Published in final edited form as:
Nat Struct Mol Biol. 2020 August 01; 27(8): 763–767. doi:10.1038/s41594-020-0468-7.

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

Antoni G. Wrobel[#*,1], **Donald J. Benton**[#*,1], **Pengqi Xu**[4,1], **Chloë Roustan**[2], **Stephen R. Martin**[2], **Peter B. Rosenthal**[3], **John J. Skehel**[1], **Steven J. Gamblin**[*,1]

[1]Structural Biology of Disease Processes Laboratory

[2]Structural Biology Science Technology Platform

[3]Structural Biology of Cells and Viruses Laboratory, Francis Crick Institute, NW1 1AT London, United Kingdom

[4]Precision Medicine Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, 518107, Guangdong, China

[#] These authors contributed equally to this work.

Abstract

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 that of a closely related bat virus, RaTG13. We determined cryo-EM structures for both S proteins, and for SARS-CoV-2, in furin-cleaved and uncleaved forms; and compared them to recently reported structures for uncleaved SARS-CoV-2 S. We have also characterised biochemically 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 pre-cleavage form of human S and about 1000-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 species[1]. 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[2,3] and a bat virus RaTG13[2,6], was identified as the closest known relative of SARS-CoV-2. It is not known how SARS-CoV-2 evolved to infect

[*]to whom correspondence should be addressed: antoni.wrobel@crick.ac.uk, donald.benton@crick.ac.uk, steven.gamblin@crick.ac.uk.

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

Competing interests
The authors declare no competing interests.

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

Europe PMC Funders Author Manuscripts

humans but two mechanisms have been hypothesised: 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 spike glycoprotein (S) of SARS-CoV-2 mediates attachment of the virus to cell surface receptors and fusion between virus and cell membranes[1]. The receptor for SARS-CoV-2, like that for SARS-CoV[9,10], is the human cell-surface membrane protein ACE2[11–13]. Membrane fusion activity, as for other class-1 fusion glycoproteins[14], 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 RBD[2], a second difference between the spike proteins from human and bat viruses is the presence of a four-amino-acid insertion, PRRA, which adds a furin-cleavage site between S1 and S2 subunits[11]. Similar cleavage sites have been found in related coronaviruses including HKU1 and MERS, which infect humans[16–18], and the acquisition of similar cleavage sites is associated with increased pathogenicity in other viruses such as influenza[19].

In order to examine the evolutionary origin of SARS-CoV-2 and better understand the pandemic emergence, here we have characterised 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 characterised the furin-cleaved S protein of SARS-CoV-2 virus by cryo-EM (Fig. 1a, 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 cells[16] (Extended Data Fig. 1a). We further cleaved this protein using exogenous furin for structural and biochemical characterisation (Extended Data Fig. 1a).

The particles analysed from cryo-electron micrographs fell into three populations; a closed form (34%), an intermediate form (39%) and an open form (27%) with an upright Receptor Binding Domain (RBD) (Fig. 1a). The overall structure of the closed conformation of the S trimer is three-fold symmetric and similar to structures described previously using uncleaved material[11,20]. However, in both this cleaved sample and in an uncleavable form (discussed below) the closed conformation is more compact than in the previously published closed structure[11] (Fig. 1a). In the closed conformation, the surface of the RBD, which would interact with the ACE2 receptor, is buried inside the trimer and not accessible for receptor binding. In the intermediate form (Fig. 1a) two of the three RBDs maintain a similar interaction to the closed form but the third RBD is disordered. In addition, two of the NTDs that are in closest contact to the disordered RBD have shifted their centre of gravity by 2.5–2.9Å from the closed form (in the same direction as the open form) (Extended Data Figure 2e). In the open form (Fig. 1a), 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 is now fully exposed at the top of the assembly, while the NTD of the

adjacent chain moves toward the rotated domain, with the NTD of the same chain moving 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 Figure 1d.

In this protease-cleaved material, there is a much lower proportion of the S proteins in a closed conformation: 34% compared to 83% in the uncleaved human S trimer described below, compared to 65% and 50% reported in other recent studies[11,20]. The observation here of a substantially populated (39%) intermediate form, where one of the RBDs is disordered and two of NTDs have shifted, also suggests that this conformation, possibly transient, will also 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 a higher proportion of S proteins on the virus surface 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 much lower thermal stability of the cleaved protein (Extended Data Figure 1c) discussed below, 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 spike 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 Figure 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, Extended Data Figure 2), enabling us to model 15% more of the receptor binding domain (RBD, 100% complete) and 25% more of the N-terminal domain (NTD, 98% complete) than 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 crosslinking 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 from 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% identity) (Fig. 2b). As suggested before[11], the substitutions are clustered at two interfaces: the ACE2 receptor binding surface (considered below) 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 intra-trimer hydrogen bonds; including Arg403, Gln493 and Tyr505 from one subunit interacting 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 Figure 3). Indeed, surface contact area

calculations show that in the bat virus S trimer, the monomer:monomer interactions account for 5200 Å^2 (of which 485 Å^2 between the RBDs) whereas the equivalent contact area in the closed structure of the SARS-CoV-2 S trimer is 6100 Å^2 (with 550 Å^2 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, whereas the cleaved SARS-CoV-2 has a similar stability to the (uncleaved) bat virus protein (Extended Data Figure 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 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 immobilised onto the surface of a sensor and purified ACE2 was flowed over the surface to measure binding. Amplitude analysis suggests that SARS-CoV-2 S binds approximately 1000 times more tightly to ACE2 than the bat virus protein, 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[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 the SARS-CoV-2 S–ACE2, there is a buried surface area of 840 Å^2. As well as a series of specific salt and hydrogen bonds, another 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 Å^2. Structural comparison also suggests another substitution that likely contributes to the greatly enhanced affinity of SARS-CoV-2 S binding to ACE2: spike Gln493 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 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

Europe PMC Funders Author Manuscripts

infect humans directly. Given the modular nature of the human and bat spike glycoproteins, and the number and structural locations of the amino acid sequence differences between them, our observations support the involvement of recombination[8] between distinct coronavirus genomes in the generation of SARS-CoV-2.

