



SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects

Antoni G. Wrobel^{1,5}✉, Donald J. Benton^{1,5}✉, Pengqi Xu^{2,1}, Chloë Roustan³, Stephen R. Martin³, Peter B. Rosenthal⁴, John J. Skehel¹ and Steven J. Gamblin¹✉

SARS-CoV-2 is thought to have emerged from bats, possibly via a secondary host. Here, we investigate the relationship of spike (S) glycoprotein from SARS-CoV-2 with the S protein of a closely related bat virus, RaTG13. We determined cryo-EM structures for RaTG13 S and for both furin-cleaved and uncleaved SARS-CoV-2 S; we compared these with recently reported structures for uncleaved SARS-CoV-2 S. We also biochemically characterized their relative stabilities and affinities for the SARS-CoV-2 receptor ACE2. Although the overall structures of human and bat virus S proteins are similar, there are key differences in their properties, including a more stable precleavage form of human S and about 1,000-fold tighter binding of SARS-CoV-2 to human receptor. These observations suggest that cleavage at the furin-cleavage site decreases the overall stability of SARS-CoV-2 S and facilitates the adoption of the open conformation that is required for S to bind to the ACE2 receptor.

Coronaviruses infect a range of mammalian and avian species¹. SARS-CoV-2, the agent of the COVID-19 pandemic^{2,3}, belongs to the *Sarbecovirus* subgenus of betacoronaviruses, members of which mostly infect bats^{4,5}. Hence, bat coronaviruses were identified as a likely evolutionary precursor of SARS-CoV-2 (refs. 2,3), and the bat virus RaTG13 (refs. 2,6) was identified as the closest known relative of SARS-CoV-2. It is not known how SARS-CoV-2 evolved to infect humans, but two mechanisms have been hypothesized: selection in an animal host before zoonotic transfer (possibly via an intermediate host), or natural selection in humans following direct zoonotic transmission from bats^{7,8}.

The S protein of SARS-CoV-2 mediates attachment of the virus to cell-surface receptors and fusion between virus and cell membranes¹. The receptor for SARS-CoV-2, like that for SARS-CoV^{9,10}, is the human cell-surface-membrane protein angiotensin-converting enzyme 2 (ACE2)^{11–13}. Membrane-fusion activity, as for other class I fusion glycoproteins¹⁴, requires S to be proteolytically cleaved into S1 and S2 subunits that remain associated following cleavage^{13,15–17}. In addition to substitutions in the receptor-binding domain (RBD)², a second difference between the S proteins from human and bat viruses is the presence of a four-amino-acid insertion, PRRA, which adds a furin-cleavage site between the S1 and S2 subunits¹¹. Similar cleavage sites have been found in related coronaviruses, including HKU1 and Middle East respiratory syndrome coronavirus (MERS-CoV), which infect humans^{16–18}, and the acquisition of similar cleavage sites is associated with increased pathogenicity in other viruses such as influenza virus¹⁹.

In order to examine the evolutionary origin of SARS-CoV-2 and to better understand the emergence of the COVID-19 pandemic, here we have characterized the S proteins of SARS-CoV-2 and RaTG13, determined their affinities for human ACE2 and investigated the effects of furin cleavage on the structure of S from SARS-CoV-2.

Results

Structure of protease-cleaved SARS-CoV-2 S glycoprotein. We first characterized the furin-cleaved S protein of SARS-CoV-2 virus by cryo-electron microscopy (cryo-EM) (Fig. 1a and Table 1). We produced a form of SARS-CoV-2 S protein with the furin-cleavage site intact. This protein, which we expressed in mammalian cells, was secreted in a partially cleaved form, presumably due to the naturally expressed proteases within these cells¹⁶ (Extended Data Fig. 1a). We further cleaved this protein using exogenous furin for structural and biochemical characterization (Extended Data Fig. 1a).

The particles analyzed from cryo-electron micrographs fell into three populations: a closed form (34%), an intermediate form (39%) and an open form (27%) with an upright RBD (Fig. 1a). The overall structure of the closed conformation of the S trimer has three-fold symmetry and is similar to structures from uncleaved material, described previously^{11,20}. However, for both this cleaved sample and for the uncleavable form (discussed in the next section), the closed conformation is more compact than the previously published uncleaved closed structure¹¹ (Fig. 1). In the closed conformation, the surface of the RBD, which would interact with the ACE2 receptor, is buried inside the trimer and is not accessible for receptor binding. In the intermediate form (Fig. 1), two of the three RBDs maintain a similar interaction to the closed form, but the third RBD is disordered. In addition, two of the amino-terminal domains (NTDs) that are in closest contact to the disordered RBD have shifted their center of mass by 2.5–2.9 Å from the closed form (in the same direction as the open form) (Extended Data Fig. 2e). In the open form (Fig. 1), two of the RBDs remain closely associated, similar to the closed and intermediate forms. But the third RBD rotates ~60° such that the ACE2-interacting surface becomes fully exposed at the top of the assembly, whereas the NTD of the adjacent chain moves toward the rotated domain, with the NTD of the same chain moving

¹Structural Biology of Disease Processes Laboratory, Francis Crick Institute, London, UK. ²Precision Medicine Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, Guangdong, China. ³Structural Biology Science Technology Platform, Francis Crick Institute, London, UK. ⁴Structural Biology of Cells and Viruses Laboratory, Francis Crick Institute, London, UK. ⁵These authors contributed equally: Antoni G. Wrobel, Donald J. Benton. ✉e-mail: antoni.wrobel@crick.ac.uk; donald.benton@crick.ac.uk; steven.gamblin@crick.ac.uk

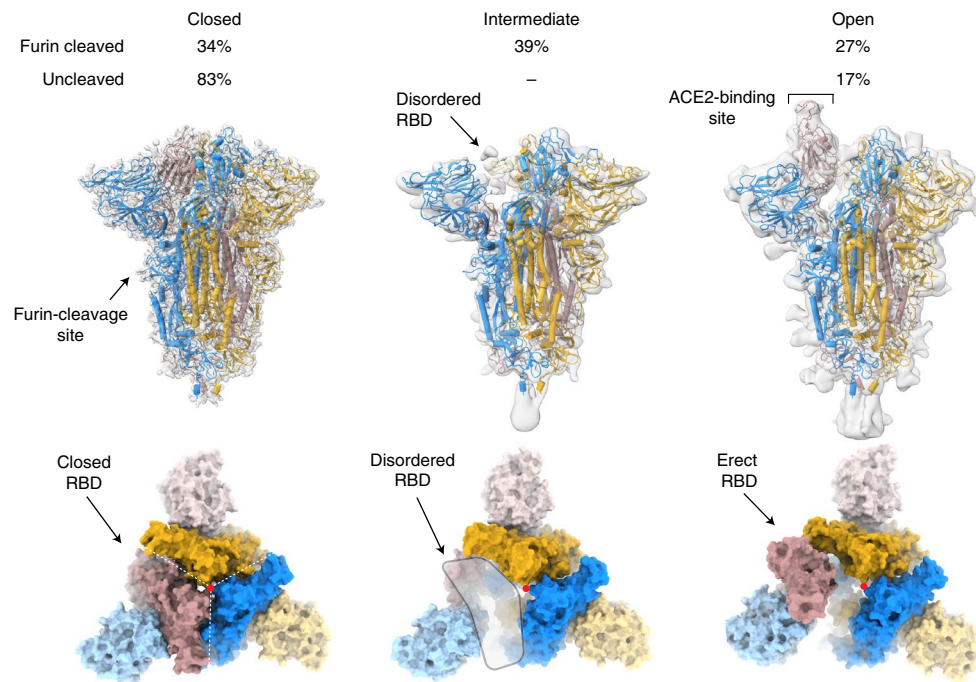


Fig. 1 | Structure of protease-cleaved SARS-CoV-2 S glycoprotein. Three structures are calculated from micrographs of furin-cleaved material: closed, intermediate and open forms, of approximately equal proportions. In the uncleaved material, most of the population represents the closed form, and a small proportion is in the open conformation. Density maps are shown for the three types of particles, overlaid with a ribbon representation of the built molecular models, viewed with the long axis of the trimer vertical (top); the three monomers are colored blue, yellow and brown. An orthogonal view (bottom) looking down the long axis (indicated by a red dot) is shown. The coloring is as in the top panel, with the NTDs in a lighter hue.

away to accommodate this rotation. The changes in domain orientations between the closed and open forms are shown for a selected monomer in Extended Data Fig. 1d.

