands in CR are labeled. R671 (S2' cleavage site) is shown as stick and colored magenta. The magenta dashed box shows detailed interactions between R671 and CR.

of HKU2 and SADS-CoV spike glycoproteins at atomic resolutions. Pairwise comparisons demonstrated nearly identical overall structures, and differences mainly locate in the loops of NTD and CTD of the S1 subunit between the spikes of HKU2 and SADS-CoV (Supplementary Fig. 6). A series of structural analysis and comparisons were also performed at the domain level between HKU2 spike with other coronavirus spikes with determined structures. Our results show that HKU2 and SADS-CoV spikes maintain primitive structural features, especially in the NTD and CTD, and provide more insights into the evolution of coronaviruses.

The HKU2 and SADS-CoV have one NTD in the S1 subunit, and their structures are more similar to the NTD1 than the NTD2 of  $\alpha$ -coronaviruses HCoV-NL63 and PEDV, whereas the only NTD of HCoV-229E is structurally more similar to the NTD2 than the NTD1 (Fig. 2). Therefore, we suggest that  $\alpha$ -coronaviruses have two subtypes of NTD. The evolution relationship between them are not clear yet. It was once suggested that the presence of two NTDs in HCoV-NL63 is a result of gene duplication<sup>39</sup>. However, the sequence identity between these two NTDs is only 15.7% in HCoV-NL63 and 12.9% in PEDV. Considering that HKU2 (SADS-CoV) and HCoV-229E have one NTD belonging to either subtype I or subtype II, a more plausible

evolution way of the NTD in  $\alpha$ -coronaviruses is the recombination of two separate primitive domains into the genome, resulting in the presence of two NTDs in the S1 subunit of  $\alpha$ -coronaviruses including HCoV-NL63 and PEDV. Of note, these two NTD subtypes may represent primitive structures that could be the evolutionary ancestors of NTDs of other genera coronaviruses. For example, in the current available spike structures, the NTDs of  $\beta$ -coronavirus are similar to the HKU2 NTD representing the subtype I in both architecture and disulfide bond positions. These  $\beta$ -coronavirus NTDs also have additional loop ceiling over the top sheet, functionally facilitating immune evasion or protein receptor binding such as in MHV<sup>37,38</sup>. The NTD of  $\gamma$ -coronavirus IBV is also architecturally similar to the subtype I, although the disulfide bond positions are not conserved (Fig. 2e). In contrast, the  $\delta$ -coronavirus PdCoV NTD is similar to the HCoV-229E NTD representing the subtype II in both architecture and disulfide bond positions (Fig. 2f). A previous study of the IBV spike proposed that  $\alpha$ -coronavirus NTDs are probably the most ancestral and the NTDs of the four genera form an evolutionary spectrum in the order of  $\alpha$ -,  $\delta$ -,  $\gamma$ -, and  $\beta$ -genus<sup>52</sup>. Our proposal here is similar to the previous one in the point that two NTD subtypes in  $\alpha$ -coronaviruses may represent primitive structures that could be the evolutionary ancestors of NTDs. However, we

argue that the evolution pathways may not be in the order of  $\alpha$ -,  $\delta$ -,  $\gamma$ -, and  $\beta$ -genus. A more plausible pathway is that the  $\beta$ -,  $\gamma$ -, and  $\delta$ -coronavirus NTDs may evolve independently and parallelly from subtype I ( $\beta$ - and  $\gamma$ -coronavirus NTDs) or subtype II ( $\delta$ -coronavirus NTDs) (Supplementary Fig. 8b).

The HKU2 and SADS-CoV CTD structure has a one-layer core consisting of a twisted five-stranded antiparallel  $\beta$ -sheet. Interestingly,  $\beta$ -coronavirus CTDs also have the similar one-layer core structure and three strictly conserved disulfide bonds are also present in the core of HKU2 CTD. Currently, all identified RBM of  $\beta$ -coronavirus CTDs are within an inserted domain between two strands of the core sheet, and this insertion responsible for receptor binding of  $\beta$ -coronaviruses is replaced by a short loop in HKU2 CTD. This result firstly indicates HKU2 CTD represent a primitive structure in the one-layer CTD family, while the inserted domain in  $\beta$ -coronaviruses results from a recombinant event during evolution (Supplementary Fig. 8c). The second indication is that HKU2 and SADS-CoV may not utilize the CTD to bind protein receptors that have not been identified yet, and their different receptor usage may be determined by the NTD that harbors almost 50% of residue difference between them. Of note, the CTDs from other  $\alpha$ -coronaviruses,  $\gamma$ -coronavirus IBV, and  $\delta$ -coronavirus PdCoV all belong to the two-layer subtype consisting of two layers  $\beta$ -sheets, although with structural variations in different viruses. These results further confirmed the previous phylogenetic analysis suggesting that HKU2 and SADS-CoV probably resulted from a recombination of an  $\alpha$ -coronavirus genomic backbone with an unrecognized  $\beta$ -coronavirus spike gene<sup>8,11,59</sup>.