The structure of the SARS-CoV-2 spike protein presented here is at high resolution and nearly complete, with many more external loops included than previously reported structures, 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 Methods

Design of protein constructs

The constructs corresponding to the ectodomain (residues 1-1208) of SARS-CoV-2 spike protein (NCBI reference sequence YP_009724390.1) and the ectodomain (residues 1-1204) of bat RaTG13 spike protein (QHR63300.2) were both codon-optimised for human expression, synthesised, and cloned into pcDNA.3.1(+) vector by GenScript with an N-terminal μ-phosphatase secretion leader sequence and a C-terminal hexa-histidine tag preceded by a foldon trimerisation 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 conformation[24].

The ectodomain (residues 19-615) of human ACE2 (NM_021804.2) was optimised for human expression, synthesised, 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% CO_2 atmosphere, at 37°C, shaking at 125 rpm. 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 1 litre of culture and ExpiFectamine 293 (Gibco) according to the manufacturer’s instructions. The supernatants were harvested twice: after 3-4 and 6-7 days; clarified, filtered, and incubated with appropriate affinity resin.

Spike proteins were bound to 5-7 mL of TALON cobalt beads (Takara) per litre of culture, washed briefly, and eluted with imidazole. ACE2 was bound to 4-6 mL of Streptactin XT resin (iba) per litre of cell culture supernatant, which had been pre-treated with the BioLock solution (iba). The beads were briefly washed and the protein 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 spike protein. Two units of the enzyme were used per 25 μg of the spike protein and the reaction was performed at 25°C in presence of 1 mM $CaCl_2$ and 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. 20μl reactions consisted of 5 μg of protein with SYPRO Orange (Sigma) present at a 5x concentration, diluted from the 5000x 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 spike 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 1000 rpm. 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 immobilised on them at 15-30 μg/mL for 30-40 minutes. ACE2 binding was measured using a 3-5 min association step followed by a 10-15 minute 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. The equilibrium dissociation constant (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 spike was frozen at 0.4 mg/mL 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 optimisation and crosslinking had to be performed to obtain RaTG13 spike protein suitable for cryoEM. The protein in MES buffer (50 mM MES pH 6.0, 100 mM NaCl) was treated with BS^3 (Thermo Scientific) at a final concentration of 0.5 mM on ice for an hour and then a Grafix protocol[25] was performed to achieve double cross-linking: the BS^3-crosslinked 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 spun at 35,000 rpm 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 crosslinked protein identified with SDS-PAGE were pooled, concentrated, and gel-filtered into 50 mM MES pH 6.0, 100 mM NaCl. The resulting bat spike protein as well as the furin-treated human spike were frozen at 0.1 mg/mL 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 μl of a sample to a grid equilibrated to 4˚C in 100% humidity, followed by a 4-5 s blot using a Vitrobot MkIII, 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/Å^2, 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/Å^2, fractionated into 32 frames, with a calibrated pixel size of 1.08 Å. All datasets were collected using defoci between 1.5-3 μm.

Cryo-EM data processing

Movie frames were aligned using MotionCor2[26] implemented in RELION[27] and Contrast Transfer Function fitted using CTFfind4[28]. All subsequent data processing was carried out using both RELION and cryoSPARC[29]. Particles for the furin-uncleavable SARS-CoV-2 spike dataset were picked using crYOLO[30] 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 2D classification, retaining classes with clear secondary structure. These particles were classified using RELION 3D 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 blocres[31] implemented in cryoSPARC. Maps were filtered by local resolution and globally sharpened[32] in cryoSPARC. Additional information is available in Table 1 and Extended Data Fig. 5.