In this protease-cleaved material, there is a much lower proportion of S protein in a closed conformation: 34% compared with 83% in the uncleaved human S trimer described in the next section, and compared with the 65% and 50% that were reported recently^{11,20}. The observation here of a substantially populated (39%) intermediate form, in which one of the RBDs is disordered and two of NTDs have shifted, suggests that this conformation, which is possibly transient, will lead to a receptor-binding-competent form. Thus, we suspect that, in addition to its requirement for membrane fusion, efficient protease cleavage might be selected to ensure there is a higher proportion of S protein on the virus surface that is capable of binding to receptor. Although the loop containing the cleavage site (residues 676–689) is disordered, in both cleaved and uncleaved forms, the observation of the intermediate form and the much lower thermal stability of the cleaved protein (Extended Data Fig. 1c) discussed in the next section suggest that cleavage reduces the overall stability of S. This reduction in stability may facilitate the movement of the NTDs and the RBDs, enabling, finally, the adoption of the open, receptor-binding-competent form.

Comparison with the S glycoprotein from the bat virus RaTG13.

Next, we determined the cryo-EM structures of S from the closest known bat virus (RaTG13) and of uncleaved SARS-CoV-2 S (Extended Data Fig. 1b). The bat virus protein was expressed in mammalian cells, but was found to be unstable during preparation of EM grids and required chemical cross-linking to produce particles for data collection and analysis. The resulting micrographs yielded a high-resolution single-particle reconstruction at 3.1-Å resolution (Table 1). The uncleaved SARS-CoV-2 S was particularly stable and gave rise to the best-quality density maps at 2.6 Å (Table 1 and Extended Data Fig. 2), enabling us to model 15%

more of the RBD (100% complete) and 25% more of the NTD (98% complete) than were modeled in earlier studies^{11,20}, influencing the overall appearance of the trimer. The structure of the bat virus S protein is similar to that of the uncleaved SARS-CoV-2 closed form (Fig. 2a,d). It may be that the chemical cross-linking required to obtain the structure of bat virus S is responsible for all particles being in the closed conformation.

Comparison of the bat virus S protein sequence with that of SARS-CoV-2 S reveals a high degree of conservation (97.8% in the ectodomain) but with a relatively high proportion of substitutions in the RBD (89.6%) (Fig. 2b). As suggested before¹¹, the substitutions are clustered at two interfaces: the ACE2-receptor-binding surface (considered in the section “Binding of ACE2 to bat virus and SARS-CoV-2 S glycoproteins”) and the RBD–RBD interfaces of the trimeric S. Analysis of the latter interface in the SARS-CoV-2 trimer reveals an extensive network of potential intratrimer hydrogen bonds, including the interaction of Arg403, Gln493 and Tyr505 from one subunit with Ser373, Ser371 and Tyr369 from another (Fig. 2c). The corresponding residues in the bat structure, and other intersubunit contacts, suggest a lower surface complementarity. Of note, the bat virus S protein has an N-glycosylation site at Asn370, where a bulky fucosylated glycan wedges between adjacent domains (Extended Data Fig. 3). Indeed, calculations of surface contact area show that, in the bat virus S trimer, the monomer–monomer interactions account for 5,200 Å² (of which 485 Å² is between the RBDs), whereas the equivalent contact area in the closed structure of the SARS-CoV-2 S trimer is 6,100 Å² (with 550 Å² between the RBDs). Thermal-stability data show that the uncleaved SARS-CoV-2 S trimer has a markedly higher stability than the bat virus protein does, whereas the cleaved SARS-CoV-2 has a similar stability to the (uncleaved) bat virus protein (Extended Data Fig. 1c). Perhaps the higher stability of SARS-CoV-2 S is required to offset some of the loss of stability that occurs upon cleavage. These structural and biochemical data together suggest that the furin-cleavage site might

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

	Uncleavable closed (EMDB-11203, PDB 6ZGE)	Cleaved closed (EMDB-11207, PDB 6ZGI)	Intermediate (EMDB-11206, PDB 6ZGH)	Cleaved open (EMDB-11205, PDB 6ZGG)	Bat virus (EMDB-11204, PDB 6ZGF)
Data collection and processing					
Voltage (kV)	300	300	300	300	300
Electron exposure (e ⁻ /Å ²)	33.6	54.4	54.4	54.4	54.4
Defocus range (μm)	-1.5 to -3	-1.5 to -3	-1.5 to -3	-1.5 to -3	-1.5 to -3
Pixel size (Å)	1.09	1.08	1.08	1.08	1.08
Symmetry imposed	C3	C3	C1	C1	C3
Final particle images (no.)	95,000	95,000	107,000	73,000	62,000
Map resolution (Å) FSC threshold = 0.143	2.6	2.9	6.8	3.8	3.1
Map resolution range (Å)	2.4-3.2	2.8-3.6	6-10	3.5-7.5	3-3.8
Refinement					
Initial model used (PDB code)	6VXX	-	-	-	-
Model resolution (Å) FSC threshold = 0.5	2.7	3.0	8.2	4.1	3.2
Map sharpening <i>B</i> factor (Å ²)	-109.8	-91.2	-207.0	-61.7	-88.2
Model composition					
Non-hydrogen atoms	26,991	26,901	23,735	25,111	26,169
Protein residues	3,294	3,294	3,032	3,207	3,180
Ligands	81	78	-	-	93
<i>B</i> factors (Å²)					
Protein	35.1	47.8	-	-	37.1
Ligand	83.0	77.5	-	-	52.2
R.m.s. deviations					
Bond lengths (Å)	0.009	0.006	0.003	0.003	0.007
Bond angles (°)	0.902	0.753	0.855	0.863	1.13
Validation					
MolProbity score	1.33	1.21	1.69	1.79	1.7
Clashscore	2.20	1.64	3.20	3.79	5.16
Poor rotamers (%)	0.73	0.73	0.90	0.98	0.75
Ramachandran plot					
Favored (%)	95.22	95.68	88.94	87.13	93.57
Allowed (%)	4.69	4.23	10.75	12.36	6.43
Disallowed (%)	0.09	0.09	0.3	0.5	0

confer the human virus with an advantage, as the cleavage facilitates a higher proportion of the open, receptor-binding-competent conformation.

Binding of ACE2 to bat virus and SARS-CoV-2 S glycoproteins.