In contrast with the NTD and CTD having unique structural features, the SD1 and SD2 of the S1 subunit and the S2 subunits of HKU2 and SADS-CoV are structurally conserved as those of other coronaviruses. In the evolutionary aspect, it is not surprising because this region either connects the CTD to the S2 subunit (SD1 and SD2) or mediates the membrane fusion (S2 subunits), whereas the NTD and CTD are key factors determining tissue tropism and host range of coronaviruses<sup>23</sup>. Even highly conserved in overall structure, the S2 subunit in HKU2 and SADS-CoV still have a secondary structure arrangement in the CR after the FP, resulting in a more buried S2' cleavage site (Fig. 6c). It indicates that although the membrane fusion mechanism is highly conserved, the dynamic fusion procedures of HKU2 and SADS-CoV may still have their unique features that need to be addressed in the future.

The ongoing COVID-19 pandemic caused by SARS-CoV-2 has raised worldwide alert<sup>4–6</sup>. We compared the domain structures of HKU2 and SARS-CoV-2 S proteins (Supplementary Fig. 9). The NTD of SARS-CoV-2 has the two conserved disulfide bonds in subtype I; the CTD core of SARS-CoV-2 has one layer of  $\beta$ -sheets; the SD1 and SD2 of SARS-CoV-2 have the same disulfide bonds with HKU2, although the additional helix of SD2 is not built in the SARS-CoV-2 S model owing to poor density in this region. SARS-CoV-2 has the typical  $\beta$ -coronavirus structural features in each domain, and the inserted RBM in the CTD could be the major impetus for cross-species transmission from bats to human.

In comparing the structures of HKU2 and SADS-CoV spikes with other coronavirus spikes, we observed that additional NTDs could be recombined, ceiling loops could be inserted into NTD core, and extra domain containing RBM could be inserted into the CTD core structure. These phenomena indicate the sub-domains are gradually recruited into the S1 subunit during evolution, and the recruitments are required for cross-species transmission, adapting to different host range, and responding to the updating of host immune system, which provides a vivid example for the co-evolution of virus and host.

## Methods

**Expression and purification of HKU2 and SADS-CoV spike ectodomains.** The cDNAs encoding HKU2 spike (YP\_001552236) and SADS-CoV spike (AVM41569.1) were synthesized with codons optimized for insect cell expression. HKU2 ectodomains (1–1066) and SADS-CoV ectodomains (1–1068) were separately cloned into pFastBac-Dual vector (Invitrogen) with C-terminal foldon tag for trimerization and Strep tag for purification in the frame of pH promoter. They were expressed in Hi5 insect cell using Bac-to-Bac baculovirus system (Invitrogen). Briefly, the construct was transformed into DH10Bac competent cells and the extracted bacmid was transfected into Sf9 cell using Cellfectin II reagent (Invitrogen). The baculoviruses were harvested after 7–9 days. The high-titer viruses were generated after one more amplification that used to infect Hi5 cells at a density of  $1.5 \times 10^6$ /ml. After 60 h in culture, the cell medium containing spike ectodomains were concentrated and exchanged to binding buffer (50 mM Tris, pH 8.0, 150 mM NaCl). The spike ectodomains were purified by StrepTactin beads (IBA) and then purified by gel-filtration chromatography using Superose 6 gel filtration column (GE Healthcare) pre-equilibrated with HBS buffer (10 mM HEPES pH 7.2, 150 mM NaCl). Purified ectodomains were concentrated for electron microscopy analysis.

**Cryo-electron microscopy.** Aliquots of spike ectodomains (4  $\mu$ l, 0.33 mg/ml, in buffer containing 10 mM HEPES pH 7.2, 150 mM NaCl) were applied to glow-discharged holey carbon grids (Quantifoil grid, Au 300 mesh, R1.2/1.3). The grids were then blotted and then plunge-frozen in liquid ethane using FEI Vitrobot system (FEI).

Images were recorded using FEI Titan Krios microscope operating at 300 kV with a Gatan K2 Summit direct electron detector (Gatan Inc.) at Tsinghua University. The automated software (AutoEMation) was used to collect 7663 movies for HKU2 and 4568 movies for SADS-CoV at  $\times 130,000$  magnification and at a defocus range between 1 and 3  $\mu$ m. Each movie has a total accumulate exposure of  $49.784 e^-/\text{\AA}^2$  fractionated in 32 frames of 175 ms exposure. Data collection statistics are summarized in Table 1.