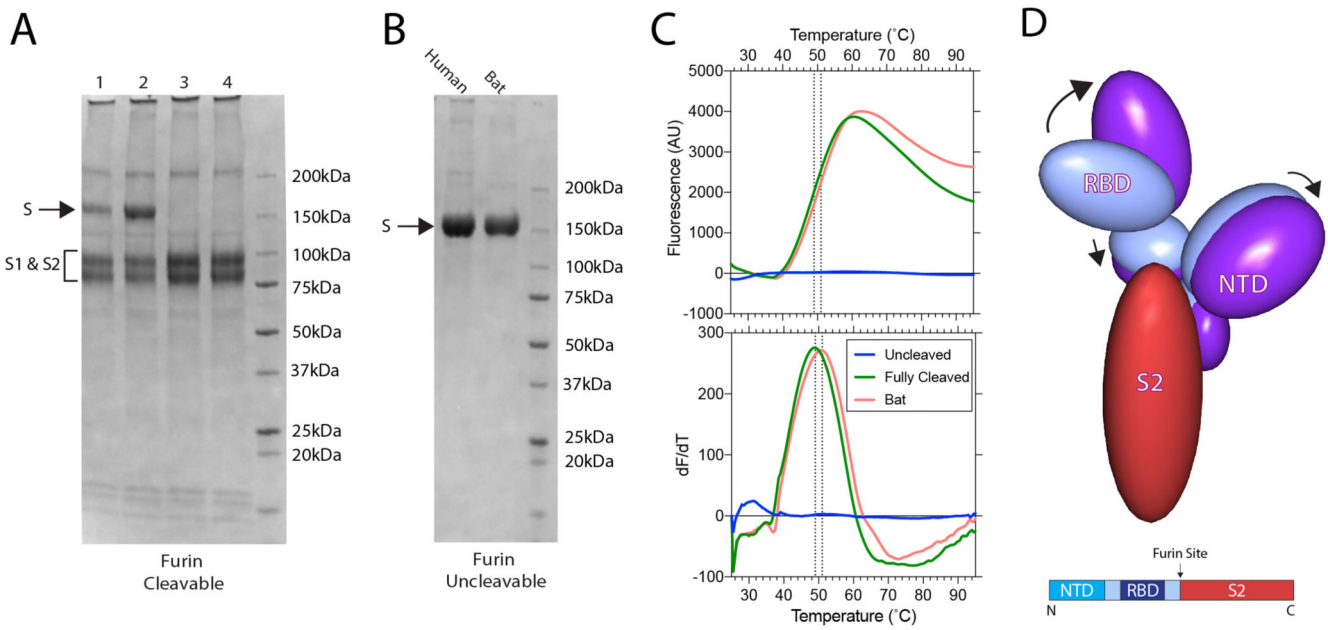
Model building

The model for the uncleavable SARS-CoV-2 spike 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 Coot[33]. 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 PHENIX[34].

The intermediate and open structures was 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 Namdinator[35], followed by geometry normalization using PHENIX. Additional information is available in Table 1.

Europe PMC Funders Author Manuscripts

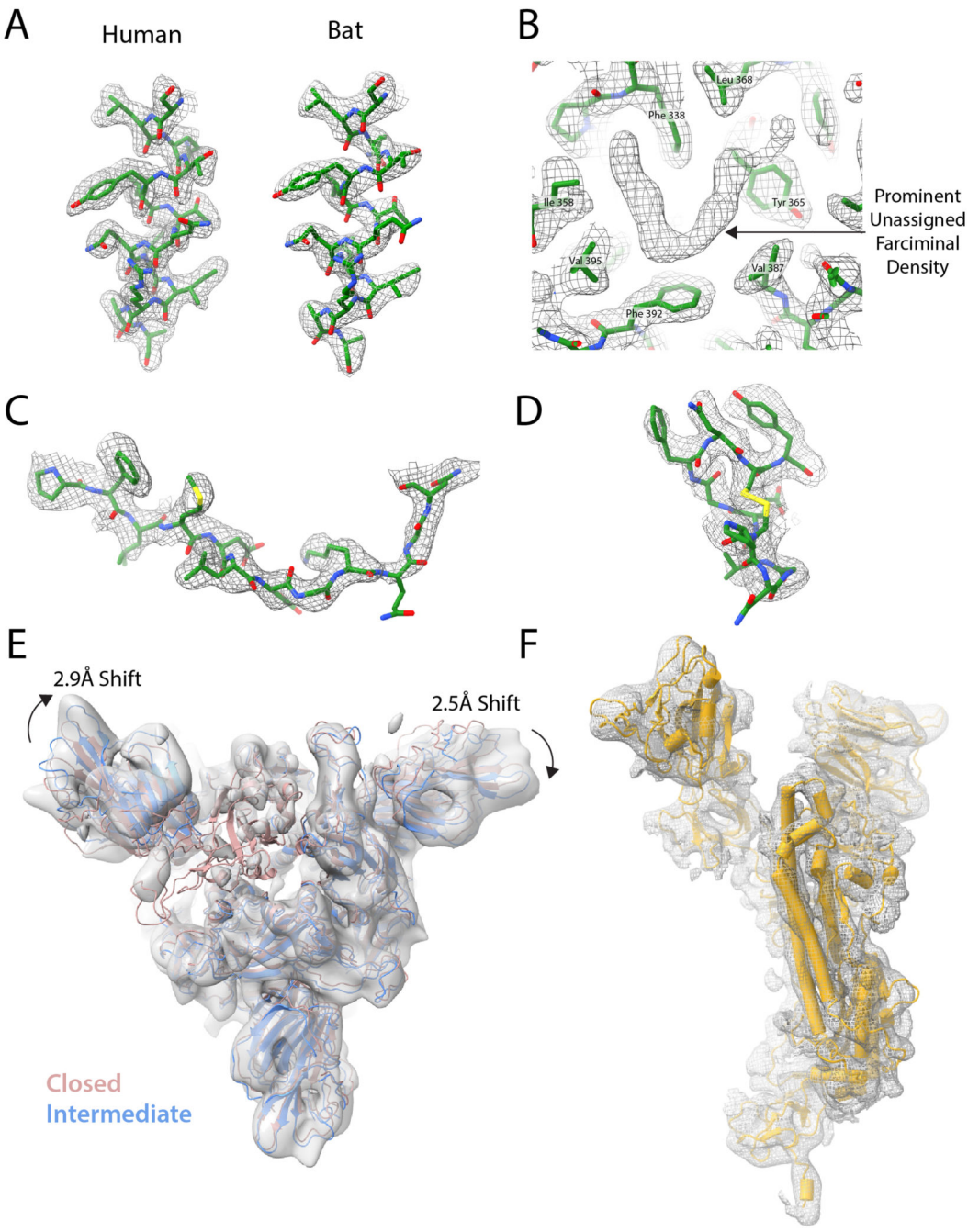
Extended Data



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 bat and human virus S proteins. (C) Differential Scanning Fluorimetry measurement of melting temperature for uncleavable human and bat, and fully-furin-cleaved SARS-CoV-2 proteins. (D) The changes in domain orientation, between the closed and open forms, 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.