As mentioned above, the second region with a high sequence difference between the bat virus and SARS-CoV-2 S RBDs is the receptor-binding site. To quantitate the impact of these differences on binding to the human ACE2 receptor, we measured binding with surface biolayer interferometry. The S protein, either from human or bat viruses, was immobilized onto the surface of a sensor, and purified ACE2 was flowed over the surface to measure binding. Amplitude analysis suggested that SARS-CoV-2 S binds approximately 1,000 times more tightly to ACE2 than the bat virus protein does, with K_d values of <100 nM and >40 μM, respectively (Fig. 3a).

Previous studies have determined the structural interaction of the isolated RBD of SARS-CoV-2 S with human ACE2 (refs. 21–23). This information (PDB 6VW1)²¹ enabled us to model and compare

the ACE2 domain bound to the RBD domain of our SARS-CoV-2 and bat virus S trimers; it should be noted that, due to conformational plasticity of side chains, analysis of isolated domain can only partially address potential binding interactions. In the case of SARS-CoV-2 S-ACE2, there is a buried surface area of 840 Å². As well as a series of specific salt and hydrogen bonds, a notable feature is that Phe486 from SARS-CoV-2 S inserts into a hydrophobic pocket on the surface of ACE2 formed by residues including Phe28, Leu79, Met82 and Tyr83. In contrast, in the bat virus S protein, hydrophobic Phe486 is replaced by a less-bulky Leu486 (Fig. 3b), which may account in part for the smaller buried surface of the bat virus S-ACE2 complex of 760 Å². Structural comparison also suggests another substitution that likely contributes to the greatly enhanced affinity of SARS-CoV-2 S binding to ACE2: Gln493 of S makes a potential hydrogen bond with Glu35 of ACE2, which forms an intramolecular salt bridge with Lys31; in turn, ACE2 Lys31 forms a salt bridge with S Glu484. In contrast, the residue equivalent to SARS-CoV-2 Gln493 in the bat virus S is a tyrosine that is unlikely

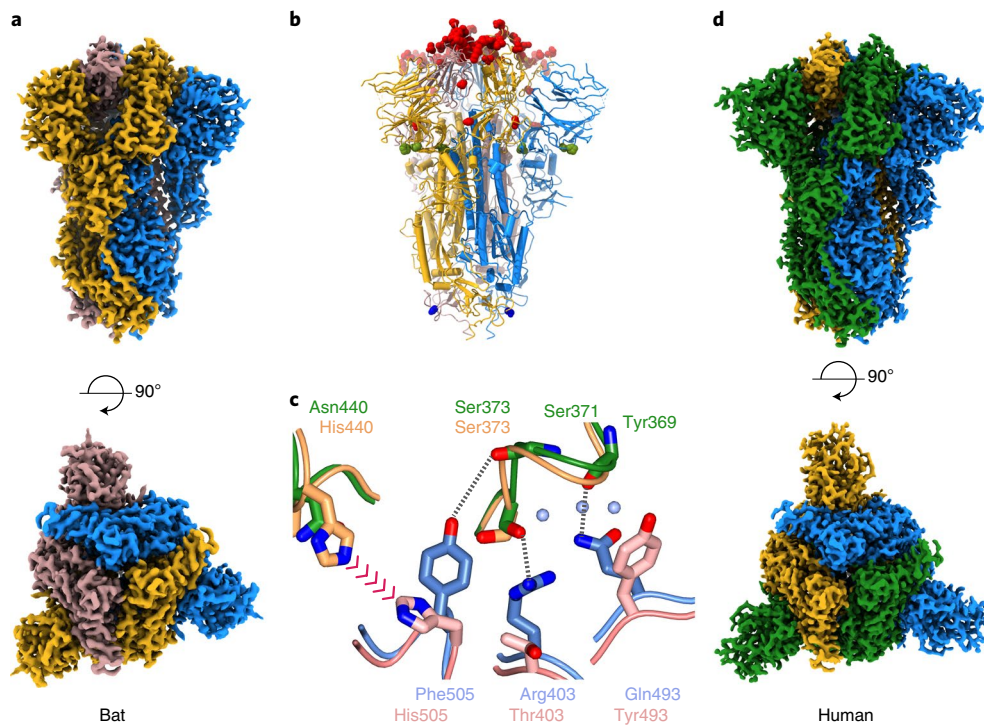


Fig. 2 | Structural comparison of the S glycoproteins from the bat virus RaTG13 and from SARS-CoV-2. **a**, The density map for the bat virus trimeric S is shown, with the long axis vertical in the top panel and in an orthogonal view in the bottom panel. All of the particles are in the closed conformation, likely because of the cross-linking of the material. The three monomers are colored blue, yellow and brown. **b**, Molecular model of the bat virus S protein, colored as in **a**, with substitutions between the bat virus and SARS-CoV-2 highlighted. Most of the changes are in the RBD and are colored red; there are four substitutions in S1 outside of the RBD, which are shown in green, and a single substitution in S2 is shown in blue. **c**, Overlay of the molecular structure of a portion of the RBD-RBD interface; the two bat virus S monomers are colored gold (top) and pink (bottom), and the two superposed SARS-CoV-2 S RBD chains are shown in green (top) and blue (bottom). Analysis suggests that the residues at the interface of SARS-CoV-2 S RBD chains support several additional stabilizing interactions and avoid the potential steric repulsion between His505 and His440, seen in the structure of the bat virus. **d**, The density map for the uncleaved SARS-CoV-2 S protein, in the closed conformation, shown in the same orientation as in **a**, with the subunits colored blue, green and yellow. This sample gave the best quality maps and enabled the most extensive build of the polypeptide chain.

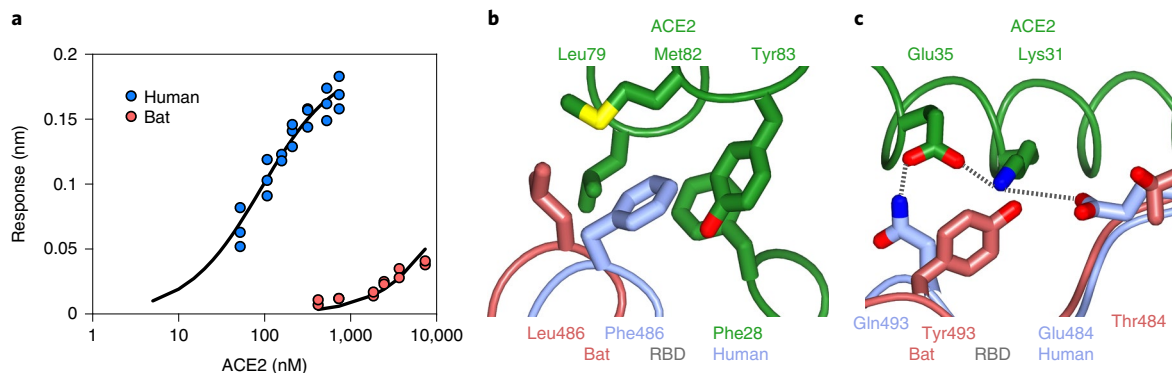


Fig. 3 | Binding of ACE2 receptor to bat virus and SARS-CoV-2 S proteins. **a**, Plot of surface biolayer amplitude measurement as a function of ACE2 concentration with the data for S from SARS-CoV-2 (blue, K_d calculated as 91 ± 18 nM) and from the bat virus (red, K_d estimated to be >40 μ M). K_d for the SARS-CoV-2 protein was calculated from kinetic constants ($k_{off} = 0.0105$ s $^{-1}$ and $k_{on} = 1.56 \times 10^5$ m $^{-1}$ s $^{-1}$) and was 67.5 ± 9 nM. **b, c**, Ribbon representation of modeled molecular interactions between ACE2 (green) with RBD from S in SARS-CoV-2 (blue) (both PDB 6VW1)²¹ and bat virus (red, this study). **b**, Details of a hydrophobic pocket on ACE2 that accommodates a phenylalanine residue from the SARS-CoV-2 S RBD. **c**, Two salt bridges and a charged hydrogen bond linking SARS-CoV-2 S RBD to ACE2, while the interface with bat virus S RBD is not able to make these interactions and presents a potential steric clash between Tyr493 and ACE2 Lys31.

to bond to ACE2 Glu35, and SARS-CoV-2 Glu484 is replaced by a threonine that would not bond to ACE2 Lys31 (Fig. 3c). Moreover, SARS-CoV-2 Gln498 is replaced by a Tyr498 that cannot form a hydrogen bond to ACE2 Tyr41.