Whole frames in each movie were corrected for beam-induced motion using MotionCo2<sup>60</sup>. The final image was bin averaged to give a pixel size of 1.061  $\text{\AA}$ . The parameters of contrast transfer function (CTF) was estimated for each micrograph using GCTF<sup>61</sup>. Particles were automatically picked using Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) and extracted using RELION<sup>62</sup>. Initially,  $\sim 1,400,000$  particles for HKU2 and  $\sim 900,000$  particles for SADS were subjected to 2D classification. After two or three additional 2D classifications, the best class consisting  $\sim 750,000$  particles (HKU2) and  $\sim 320,000$  particles (SADS-CoV) were applied for creating 3D initial model, 3D refinement, and 3D classification. A total of 421,490 particles (HKU2) and 152,334 particles (SADS-CoV) of best classes were selected and subjected to 3D refinement with C3 symmetry to generate density map. The reported resolutions based on the gold-standard FSC cutoff of 0.143 criterion were 2.38  $\text{\AA}$  for HKU2 spike and 2.83  $\text{\AA}$  for SADS-CoV spike after RELION post processing. Local resolution variations were estimated using ResMap<sup>63</sup>. Data processing statistics are summarized in Table 1.

**Model building and refinement.** As for the HKU2 spike model building, the initial model of HKU2 S1 NTD was generated using the SWISS-MODEL<sup>64</sup> and fit into the map using UCSF Chimera<sup>65</sup>. The other part of HKU2 spike ectodomains was obtained using Map to Model in PHENIX suite<sup>66</sup>. As for the SADS-CoV spike model building, the initial model was obtained by fit HKU2 S2, NTD, and CTD separately into the map using UCSF Chimera<sup>65</sup>. Manual model rebuilding was carried out using Coot<sup>67</sup> and refined with PHENIX real-space refinement<sup>66</sup>. The quality of the final model was analyzed with Molprobity<sup>68</sup> and EMRinger<sup>69</sup>. The validation statistics of the structural models are summarized in Table 1.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The atomic coordinates of HKU2 spike and SADS-CoV spike have been deposited in the Worldwide Protein Data Bank with the accession codes 6M15 and 6M16, respectively; the corresponding maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-30037 and EMD-30038, respectively.

Received: 8 February 2020; Accepted: 29 May 2020;

Published online: 17 June 2020

## References

1. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).

2. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **1282**, 1–23 (2015).
3. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
4. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
5. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
7. Wang, Q., Vlasova, A. N., Kenney, S. P. & Saif, L. J. Emerging and re-emerging coronaviruses in pigs. *Curr. Opin. Virol.* **34**, 39–49 (2019).
8. Fu, X. et al. Newly emerged porcine enteric alphacoronavirus in southern China: Identification, origin and evolutionary history analysis. *Infect. Genet. Evol.* **62**, 179–187 (2018).
9. Xu, Z. et al. Isolation and characterization of a highly pathogenic strain of porcine enteric alphacoronavirus causing watery diarrhoea and high mortality in newborn piglets. *Transbound. Emerg. Dis.* **66**, 119–130 (2019).
10. Zhou, P. et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **556**, 255–258 (2018).
11. Pan, Y. et al. Discovery of a novel swine enteric alphacoronavirus (SeACoV) in southern China. *Vet. Microbiol.* **211**, 15–21 (2017).
12. Gong, L. et al. A new bat-HKU2-like coronavirus in swine, China, 2017. *Emerg. Infect. Dis.* **23**, 1607–1609 (2017).
13. Wang, L., Su, S., Bi, Y., Wong, G. & Gao, G. F. Bat-origin coronaviruses expand their host range to pigs. *Trends Microbiol.* **26**, 466–470 (2018).
14. Parrish, C. R. et al. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* **72**, 457–470 (2008).
15. Lu, G., Wang, Q. & Gao, G. F. Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* **23**, 468–478 (2015).
16. Peck, K. M., Burch, C. L., Heise, M. T. & Baric, R. S. Coronavirus host range expansion and middle east respiratory syndrome coronavirus emergence: biochemical mechanisms and evolutionary perspectives. *Annu. Rev. Virol.* **2**, 95–117 (2015).
17. Woo, P. C., Lau, S. K., Huang, Y. & Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med.* **234**, 1117–1127 (2009).
18. Woo, P. C. et al. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *J. Virol.* **86**, 3995–4008 (2012).
19. Fan, Y., Zhao, K., Shi, Z. L. & Zhou, P. Bat coronaviruses in China. *Viruses* **11**, 210 (2019).
20. Yang, Y. L. et al. Broad cross-species infection of cultured cells by bat HKU2-related swine acute diarrhoea syndrome coronavirus and identification of its replication in murine dendritic cells in vivo highlight its potential for diverse interspecies transmission. *J. Virol.* **93**, e01448–19 (2019).
21. Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**, 237–261 (2016).
22. Tortorici, M. A. & Vesler, D. Structural insights into coronavirus entry. *Adv. Virus Res.* **105**, 93–116 (2019).
23. Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* **89**, 1954–1964 (2015).
24. Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* **14**, e1007236 (2018).
25. Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl Acad. Sci. USA* **106**, 19970–19974 (2009).
26. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
27. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**, 215–220 (2020).
28. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
29. Wang, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894–904.e9 (2020).
30. Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
31. Wang, N. et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* **23**, 986–993 (2013).
32. Lu, G. et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).
33. Reguera, J. et al. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* **8**, e1002859 (2012).
34. Li, Z. et al. The human coronavirus HCoV-229E S-protein structure and receptor binding. *Elife* **8**, e51230 (2019).
35. Wong, A. H. M. et al. Receptor-binding loops in alphacoronavirus adaptation and evolution. *Nat. Commun.* **8**, 1735 (2017).
36. Tortorici, M. A. et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **26**, 481–489 (2019).
37. Peng, G. et al. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl Acad. Sci. USA* **108**, 10696–10701 (2011).
38. Shang, J. et al. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog.* **16**, e1008392 (2020).
39. Walls, A. C. et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
40. Wrapp, D. & McLellan, J. S. The 3.1-Angstrom Cryo-electron Microscopy Structure of the Porcine Epidemic Diarrhoea Virus Spike Protein in the Prefusion Conformation. *J. Virol.* **93**, e00923–19 (2019).
41. Yang, T. J. et al. Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans. *Proc. Natl Acad. Sci. USA* **117**, 1438–1446 (2020).
42. Walls, A. C. et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114–117 (2016).
43. Walls, A. C. et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl Acad. Sci. USA* **114**, 11157–11162 (2017).
44. Kirchdoerfer, R. N. et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).
45. Yuan, Y. et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).
46. Gui, M. et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017).
47. Kirchdoerfer, R. N. et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 15701 (2018).
48. Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl Acad. Sci. USA* **114**, E7348–E7357 (2017).
49. Park, Y. J. et al. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nat. Struct. Mol. Biol.* **26**, 1151–1157 (2019).
50. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e286 (2020).
51. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
52. Shang, J. et al. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. *PLoS Pathog.* **14**, e1007009 (2018).
53. Shang, J. et al. Cryo-Electron Microscopy Structure of Porcine Deltacoronavirus Spike Protein in the Prefusion State. *J. Virol.* **92**, e01556–17 (2018).
54. Xiong, X. et al. Glycan Shield and Fusion Activation of a Deltacoronavirus Spike Glycoprotein Fine Tuned for Enteric Infections. *J. Virol.* **92**, e01628–17 (2018).
55. Wang, Q. et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **16**, 328–337 (2014).
56. Han, X. et al. Structure of the S1 subunit C-terminal domain from bat-derived coronavirus HKU5 spike protein. *Virology* **507**, 101–109 (2017).
57. Huang, C. et al. Putative receptor binding domain of bat-derived coronavirus HKU9 spike protein: evolution of betacoronavirus receptor binding motifs. *Biochemistry* **55**, 5977–5988 (2016).
58. Dong, B. Q. et al. Detection of a novel and highly divergent coronavirus from asian leopard cats and Chinese ferret badgers in Southern China. *J. Virol.* **81**, 6920–6926 (2007).
59. Lau, S. K. et al. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* **367**, 428–439 (2007).
60. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
61. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
62. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
63. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).



64. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
65. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
66. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
67. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
68. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
69. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

### Acknowledgements

We thank the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for the cryo-EM facility and the computational facility support, and L. Zhao, X. Li, J. Wen, H. Zhou, Y. Wang, A. Jia, and S. Zhang for technical support. This work was supported by funds from Beijing Advanced Innovation Center for Structural Biology at Tsinghua University and the National Key Plan for Scientific Research and Development of China (grant number 2016YFD0500307).

### Author contributions

J.Y. carried out protein expression, purification, electron microscopy sample preparation, data collection, image processing, and model building with the help of S.Q. Author X.W. conceived, designed, and directed the study. X.W., J.Y., and R.G. analyzed the structure, made the figures, and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-16876-4>.

**Correspondence** and requests for materials should be addressed to X.W.

**Peer review information** *Nature Communications* thanks Jean K Millet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020