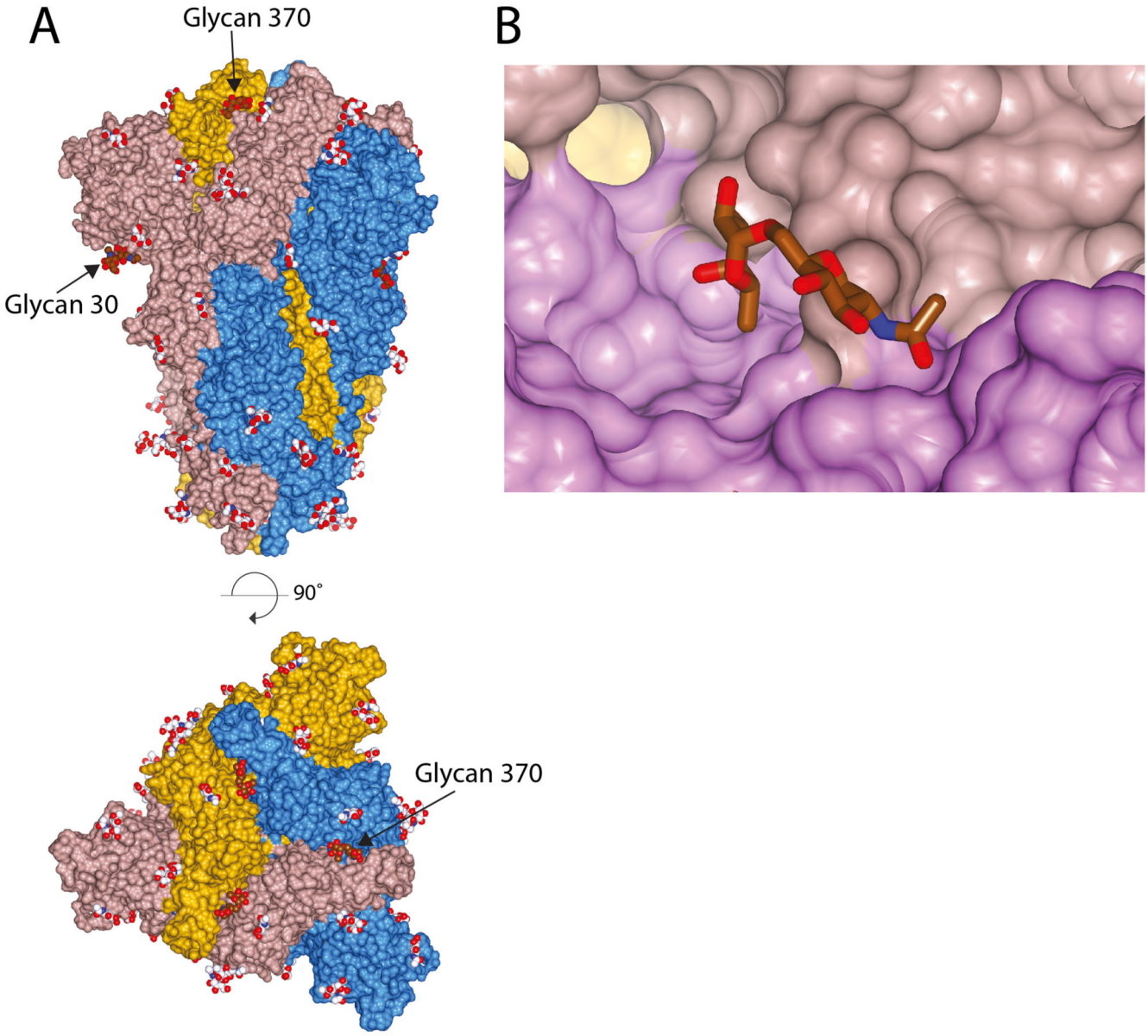
Europe PMC Funders Author Manuscripts



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 human and bat structures. (B) Unassigned farciminal density seen in maps of the closed conformation of human spike protein. (C+D) Typical density for examples of previously unbuilt external loops in uncleaved closed human 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 spike. The fitted model (blue) is compared to that of the closed conformation (pink). Shifts in the NTDs are indicated with