Discussion

Together, our structural and biochemical data indicate that a bat virus, similar to RaTG13, would not be able to bind effectively to human ACE2 receptor and would be unlikely to infect humans

directly. Given the modular nature of the human and bat S glycoproteins, and the number and structural locations of the amino-acid-sequence differences between them, our observations support the involvement of recombination⁸ between distinct coronavirus genomes in the generation of SARS-CoV-2.

The structure of the SARS-CoV-2 S protein presented here is at high resolution and is nearly complete, and has many more external loops included than previously reported structures do, providing important insights for vaccine design. Furthermore, our study suggests that the presence of the furin-cleavage site in the S protein of SARS-CoV-2 facilitates the conformational change required for RBD exposure and binding to surface receptors.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-020-0468-7>.

Received: 13 June 2020; Accepted: 24 June 2020;

Published online: 09 July 2020

References

- Masters, P. S. & Perlman S. in *Fields Virology* 6th edn, Vol. 1 (eds Howley, P. M. & Knipe, D. M.) 825–858 (Lippincott Williams & Wilkins, 2013).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
- Li, W. et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
- Yang, L. et al. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg. Infect. Dis.* **19**, 989–991 (2013).
- Ge, X. Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Viol. Sin.* **31**, 31–40 (2016).
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
- Li, X. et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci. Adv.* **6**, eabb9153 (2020).
- Li, F., Li, W., Farzan, M. & Harrison, S. C. Structural biology: structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
- Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
- Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **180**, 281–292 (2020).
- Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
- Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
- Harrison, S. C. Viral membrane fusion. *Virology* **479–480**, 498–507 (2015).
- Belouzard, S., Chu, V. C. & Whittaker, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl Acad. Sci. USA* **106**, 5871–5876 (2009).
- Millet, J. K. & Whittaker, G. R. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015).
- Millet, J. K. & Whittaker, G. R. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl Acad. Sci. USA* **111**, 15214–15219 (2014).
- Chan, C. M. et al. Spike protein, S, of human coronavirus HKU1: role in viral life cycle and application in antibody detection. *Exp. Biol. Med.* **233**, 1527–1536 (2008).
- Steinhauer, D. A. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* **258**, 1–20 (1999).
- Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
- Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
- Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
- Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Design of protein constructs. The constructs corresponding to the ectodomain (residues 1–1208) of SARS-CoV-2 S protein (NCBI reference sequence YP_009724390.1) and the ectodomain (residues 1–1204) of bat RaTG13 S protein (QHR63300.2) were both codon-optimized for human expression, synthesized and cloned into pcDNA.3.1(+) vector by GenScript with an N-terminal μ -phosphatase secretion leader sequence and a carboxy-terminal hexa-histidine tag preceded by a foldon trimerization tag and a TEV-cleavage site, all separated by short glycine-rich linkers. Both constructs were made as '2P' mutants (K986P and V987P for YP_009724390, and K982P and V983P for QHR63300.2) for increased yield and to prevent the proteins from assuming the post-fusion conformation²⁴.

The ectodomain (residues 19–615) of human ACE2 (NM_021804.2) was optimized for human expression, synthesized and cloned into pcDNA.3.1(+) vector by GenScript with an N-terminal Ig-kappa chain secretion leader sequence and a C-terminal Twin-Strep tag preceded by a DYK tag.

Protein expression and purification. Proteins were expressed in Expi293F cells (Gibco) cultured in suspension in humidified, 8% CO₂ atmosphere, at 37°C with shaking at 125 r.p.m. Cell cultures were grown in FreeStyle 293 Expression Medium to a density of 3 million cells per ml at the time of transfection and transfected with 1 mg of DNA per liter of culture and ExpiFectamine 293 (Gibco) according to the manufacturer's instructions. The supernatants were collected twice: after 3–4 and 6–7 days, and were clarified, filtered and incubated with appropriate affinity resin.

S proteins were bound to 5–7 ml of TALON cobalt beads (Takara) per liter of culture, washed briefly and eluted with imidazole. ACE2 was bound to 4–6 ml of Strep-Tactin XT resin (iba) per litre of cell-culture supernatant, which had been pretreated with the BioLock solution (iba). The beads were briefly washed, and the protein was eluted with Strep-Tactin XT Elution Buffer BXT (iba). All proteins were then concentrated and either flash-frozen or gel-filtered on a Superdex 200 Increase 10/300 GL column (GE Life Sciences) into a buffer containing 20 mM Tris pH 8.0 and 150 mM NaCl.

Furin treatment. Recombinant furin (New England Biolabs) was used to cleave the SARS-CoV-2 S protein. Two units of the enzyme were used per 25 μ g of the S protein, and the reaction was performed at 25°C in the presence of 1 mM CaCl₂ and was stopped by addition of 4 mM EDTA. SDS-PAGE was used to track the progression of the reaction.

Thermal-stability measurements. Protein melting temperatures were measured using Differential Scanning Fluorimetry. Twenty-microliter reactions consisted of 5 μ g protein with SYPRO Orange (Sigma) present at a 5 \times concentration, diluted from the 5,000 \times concentrate. Fluorescence was measured between 25°C and 95°C every 0.5°C in 140 cycles, using an Agilent Stratagene Mx3005P. Each experiment was repeated at least three times.

Biolayer interferometry. Human ACE2 binding to coronavirus S proteins was measured on an Octet Red 96 instrument (ForteBio) in the buffer containing 20 mM Tris pH 8.0 and 150 mM NaCl, at 25°C, with shaking at 1,000 r.p.m. NiNTA (NTA, ForteBio) sensors were used with the bat and furin-uncleavable human S proteins. The sensors were pre-equilibrated in the buffer, and S proteins were immobilized on them at 15–30 μ g ml⁻¹ for 30–40 min. ACE2 binding was measured using a 3- to 5-min association step followed by a 10- to 15-min dissociation step. Each experiment was repeated at least three times. Association phases were analyzed as a single exponential function, and plots of the observed rate (k_{obs}) versus ACE2 concentration gave the association and dissociation rate constants (k_{on} and k_{off}) as the slope and intercept, respectively. K_d was determined as k_{off}/k_{on} and, where possible, independently by analysis of the variation of maximum response with ACE2 concentration.