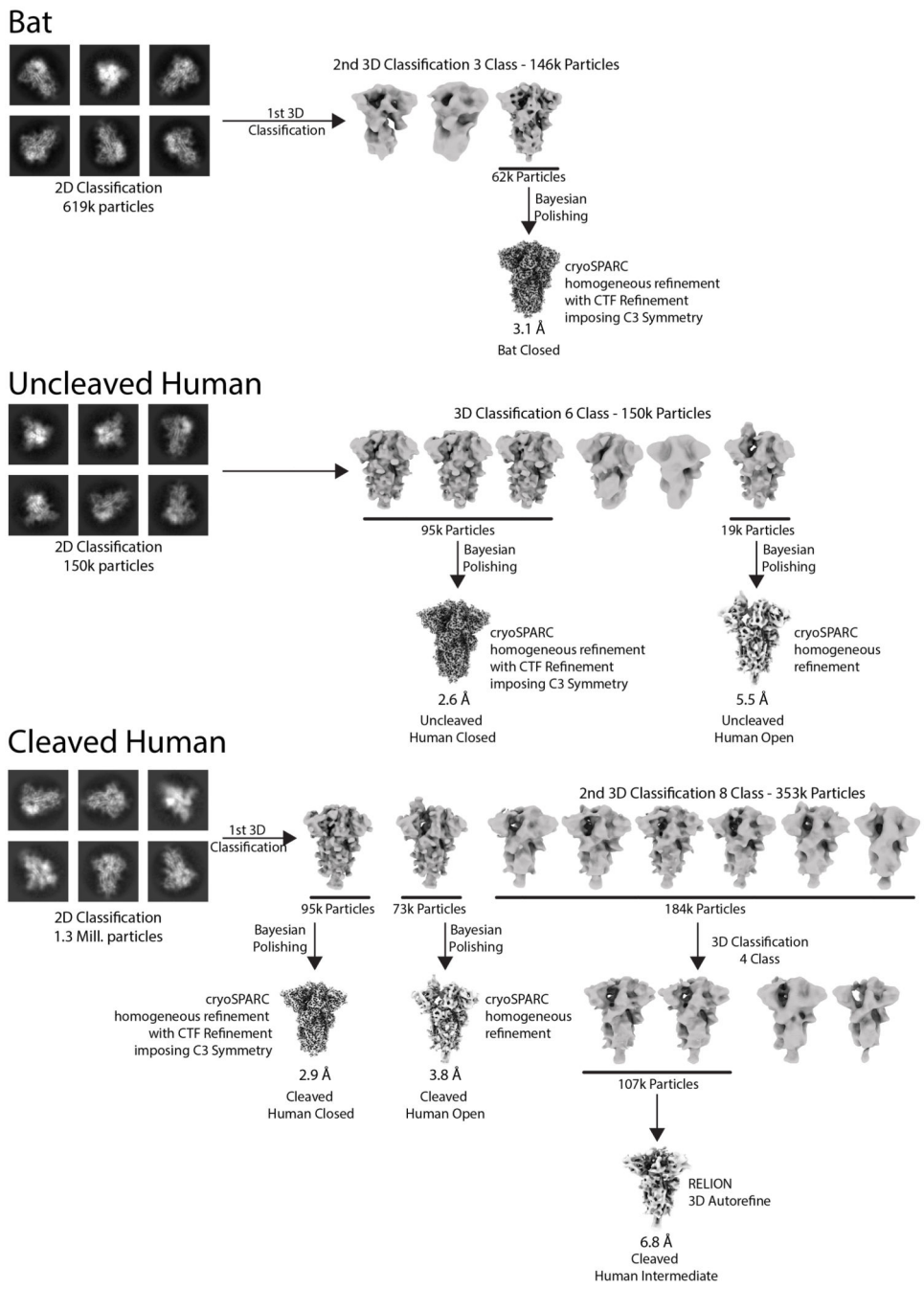
arrows, with values calculated from translations of their centres of mass. (F) Example EM density of a monomer of the Intermediate conformation map with the built structure shown in yellow.



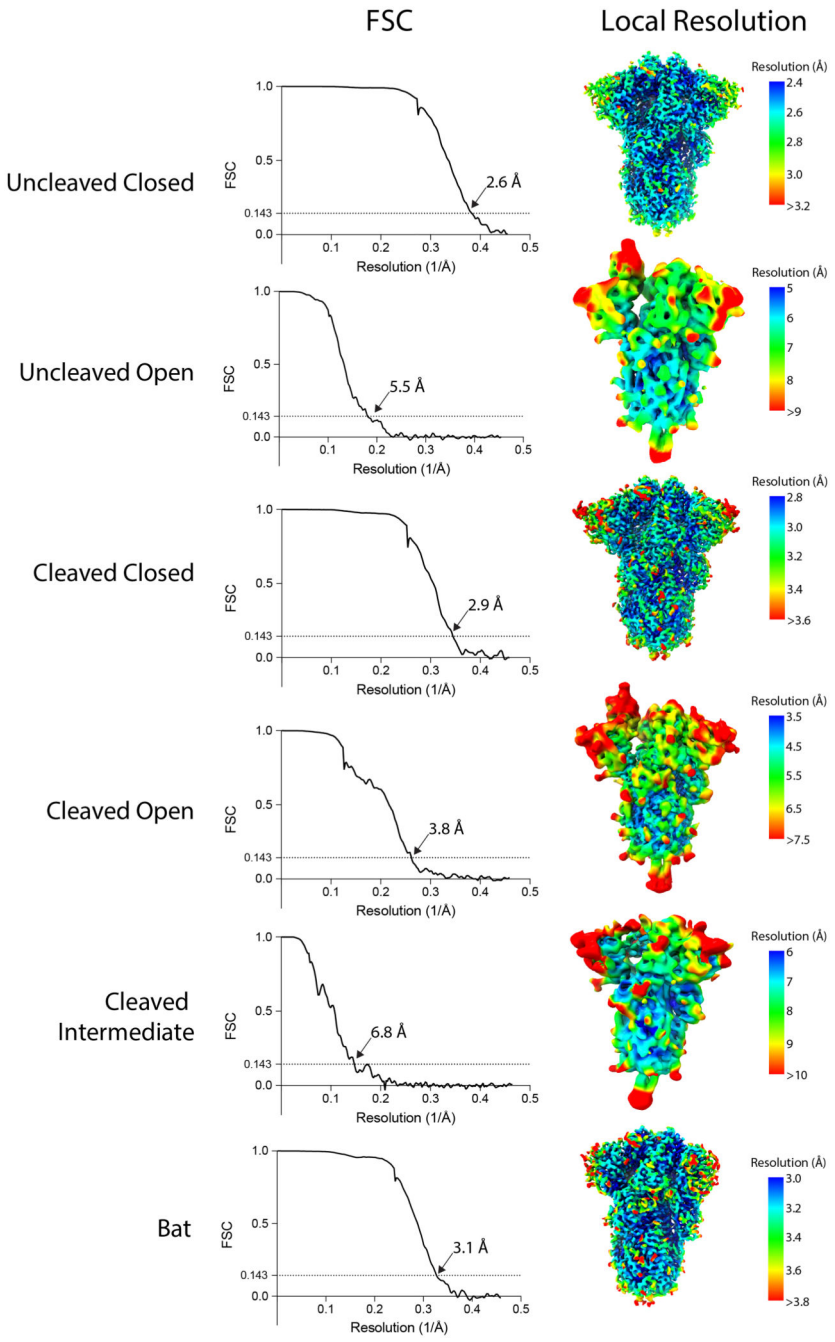
Extended Data Fig. 3. N-linked glycosylation of bat spike protein.

(A) All buildable glycans shown with bat-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.

Europe PMC Funders Author Manuscripts



Extended Data Fig. 4. Fourier Shell Correlations (FSCs) and local resolution estimates for calculated maps.



Extended Data Fig. 5. Processing schemes for cryo-EM data processing.

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 & the Sanming Project of Medicine in Shenzhen (SZSM201911003)

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 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).

References

1. Lai M, Perlman S, Anderson L. Coronaviridae. Fields Virology. 2013:1305–1336.
2. Zhou P, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579:270–273. [PubMed: 32015507]
3. Wu F, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579:265–269. [PubMed: 32015508]
4. Li W, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science (80-). 2005; 310:676–679.
5. Yang L, et al. Novel SARS-like betacoronaviruses in bats, China, 2011. Emerg Infect Dis. 2013; 19:989–991. [PubMed: 23739658]
6. Ge XY, et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol Sin. 2016; 31:31–40. [PubMed: 26920708]
7. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nature Medicine. 2020; 26:450–452.
8. Li X, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci Adv. 2020; eabb9153 doi: 10.1126/sciadv.abb9153 [PubMed: 32937441]
9. Li F, Li W, Farzan M, Harrison SC. Structural biology: Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science (80-). 2005; 309:1864–1868.
10. Li W, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003; 426:450–454. [PubMed: 14647384]
11. Walls, AC; , et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. 2020.
12. 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. 2020; 5:562–569. [PubMed: 32094589]
13. Hoffmann M, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020; 181:271–280. e8 [PubMed: 32142651]
14. Harrison SC. Viral membrane fusion. Virology. 2015; doi: 10.1016/j.Virol2015.03.043
15. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A. 2009; 106:5871–5876. [PubMed: 19321428]
16. Millet JK, Whittaker GR. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res. 2015; 202:120–134. [PubMed: 25445340]
17. Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci U S A. 2014; 111:15214–15219. [PubMed: 25288733]
18. Chan CM, et al. Spike protein, S, of human coronavirus HKU1: Role in viral life cycle and application in antibody detection. Exp Biol Med. 2008; 233:1527–1536.

19. Steinhauer DA. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. Virology. 1999; 258:1–20. [PubMed: 10329563]
20. Wrapp D, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science (80-). 2020; doi: 10.1126/science.aax0902
21. Shang J, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020; :1–4. DOI: 10.1038/s41586-020-2179-y
22. Yan R, et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science (80-). 2020; 367:1444–1448.
23. Lan J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020; 581:215–220. [PubMed: 32225176]
24. Pallesen J, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc Natl Acad Sci U S A. 114:E7348–E7357.2017; [PubMed: 28807998]
25. Kastner B, et al. GraFix: Sample preparation for single-particle electron cryomicroscopy. Nat Methods. 2008; 5:53–55. [PubMed: 18157137]
26. Zheng SQ, et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods. 2017; 14:331–332. [PubMed: 28250466]
27. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012; 180:519–530. [PubMed: 23000701]
28. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015; 192:216–221. [PubMed: 26278980]
29. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017; 14:290–296. [PubMed: 28165473]
30. Wagner T, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol. 2019; 2
31. Cardone G, Heymann JB, Steven AC. One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. J Struct Biol. 2013; doi: 10.1016/Jjsb.2013.08.002
32. Rosenthal PB, Henderson R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. J Mol Biol. 2003; 333:721–745. [PubMed: 14568533]
33. Emsley P, Lohkamp B, Scott WG, Cowtan K. IUCr. Features and development of *Coot* . Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
34. Adams PD, et al. *PHENIX* a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
35. Kidmose RT, et al. Namdinator -Automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. IUCr J. 2019; 6:526–531.