Cryo-EM sample preparation and data collection. The furin-uncleavable human S was frozen at 0.4 mg ml⁻¹ in a buffer (20 mM Tris pH 8.0 and 150 mM NaCl) supplemented with 0.1% OG on an R2/2 200 mesh Quantifoil grid. Extensive buffer optimization and cross-linking had to be performed to obtain RaTG13 S protein suitable for cryo-EM. The protein in MES buffer (50 mM MES pH 6.0, 100 mM NaCl) was treated with BS³ (Thermo Scientific) at a final concentration of 0.5 mM on ice for an hour, and then a Grafix protocol²⁵ was performed to achieve double cross-linking: the BS³-cross-linked sample was loaded on a 10–30% glycerol and 0–0.15% glutaraldehyde gradient containing 50 mM MES pH 6.0, 100 mM NaCl and was spun at 35,000 r.p.m. in a SW41 rotor (Beckmann) for 20 h at 4°C. The reaction was then quenched with a final concentration of 90 mM Tris pH 7.5, and fractions containing cross-linked protein identified with SDS-PAGE were pooled, concentrated and gel-filtered into 50 mM MES pH 6.0, 100 mM NaCl. The resulting bat S protein and the furin-treated human S were frozen at a concentration of 0.1 mg ml⁻¹ on R2/2 200 mesh Quantifoil grids coated with a thin layer of continuous carbon. All grids were freshly glow discharged for 30 s at 25 mA prior to freezing. All samples were prepared by applying 4 μ l of a sample to a grid equilibrated to 4°C in 100% humidity, followed by a 4- to 5-s blot using a Vitrobot MkIII and plunge freezing into liquid ethane.

Data were collected using EPU software (Thermo Scientific) on Thermo Scientific Titan Krios microscopes operating at 300 kV. For the furin-uncleavable SARS-CoV-2 S dataset, the micrographs were collected using Falcon 3 detector (Thermo Scientific) operating in electron-counting mode. Exposures were 60 s with a total dose of 33.6 e/Å², fractionated into 30 frames, with a calibrated pixel size of 1.09 Å. For the RaTG13 S and furin-treated SARS-CoV-2 S datasets, micrographs were collected using a Gatan K2 detector mounted on a Gatan GIF Quantum energy filter operating in zero-loss mode with a slit width of 20 eV. Exposures were 8 s with a total dose of 54.4 e/Å², fractionated into 32 frames, with a calibrated pixel size of 1.08 Å. All datasets were collected using defoci between 1.5 and 3 μ m.

Cryo-EM data processing. Movie frames were aligned using MotionCor2 (ref. ²⁶) implemented in RELION²⁷ and Contrast Transfer Function fitting using CTFind4 (ref. ²⁸). All subsequent data processing was carried out using both RELION and cryoSPARC²⁹. Particles for the furin-uncleavable SARS-CoV-2 S dataset were picked using crYOLO³⁰ with a model trained on manually picked micrographs. For the datasets on the carbon support, particles were picked using RELION auto-picking. All datasets were subject to two rounds of RELION two-dimensional classification, retaining classes with clear secondary structure. These particles were classified using RELION three-dimensional classification with initial models generated using ab initio reconstructions in cryoSPARC. The details of these classifications for each of the three datasets are detailed in Extended Data Fig. 4.

Final refinements were carried out using cryoSPARC homogeneous refinement for all models except the intermediate conformation, which was refined using RELION. Local resolution was estimated using blocres³¹ implemented in cryoSPARC. Maps were filtered by local resolution and globally sharpened³² in cryoSPARC. Additional information is available in Table 1 and Extended Data Fig. 5.

Model building. The model for the uncleavable SARS-CoV-2 S protein in the closed conformation was started using the published structure (PDB 6VXX). The model was fitted to the density, and extra regions were manually built using Coot³³. This model of the uncleavable SARS-CoV-2 structure in the closed conformation was then used as the basis for building the RaTG13 structure, which was mutated at the relevant residues in Coot. Both models were real-space refined and validated using PHENIX³⁴.

The intermediate and open structures were generated by fitting the uncleavable closed human protein to the density. The open structure required manual erection of the RBD in Coot. Both the open and intermediate models were refined using Namdinator³⁵, followed by geometry normalization using PHENIX. Additional information is available in Table 1.

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

Data availability

Maps and atomic models have been deposited in the Electron Microscopy Data Bank (EMDB) and wwPDB with the following accession codes: EMD-11203 and PDB 6ZGE (SARS-CoV-2 S uncleaved closed); EMD-11207 and PDB 6ZGG (SARS-CoV-2 S cleaved, closed); EMD-11206 and PDB 6ZGH (SARS-CoV-2 S cleaved, intermediate); EMD-11205 and PDB 6ZGG (SARS-CoV-2 S cleaved, open); EMD-11204 and PDB 6ZGF (bat coronavirus RaTG13 S). Source data are provided with this paper.

References

- 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).
- Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
- Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
- Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
34. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
35. Kidmose, R. T. et al. Namdinator — automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526–531 (2019).

Acknowledgements

We would like to acknowledge A. Nans of the Structural Biology Science Technology Platform for assistance with data collection, P. Walker and A. Purkiss of the Structural Biology Science Technology Platform and the Scientific Computing Science Technology Platform for computational support and L. Calder for discussions and assistance with experiments. We thank P. Cherepanov, G. Kassiotis and S. Kjaer for discussions. This work was funded by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001078 and FC001143), the UK Medical Research Council (FC001078 and FC001143) and the Wellcome Trust (FC001078 and FC001143). P.X. is also supported by the 100 Top Talents Program of Sun Yat-sen University and the Sanming Project of Medicine in Shenzhen (SZSM201911003)

Author contributions

A.G.W., D.J.B., P.X., C.R. and S.R.M. performed research and collected and analyzed data; A.G.W., D.J.B., P.B.R., J.J.S. and S.J.G. conceived and designed the research and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

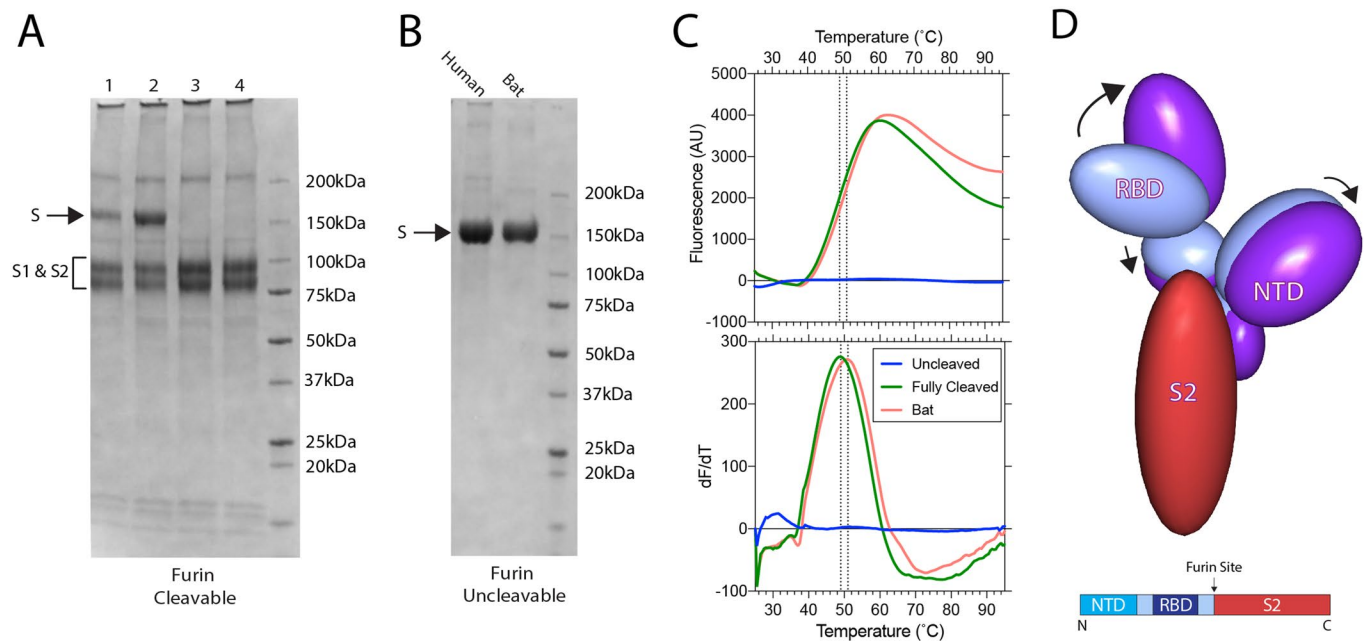
Extended data is available for this paper at <https://doi.org/10.1038/s41594-020-0468-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-020-0468-7>.