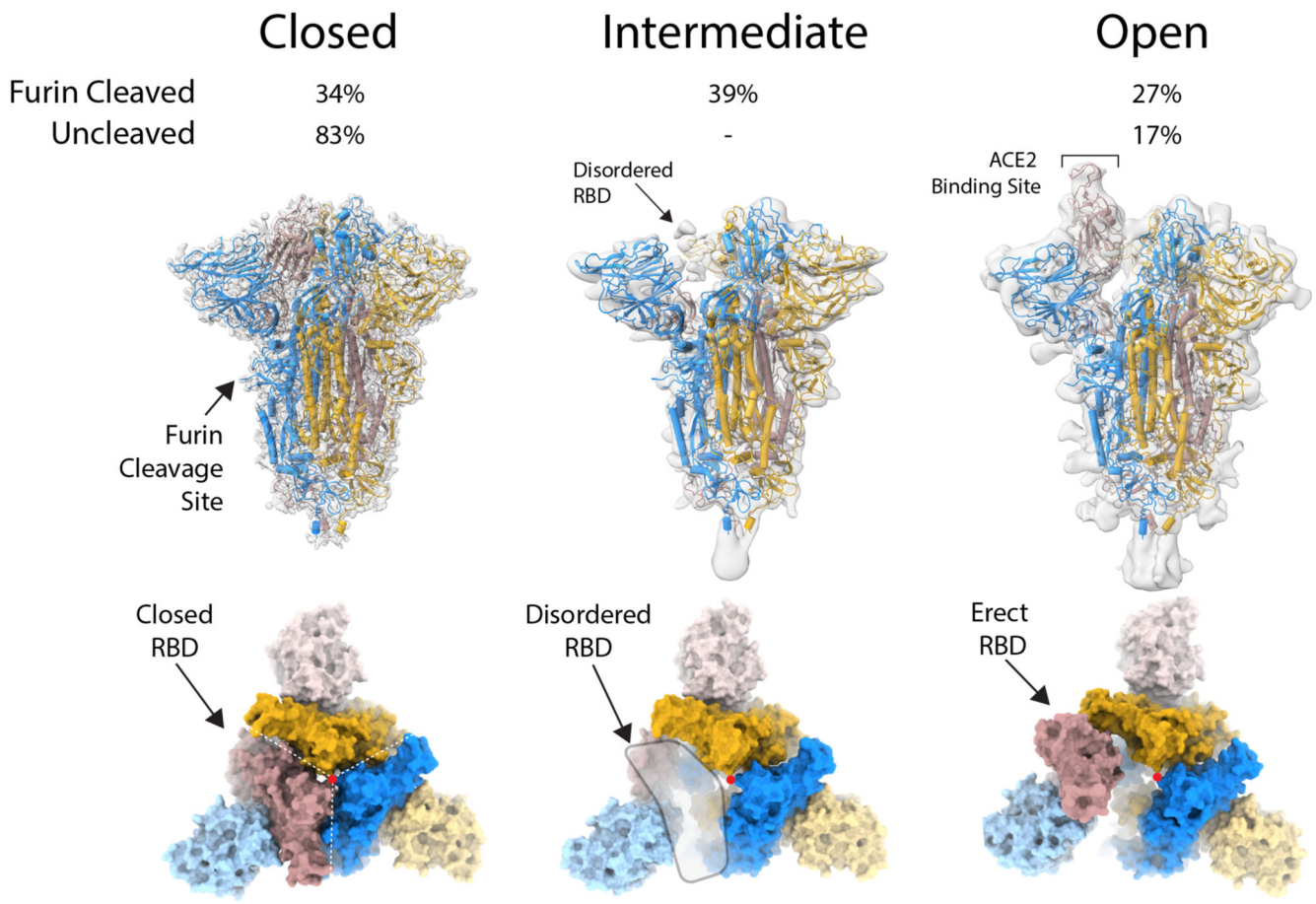


Figure 1. Structure of protease-cleaved SARS-CoV-2 spike glycoprotein.

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 with a small proportion in the open conformation. Density maps for the three types of particles, overlaid with a ribbon representation of the built molecular models, as viewed with the long axis of the trimer vertical (middle panel); the three monomers are coloured blue, yellow and brown. An orthogonal view (lower panel) looking down the long axis (indicated by a red dot), the colouring is as in the middle panel with the NTDs in a lighter hue.

Europe PMC Funders Author Manuscripts

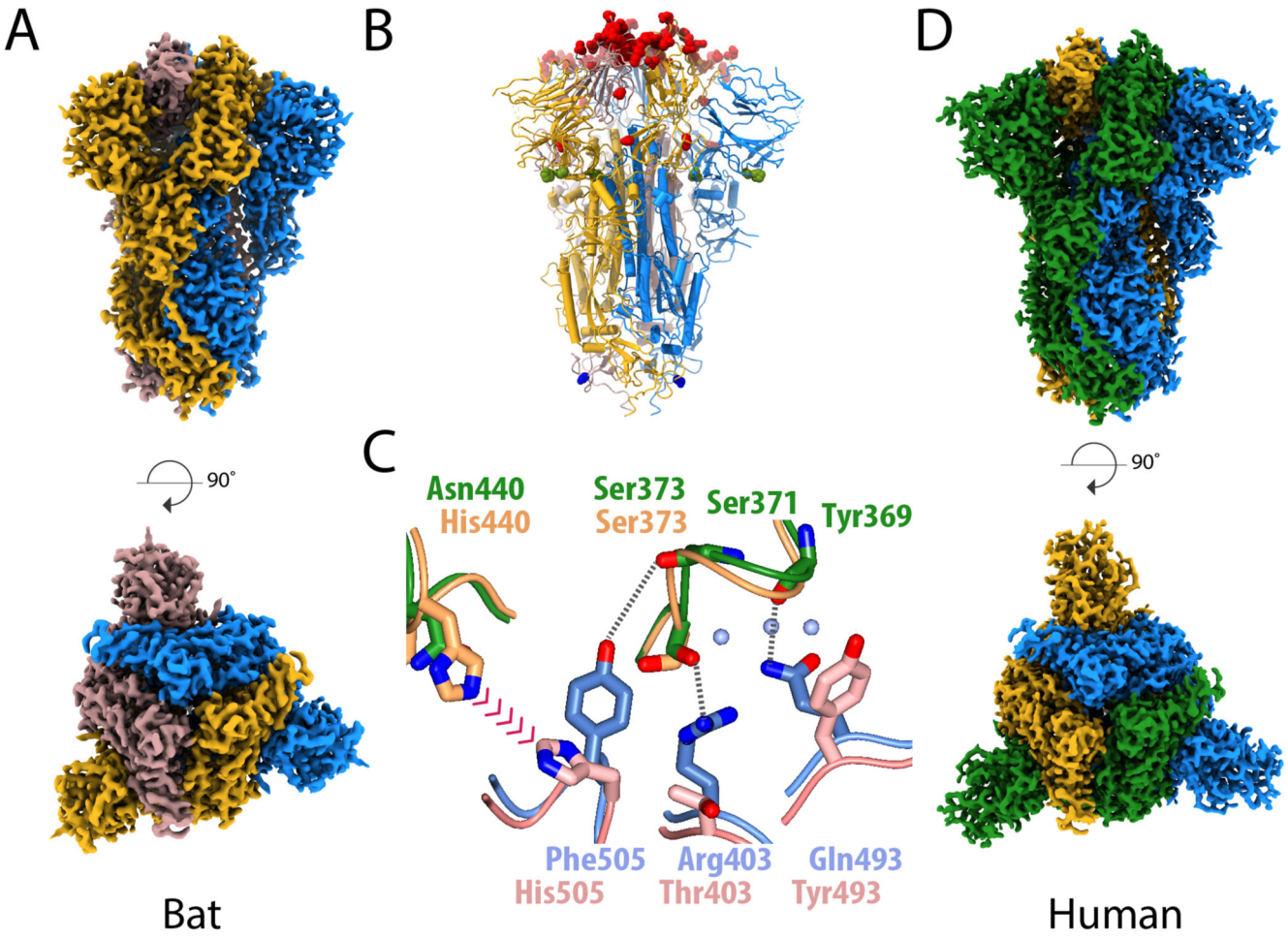


Figure 2. Structural comparison of the spike glycoprotein from the bat virus RaTG13 and SARS-CoV-2.

(A) The density map for the bat virus trimeric S is shown with the long axis vertical in the top panel and viewed orthogonally 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 coloured blue, yellow and brown. (B) Molecular model of the bat virus S protein, coloured as in (A), with substitutions between the bat virus and SARS-CoV-2 highlighted. Most of the changes are in the RBD and coloured red, there are four substitutions in S1 outside of the RBD, which are shown in green, and a single substitution in S2 shown in blue. (C) Overlay of the molecular structure of a portion of the RBD–RBD interface; the two bat virus S monomers are coloured gold (upper) and pink (lower) while the two superposed SARS-CoV-2 S RBD chains are shown in green (upper) and blue (lower). Analysis suggests that the residues at the interface of SARS-CoV-2 S RBD chains support several additional stabilising interactions and avoid potential steric repulsion between His505 and His440 seen in the bat virus structure. (D) The density map for the uncleaved SARS-CoV-2 S protein, in the closed conformation, shown in the same orientation as (A) with the subunits coloured blue, green and yellow. This sample gave the best quality maps and enabled the most extensive build of the polypeptide chain.

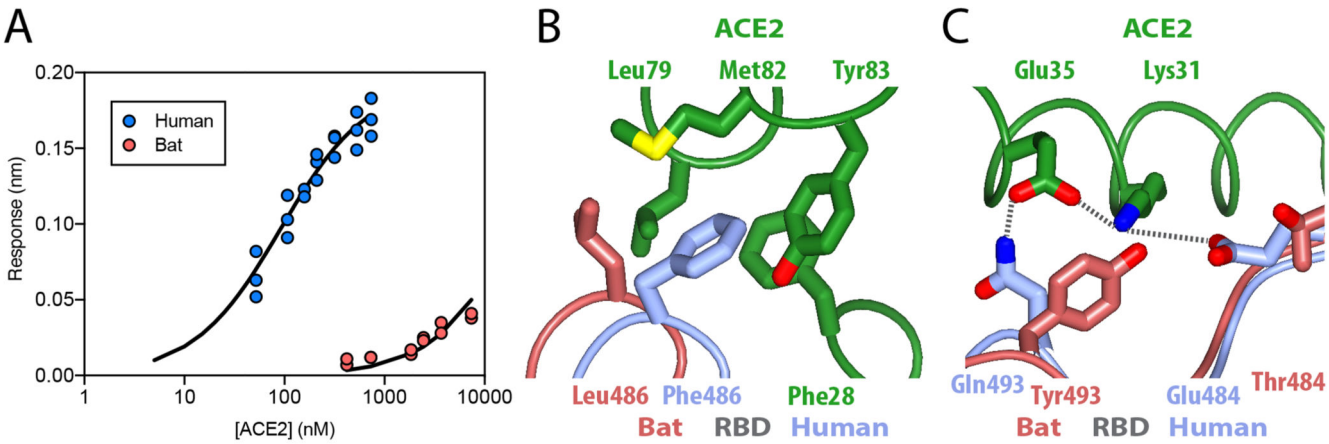


Figure 3. Binding of ACE2 receptor to bat virus and SARS-CoV-2 spike protein.

(A) Plot of surface biolayer amplitude measurement as a function of ACE2 concentration with the data for spike from SARS-CoV-2 (blue, Kd calculated as 91 ± 18 nM) and from the bat virus (red, Kd estimated to be >40 μM). The equilibrium dissociation constant for SARS-CoV-2 protein calculated from kinetic constants (koff = 0.0105 s^{-1} and kon = 1.56 x 10^5 $M^{-1}s^{-1}$) was 67.5 +/- 9 nM. (B & C) Ribbon representation of modelled molecular interactions between ACE2 (green) with RBD from spike in SARS-CoV-2 (blue) (both PDB 6VW1[21]) and bat virus (brown, this study). (B) Details of a hydrophobic pocket on ACE2 that accommodates a phenylalanine residue from the SARS-CoV-2 S RBD. (C) Shows 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 SARS-CoV-2 RBD Tyr493 with ACE2 Lys31.

Europe PMC Funders Author Manuscripts

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 -3.0	-1.5 to -3.0	-1.5 to -3.0	-1.5 to -3.0	-1.5 to -3.0
Pixel size (Å)	1.09	1.08	1.08	1.08	1.08
Symmetry imposed	*C*3	*C*3	*C*1	*C*1	*C*3
Final particle images (no.)	95 K	95 K	107 K	73 K	62 K
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	3.0–3.8
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 *B* factor (Å^2)	-109.8	-91.2	-207.0	-61.7	-88.2
Model composition					
Non-hydrogen atoms	26991	26901	23735	25111	26169
Protein residues	3294	3294	3032	3207	3180
Ligands	81	78	-	-	93
B factors (Å^2)					
Protein	35.1	47.8	-	-	37.13
Ligand	83.0	77.5	-	-	52.15
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	1.130
Validation					
MolProbity score	1.33	1.21	1.69	1.79	1.70
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	0.30	0.50	0.00