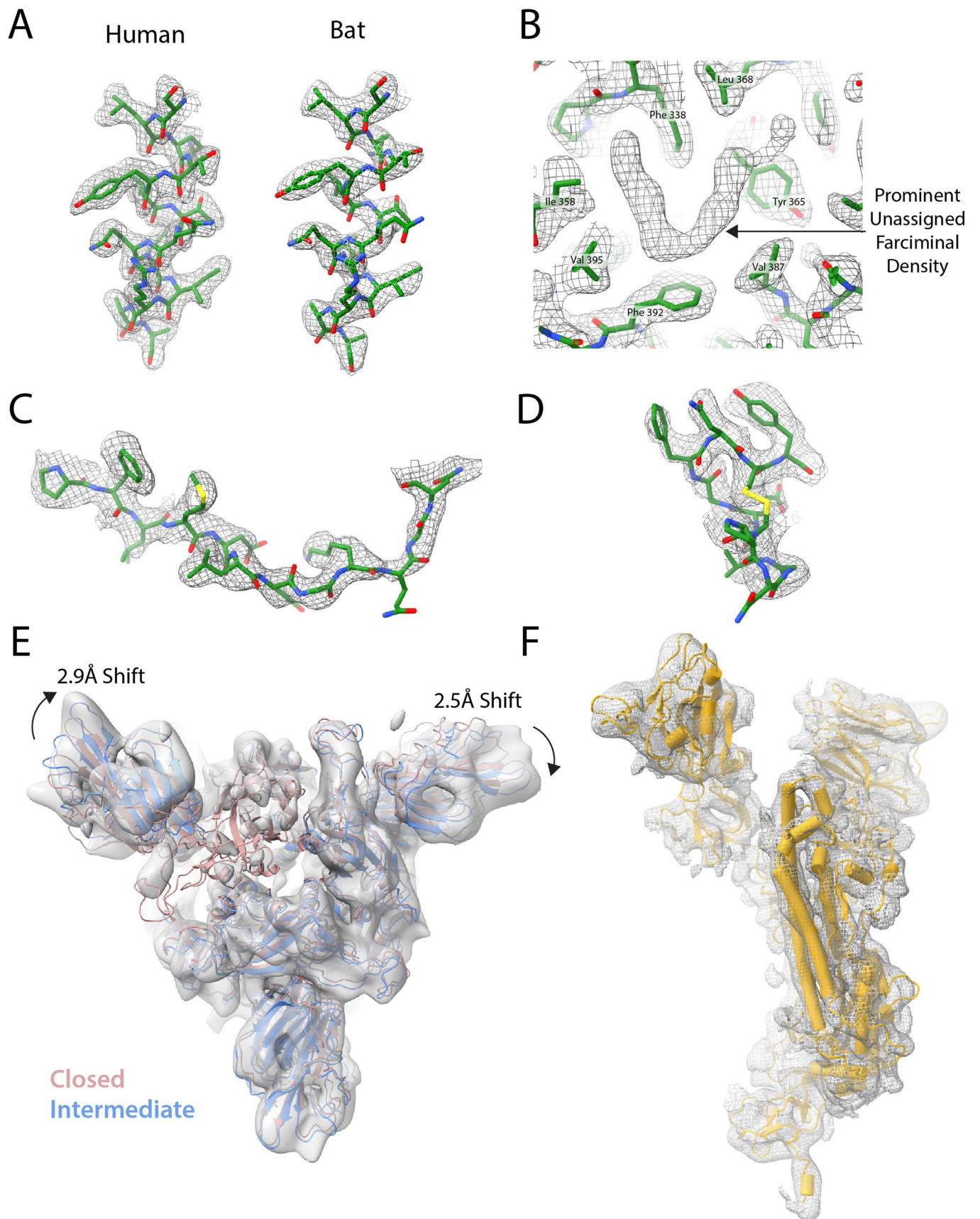
Correspondence and requests for materials should be addressed to A.G.W., D.J.B. or S.J.G.

Peer review information Inês Chen was the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

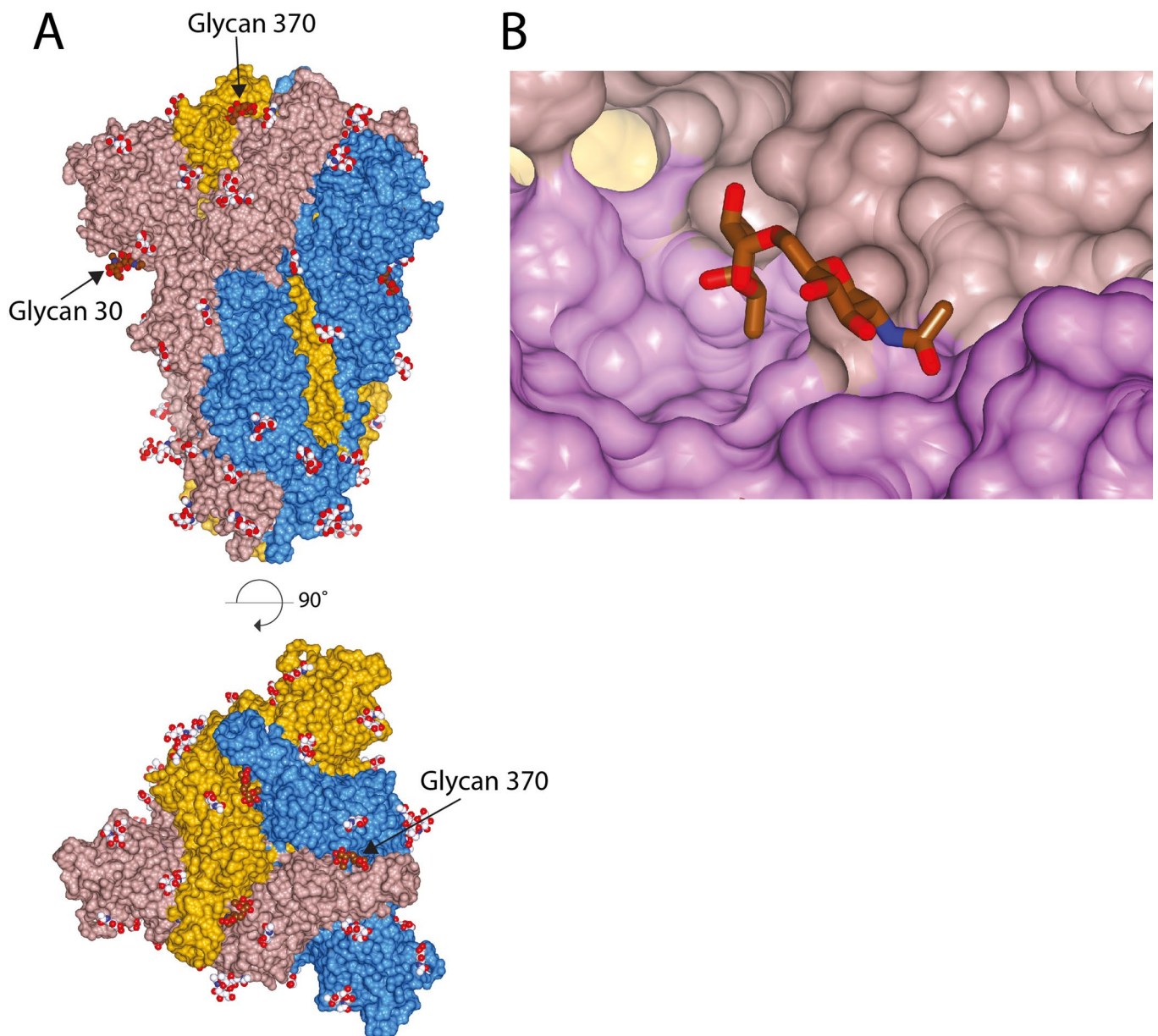


Extended Data Fig. 1 | Biochemical analyses of spike proteins. **a**, SDS-PAGE of furin-cleavable SARS-CoV-2 protein. 1: 5 hr furin cleavage; 2: Expressed protein, not cleaved *in vitro*; 3: 12 hr furin cleavage; 4: 32 hr furin cleavage. **b**, SDS-PAGE of uncleavable RaTG13 and SARS-CoV-2 S proteins. **c**, Differential Scanning Fluorimetry measurement of melting temperature for uncleavable RaTG13 and SARS-CoV-2 S, and fully-furin-cleaved SARS-CoV-2 S proteins. **d**, The changes in domain orientation, between the closed and open forms of SARS-CoV-2 S, shown schematically, for the monomer that undergoes the most substantive change in the RBD position. The image is produced by the CCP4 MG 'bloboid representation' and is calculated from the shape and centre of mass of the molecular model. Also shown is a bar representation of the domains with the furin cleavage site indicated.



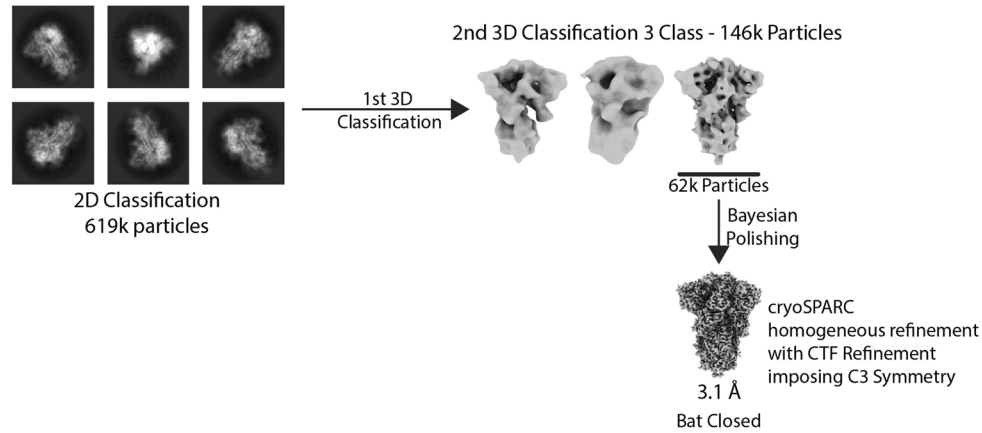
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Density features of spike protein cryoEM density. EM density shown as grey mesh, with built models shown in green. **a**, Density for residues 1003-1016 of the uncleaved closed SARS-CoV-2 (human) and RaTG13 (bat) S structures. **b**, Unassigned farciminal density seen in maps of the closed conformation of SARS-CoV-2 spike protein. **c, d**, Typical density for examples of previously unbuilt external loops in uncleaved closed SARS-CoV-2 structures: residues 174-185 of the NTD (**c**) and residues 479-489 of the RBD (**d**). **e**, EM density of Intermediate conformation of furin-cleaved SARS-CoV-2 spike. The fitted model (blue) is compared to that of the closed conformation (pink). Shifts in the NTDs are indicated with arrows, with values calculated from translations of their centres of mass. **f**, Example EM density of a monomer of the Intermediate conformation map with the built structure shown in yellow.

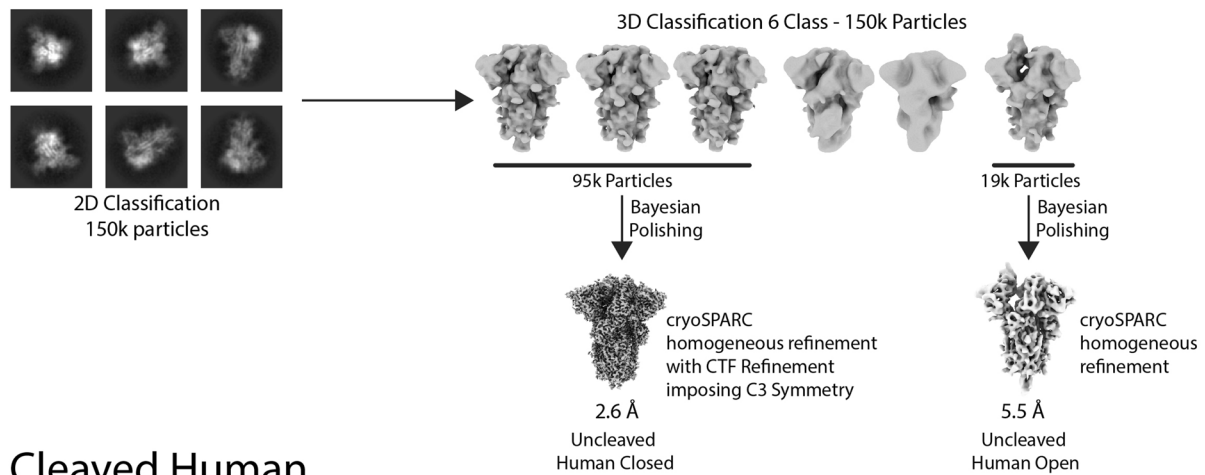


Extended Data Fig. 3 | N-linked glycosylation of RaTG13 spike protein. a, All buildable glycans shown with bat-virus-specific glycosylations, attached to Asn-30 and Asn-370, highlighted in brown **(b)** Glycan at Asn-370, shown in brown, which is inserted in the cleft between RBDs, shown in different hues of pink.

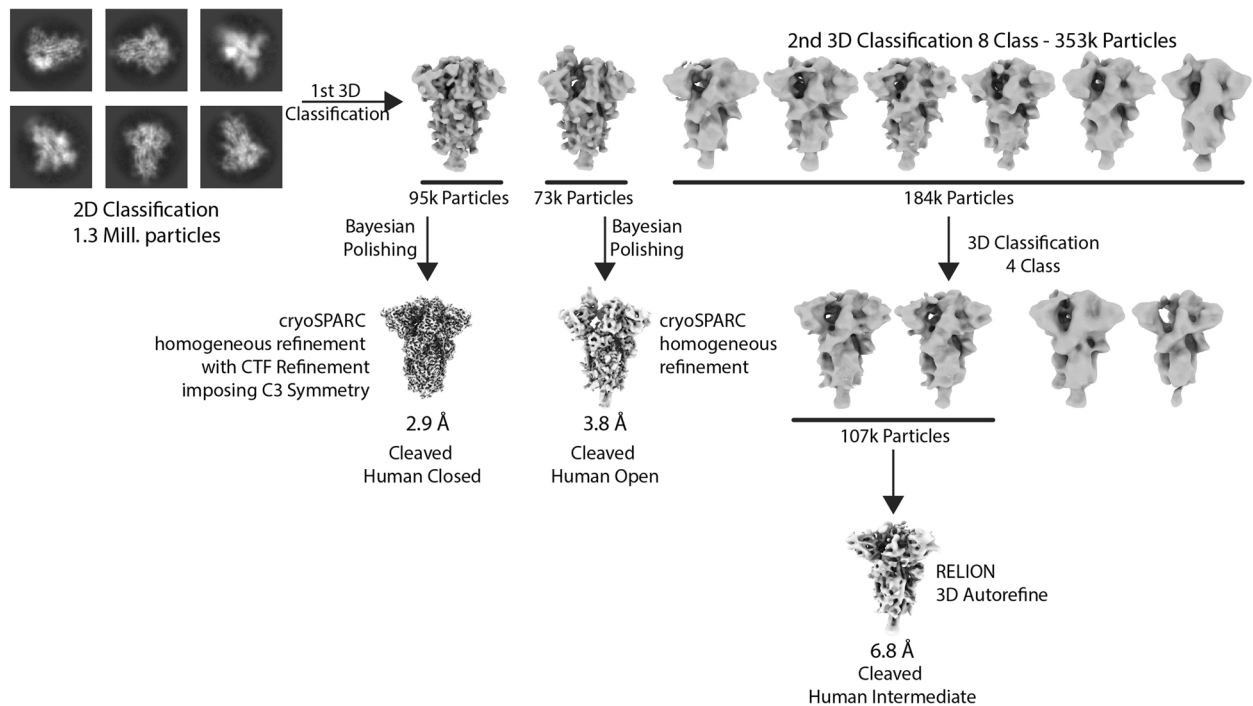
Bat



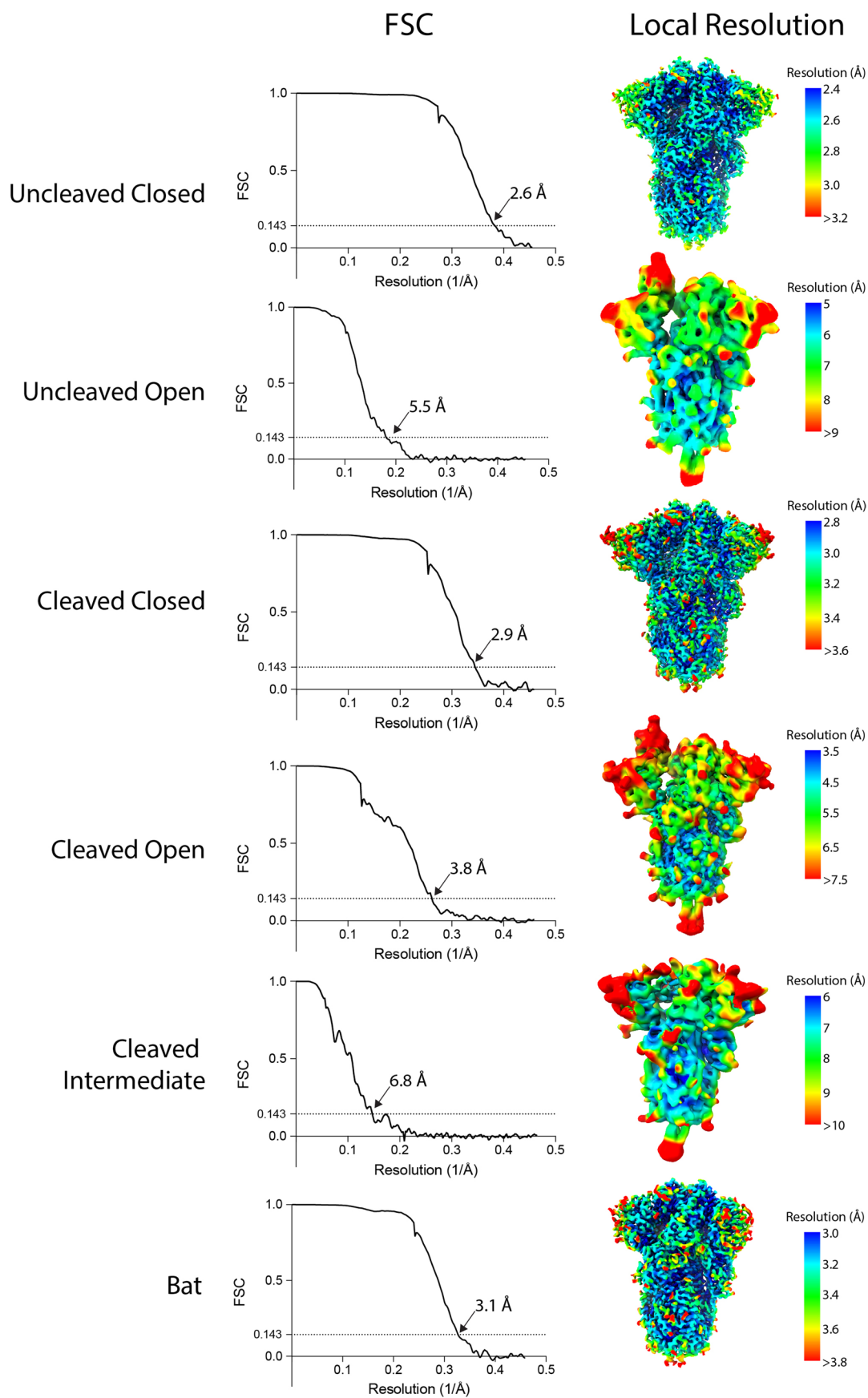
Uncleaved Human



Cleaved Human



Extended Data Fig. 4 | Processing schemes for cryo-EM data processing.



Extended Data Fig. 5 | Fourier Shell Correlations (FSCs) and local resolution estimates for calculated maps.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

CryoEM data collected using Thermo Scientific EPU

Data analysis

CryoEM data processed using following packages: RELION-3.1, cryoSPARC, CTFFind4, MotionCor2, crYOLO, PHENIX, namdinator, Coot

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Maps and models have been deposited in the Electron Microscopy Data Bank, <http://www.ebi.ac.uk/pdbe/emdb/> and models have been deposited in the Protein Data Bank <https://www.ebi.ac.uk/pdbe/> with the following identifiers: EMD-11203 and PDB 6ZGE (SARS-CoV-2 spike uncleaved closed); EMD-11207 and PDB 6ZGG (SARS-CoV-2 spike cleaved, closed); EMD-11206 and PDB 6ZGH (SARS-CoV-2 spike cleaved, intermediate); EMD-11205 and PDB 6ZGG (SARS-CoV-2 spike cleaved, open); EMD-11204 and PDB 6ZGF (bat coronavirus RaTG13 spike).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All cryoEM datasets consist of several thousand images. The number of images were sufficient to achieve the reported resolution, according to the most commonly reported resolution measures in cryoEM.
Data exclusions	cryoEM single particles were included and excluded using standard image processing classification techniques.
Replication	Biophysical measurements were repeated at least 3 times.
Randomization	Not Applicable
Blinding	Not Applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging