Evolution of the SARS-CoV-2 spike protein in the human host

Antoni Wrobel (✉ antoni.wrobel@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0002-6680-5587>

Donald Benton

Francis Crick Institute <https://orcid.org/0000-0001-6748-9339>

Chloë Roustan

The Francis Crick Institute

Annabel Borg

The Francis Crick Institute

Saira Hussain

The Francis Crick Institute

Stephen Martin

Structural Biology Science Technology Platform, the Francis Crick Institute, Mill Hill Laboratory, London, UK.

Peter Rosenthal

The Francis Crick Institute <https://orcid.org/0000-0002-0387-2862>

John Skehel

Francis Crick Institute

Steve Gamblin

The Francis Crick Institute <https://orcid.org/0000-0001-5331-639X>

Biological Sciences - Article

Keywords: SARS-CoV-2, virology, microbiology

Posted Date: May 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-535704/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on March 4th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-28768-w>.

Evolution of the SARS-CoV-2 spike protein in the human host

Antoni Wrobel (✉ antoni.wrobel@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0002-6680-5587>

Donald Benton

Francis Crick Institute <https://orcid.org/0000-0001-6748-9339>

Chloë Roustan

The Francis Crick Institute

Annabel Borg

The Francis Crick Institute

Saira Hussain

The Francis Crick Institute

Stephen Martin

Structural Biology Science Technology Platform, the Francis Crick Institute, Mill Hill Laboratory, London, UK.

Peter Rosenthal

The Francis Crick Institute <https://orcid.org/0000-0002-0387-2862>

John Skehel

Francis Crick Institute

Steve Gamblin

The Francis Crick Institute <https://orcid.org/0000-0001-5331-639X>

Biological Sciences - Article

Keywords: SARS-CoV-2, virology, microbiology

DOI: <https://doi.org/10.21203/rs.3.rs-535704/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Evolution of the SARS-CoV-2 spike protein in the human host

Antoni G. Wrobel^{#1}, Donald J. Benton^{#1}, Chloë Roustan², Annabel Borg², Saira Hussain^{3,4}, Stephen R. Martin¹, Peter B. Rosenthal⁵, John J. Skehel¹, and Steven J. Gamblin^{*1}

¹Structural Biology of Disease Processes Laboratory, ²Structural Biology Science Technology Platform, ³Worldwide Influenza Centre, ⁴RNA Virus Replication Laboratory, ⁵Structural Biology of Cells and Viruses Laboratory; Francis Crick Institute, NW1 1AT London, United Kingdom

[#] equal contribution

^{*}to whom correspondence should be addressed: antoni.wrobel@crick.ac.uk, donald.benton@crick.ac.uk, steve.gamblin@crick.ac.uk

Variants of SARS-CoV-2 have emerged which contain multiple substitutions in the surface spike glycoprotein that have been associated with increased transmission and resistance to neutralising antibodies and antisera. We have examined the structure and receptor binding properties of spike proteins from the B.1.1.7 (UK) and B.1.351 (SA) variants to better understand the evolution of the virus in humans. Both variants' spikes have the same mutation, N501Y, in their receptor-binding domains that confers tighter ACE2 binding and this substitution relies on a common earlier substitution (D614G) to achieve the tighter binding. Each variant spike has also acquired a key change in structure that impacts virus pathogenesis. Unlike other SARS-CoV-2 spikes, the spike from the UK variant is stable against detrimmerisation on binding ACE2. This feature primarily arises from the acquisition of a substitution at the S1-S2 furin site that allows for near-complete cleavage. In the SA variant spike, the presence of a new substitution, K417N, again on the background of the D614G substitution, enables the spike trimer to adopt fully open conformations that are required for receptor binding. Both types of structural change likely contribute to the increased effectiveness of these viruses for infecting human cells.

The SARS-CoV-2 spike glycoprotein is the major surface antigen of the virus. Its function is to bind the host receptor ACE2 and mediate the subsequent membrane fusion required for cell entry^{1–8}. The virus has evolved in the human host during the pandemic^{9–11} and we and others have demonstrated that the predominant D614G substitution, located on a monomer-monomer interface of the spike trimer, increases its ability to adopt the open conformation that is competent to bind receptor^{12–14}. Recently emerging variants of SARS-CoV-2 have acquired a number of other substitutions in the spike including those on monomer-monomer interfaces, at the receptor-binding site, and proximal to the furin-cleavage site (**Fig. S1**). Here we have examined the structure and receptor binding properties of the B.1.1.7 variant first described in Kent, United Kingdom (UK)^{10,15–18} and the B.1.351 variant first described in South Africa (SA)^{19,20}. The results provide a molecular explanation for the increased human transmissibility of the variants.

First, examination of the structure of ACE2-bound spike protein of the UK variant shows that all of the spikes are present as trimers (**Fig. 1a, Supp. Table 1**). This is the first cleaved SARS-CoV-2 spike protein in complex with ACE2 we have observed to be fully trimeric. All other spike structures, when mixed with ACE2, show a substantial proportion of the material dissociated into monomers⁵. For example, in our previous study of the furin-cleaved spike of the original Wuhan strain (Wuhan) complexed with ACE2 (**Fig. 1a**), we observed that more than 70% of the particles present were monomeric S/ACE2 complexes⁵. Notably, we also observed that this spike material from the UK variant is almost fully cleaved into S1 and S2 (**Fig. S2**), which has also not been reported before⁸. This observation is consistent with the fact that one of the mutations in the UK spike is the substitution P681H at the furin site which generates a more basic cleavage site (**HRRAR**).

To test the possibility that it is the near complete cleavage of the S1-S2 subunits that is responsible for the greater stability of the trimeric ACE2 bound S1/S2 complex, we expressed the UK spike in the presence of a furin inhibitor, decanoyl-Arg-Val-Lys-Arg-chloromethylketone, which resulted in a completely uncleaved protein (**Fig. S2**). Incubation of this uncleaved spike with ACE2 resulted in more than 50% of particles being monomeric S1-S2/ACE2 (**Fig. 1a, S3**). These data support the conclusion that

the near complete cleavage at the variant furin site enables spike to remain stable as a trimer with ACE2 bound.

To better understand this phenomenon, we compared monomer/ACE2 complexes from a number of SARS-CoV-2 variants (**Figs. S4-5**) and found that they all adopt a similar overall conformation that, as noted before⁵, is markedly different from the conformation adopted by the S/ACE2 complex in the trimeric species. As shown in **Fig. 1b**, the monomeric and trimeric ACE2 complexes can be aligned very closely on the ACE2/RBD components (coloured rosy brown in the figure) but, the NTD-associated subdomain (light blue) and the NTD (blue) are rotated by 95° and their centre of mass translated by 25 Å in the monomer, resulting in a conformation incompatible with the trimer.

The fact that fully cleaved UK variant spike is able to remain stable as a trimer when ACE2 bound indicates that the strain induced by receptor binding to RBD⁵ is relieved if cleavage has occurred between the S1 and S2 subunits. The cleavage site lies at the C-terminus of the NTD-subdomain (shown on **Fig. 1b**), the individual centre of mass of which is translocated 38 Å in the monomer/ACE2 complex compared to the fully ACE2-bound trimer. Cleavage appears to be able to release sufficient strain to allow the resulting ACE2 complex to remain stable as a trimer. The preponderance of monomer S/ACE2 species in other SARS-CoV-2 variants, such as Wuhan (**Fig. 1a**) and SA (**Fig. S4**), is probably accounted for by the lower levels of cleavage of these spikes (**Fig. S2**), although our structures of unbound, uncleaved UK spike (**Fig. S5-6**) suggest the trimeric state of the receptor-bound form of the UK variant spike might be further stabilised by the substitutions D1118H and A570D on the inter-monomer interfaces (**Fig. S5**). The increased stability of the receptor-bound UK spike may enhance virus infectivity by increasing the avidity of binding to cells and by the priming mechanism, which we have previously described⁵, in which ACE2 binding exposes the S2 subunit for helical rearrangements associated with membrane fusion.

Second, we studied binding of ACE2 to spike variants by surface biolayer interferometry. The data (**Fig. 2a, b; Fig. S7**) show a six-fold increase in binding strength for UK spike, and a two-fold increase for SA spike (compared to Wuhan) arising from the shared substitution N501Y in the RBD. The substitution of the

asparagine at position 501 in Wuhan for a tyrosine residue in both UK & SA variants (**Fig. 2c**) leads to an increase in hydrophobic interactions between the aromatic ring of Y501_(RBD) and the aromatic ring of Tyr-41_(ACE2) and the aliphatic moiety of Lys-353_(ACE2), in addition to a charged hydrogen bond between the phenolic hydroxyl of Tyr-501_(RBD) and Lys-353_(ACE2) (**Fig. 2d**). Consistent with the smaller increase in affinity for ACE2 of the SA spike versus UK spike, is the finding that, whereas UK has retained the same salt bridge between Lys-417_(RBD) and Asp-30_(ACE2) as Wuhan, the RBD in SA has acquired the additional substitution of an asparagine at RBD residue 417 which cannot make this salt bridge (**Fig. 2e**).

To understand the evolution of receptor binding, we also expressed UK spike with an aspartic acid, rather than glycine residue, at position 614. The substitution D614G (relative to Wuhan) occurred earlier in the evolution of SARS-CoV-2, became the predominant global form of the virus⁹ and continues to be present in the UK and SA variant forms of the virus. Remarkably, the engineered G614D UK spike (Y501, D614) shows the same binding affinity as Wuhan (N501, D614) (**Fig. 2a,b; Fig. S7**). Thus, the additional binding energy, generated by the substitutions at the receptor-binding interface, is only realised if residue 614 is also switched from aspartic acid to glycine. We have previously shown that the D614G substitution leads to destabilisation of the closed conformation thus promoting spike opening for subsequent receptor binding and membrane fusion¹². This occurs when residue 614 is an aspartic acid (on the S1 NTD-subdomain) and it creates a salt bridge with Lys-854 (on a neighbouring S2 subunit) stabilising the closed form of the trimer^{5,12,21}. Mutation to glycine at residue 614 precludes this interaction and thus disfavors the closed-form of spike. We have previously envisaged a low energy barrier to the closed to open transition and the present data show that the G614 (open) form of spike is necessary for the Y501 substitution to produce tighter binding.

Third, the most striking feature of the structure of SA spike revealed by cryoEM is that all the trimers adopt an open conformation with either one (64%) or two (36%) of the RBDs in the erect, receptor-binding-competent form (**Fig. 3a, Fig S6**). This observation is in marked contrast to the earliest SARS-CoV-2 viruses where 83% of Wuhan spike particles were in the closed form (**Fig. 3a**)⁸ and is noteworthy with

regards to the subsequent D614G form of the protein where we observed 87% of spikes in an open conformation, out of which only 18% were in the two-RBD-erect conformation (**Fig. S6**)¹². Inspection of the sequence of the SA spike and comparison of its structure with that of the closed form of Wuhan spike and the open and closed forms of G614 spikes, suggests that the complete opening of SA spike is driven by the substitution K417N on the background of G614. In Wuhan, Lysine at residue 417 on the RBD not only makes an aliphatic packing interaction with Tyr-369 of the neighbouring subunit, it also forms a salt bridge/hydrogen bond network with Glu-406 and Arg-403 of the RBD and with Ser-373 of the neighbouring RBD that stabilises the closed conformation. In contrast, the substitution of an asparagine at position 417 in SA removes an intramolecular salt-bridge and would generate a steric clash with Tyr-369 of the neighbouring RBD leading to destabilisation of the closed form and thus promotion of the open form (**Fig. 3b**). In the same way that G614 is a prerequisite for realising tighter receptor binding by the substitution N501Y described above, it likely also enables spike protein from the SA virus to achieve a fully open conformation as a result of the K417N substitution.

We showed before that the extent of cleavage at the furin site influences the proportion of spike molecules being open and thus able to bind receptor⁸. We and others also showed that the D614G substitution in the spike, which was acquired early in the pandemic, similarly acted to increase the proportion of open forms of spike^{12–14}. We now show that the emergence of UK spike, which is completely cleaved, and SA spike, which does not adopt the closed conformation, represent two related additional steps in viral adaptation to the human host. Modifications in the spike glycoprotein during evolution of the SARS-CoV-2 virus in humans therefore have made the virus more infectious by increasing the stability of trimeric spike, by tighter receptor binding, and by promotion of the open forms of spike. Although as yet uncharacterised at the molecular level, the variant of concern B.1.617^{22,23}, which recently emerged in India, contains the substitution P681R which we suspect will have a similar impact on the virus' spike structure, and thus its pathogenesis, as the P681H substitution in the UK variant.

A

Kent

Wuhan

(furin-inhibited)

(fully cleaved)

>50% monomer

100% trimer

>70% monomer

B

NTD and NTDG shift

S + ACE2

(Dissociated monomer)

S + ACE2

(Trimer)

163
164

Figure 1: Structures of UK variant spike interacting with ACE2 receptor. (A) Surface representations of predominant molecular species present when ACE2 is mixed with the UK variant spike expressed in absence (left panel) or presence (middle panel) of furin inhibitor I, compared to the predominant species when ACE2 is mixed with Wuhan spike from our previous study (right panel) (PDB ID 7A91⁵). ACE2 is coloured in green, with monomers of the spike coloured in blue, goldenrod, and rosy brown in the middle panel and rosy brown for S monomers in the right and left panels. (B) Comparison of the domain movements of S complexed to ACE2 in its trimeric (middle) compared to monomeric (left) form with (right) the detail showing the location of the unstructured loop between the NTDG and S2, where the cleavage site is located. Domains are coloured: RBD in rosy brown, RBD-associated subdomain (ganymede, RBDG) in pink, NTD ganymede (NTDG) in light blue, NTD in navy, and S2 in red.

177

Figure 2: Variant spike binding to receptor ACE2. (A) K_d of variant spikes binding to ACE2 measured using biolayer interferometry and calculated from koff/kon analysis (see Fig. S7 for details). (B) Plots of fractional saturation binding measurements with data for the UK variant showed in red, UK variant with G614D substitution in pink, and Wuhan (D614) spike in blue. Wuhan (*the data for which shown here are adapted from our previous work²⁴) and UK D614 spike show almost identical affinity towards ACE2. Similar results were obtained for G614 vs D614 mink (Y453F) spike (**Fig. S4, S7**). (C) cryoEM density of the complex of SA variant S1 with ACE2, with ACE2 coloured in green, RBD in rosy brown, RBD subdomain in plum, with the remaining S1 disseminated density in cream. (D, E) Detail of changes in the binding interfaces present in variants (left column) compared to the Wuhan strain (right column). (D) The N501Y substitution present in both the SA and UK variants allows formation of a new hydrogen bond or a salt bridge. (E) The K417N substitution present in the SA variant eliminates a salt bridge between the RBD and ACE2.

A

South Africa

Wuhan

1 Erect RBD

2 Erect RBD

Open 100%

All RBD down

Closed 83%

B

Figure 3: South African variant spike trimer structures. (A) (Left and centre) Surface representations of cryoEM structures of the SA variant spike, with monomers coloured in blue, goldenrod and rosy brown, compared to (right) the closed form of the Wuhan strain determined in our previous study (PDB ID 6ZGE⁸) shown in paler hues of the same colours. The SA variant adopts a fully open conformation. (B) The K417N substitution present in the SA variant likely destabilises the closed form of the protein by introducing a steric clash and interfering with the network of electrostatic interactions present at the trimer interface in the closed form of the Wuhan protein. Notably, R417 in the fully-closed Pangolin-CoV spike seems to stabilise the same network of RBD-RBD interactions²⁴.

Methods

Construct design

The SARS-CoV-2 spike constructs used in this study were derived by Genscript from the spike ectodomain (residues 1-1208) constructs of the Wuhan variant cloned into pcDNA.3.1(+) described before^{5,8}. The variant spikes were stabilised in the pre-fusion conformation (K986P and V987P)²⁵ with the furin-cleavage site ([PH]RRAR) either intact ('2P' constructs) or mutated to the uncleavable sequence PSRAS ('FUR2P' constructs). The following variant spikes were made (SA¹⁹, UK¹⁸ and mink²⁶) (all mutations listed with reference to the NCBI sequence YP_009724390.1): 2P UK (Δ 69-70, Δ 144, N501Y, A570D, D614G, P681H, T716A, S982A, D1118H, and K986P, V987P), FUR2P UK (Δ 69-70, Δ 144, N501Y, A570D, D614G, T716A, S982A, D1118H, and R682S, R685S, K986P V987P), FUR2P D614 UK (Δ 69-70, Δ 144, N501Y, A570D, T716A, S982A, D1118H, and R682S, R685S, K986P, V987P), 2P SA spike (L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V, and K986P, V987P), FUR2P SA spike (L18F, D80A, D215G, Δ 242-244, R246I, K417N, E484K, N501Y, D614G, A701V, and R682S, R685S, K986P, V987P), FUR2P mink (Δ 69-70, Y453F, D614G, I692V, and R682S, R685S, K986P, V987P), FUR2P D614 mink (Δ 69-70, Y453F, I692V, and R682S, R685S, K986P, V987P). The Wuhan spikes and ACE2 ectodomain construct (residues 19-615) used in this study were exactly as described before^{5,8}.

Protein expression and purification

All the variant spikes were expressed in suspension-cultured Expi293F cells (Gibco) cells and purified as described before for the D614G spike¹². In brief, Expi293F cells were cultured at 37°C, shaking at 125 rpm, in humidified 8% CO₂ atmosphere in FreeStyle 293 Expression Medium and transfected with 1 mg of spike DNA per litre of culture at cell density of 3*10⁶/mL. For expression in presence of a furin inhibitor, decanoyl-Arg-Val-Lys-Arg-chloromethylketone (also known as "furin inhibitor I") was prepared as a 23.5 mg/mL stock solution in DMSO and added to the cells at final concentration of 100 μ M half an hour prior to transfection. Next day after transfection, the cells were enhanced according to the manufacturer's instructions and transferred to 32°C²⁷.

The supernatant was harvested on the fifth day post-transfection and the spike purified using cobalt NTA beads (TAKARA). The protein was then eluted in PBS with 200 mM imidazole, concentrated, and either flash-frozen or gel filtered at room temperature into a buffer containing 150 mM NaCl, TRIS pH 8 on a Superdex 200 Increase 10/300 GL column (GE Life Sciences). Wuhan spikes and ACE2 were made exactly as described before^{5,8}.

Biolayer interferometry

Measurements of spike variants affinity towards human ACE2 ectodomain were performed at 25 °C on an Octet Red 96 instrument (ForteBio) with shaking at 1000 rpm in 150 mM NaCl, 20 mM TRIS pH 8. First, variant spikes at 40-80 µg/mL were immobilised for 40-60 minutes on NiNTA sensors pre-equilibrated in buffer. Then, the ACE2 binding was measured using 2-5 minutes association and 10-20 minutes dissociation phases. At least three independent measurements were made for each spike.

The data were analysed using kinetic and equilibrium methods. For equilibrium analysis, the data were first normalised by dividing by the maximum observable response in order to give fractional saturation as a function of ACE2 concentration. The K_D was then determined from analysis of the variation of fractional saturation with ACE2 concentration. For kinetic analysis, plots of the observed rate (k_{obs}) were derived from association phases using a single exponential function and k_{on} and k_{off} were obtained from plots of k_{obs} vs ACE2 concentration as the slope and intercept respectively.

Cryo-EM sample preparation

Samples were frozen on R2/2 400 mesh Quantifoil grids glow-discharged for 30 s at 25 mA. A sample at 0.5-0.8 mg/mL final concentration of the spike was supplemented with 0.1% (final concentration) octyl glucoside, 4 µL of it applied on a grid, blotted for 5 s to 6 s with filter paper pre-equilibrated at 4 °C in 100% humidity, and plunge frozen in liquid ethane using Vitrobot Mark III. In order to obtain ACE2-spike complexes the proteins were mixed at 2 to 1 molar ratio of ACE2 to spike trimer and incubated at room temperature for 20-40 minutes prior to grid freezing.

Cryo-EM data collection

Data were collected using EPU software (Thermo Scientific) on Titan Krios microscopes operating at 300 kV either with a Falcon 3 camera (Thermo Scientific) operating in electron-counting mode (60 s exposures, total dose of $35 \text{ e}/\text{\AA}^2$, fractionated into 30 frames, $1.09 \text{ \AA}^2$ calibrated pixel size) or K2 camera (Gatan) with GIF Quantum LS energy filter (Gatan) with a slit width of 20 eV operating in the zero-loss mode (9.4 s exposures, $49 \text{ e}/\text{\AA}^2$ total dose, fractionated into 32 frames, $1.08 \text{ \AA}^2$ calibrated pixel size). The following eight datasets were collected (**Supp. Table 1**): FUR2P SA (Falcon 3), ACE2 + FUR2P SA (K2), FUR2P UK (Falcon 3), ACE2 + FUR2P D614 UK (K2), ACE2 + 2P UK furin-uninhibited (Falcon 3), ACE2 + 2P UK furin-inhibited (Falcon 3), FUR2P mink (K2), ACE2 + FUR2P D614 mink (K2). All micrographs were collected with defocus range between 1.5 and $3 \mu\text{m}$.

Cryo-EM data processing

Collected movies were motion corrected using MotionCor2²⁸ implemented in RELION²⁹ and contrast transfer functions estimated using CTFind4³⁰. Particles were picked using crYOLO³¹ using manually trained models. Particles were extracted 2x downsampled in RELION before two rounds 2D classification in cryoSPARC³². Classes which showed clear secondary structure were retained and an initial model generated also using cryoSPARC. Particles from the selected classes were 3D classified in RELION. The selection process for the different data collections are shown in **Figs. S3, S4, S6**. Final particle stacks were re-extracted unbinned and subjected to Bayesian polishing in RELION, and refined in cryoSPARC using either Homogeneous refinement or Non-Uniform Refinement routines, both coupled to per-particle defocus refinement. Final maps had their local resolution estimated using blocres³³ implemented in cryoSPARC (**Fig. S8**), followed by local resolution filtering in cryoSPARC and b-factor sharpening³⁴.

Model building, refinement, and validation

High resolution models of the monomeric S-ACE2 complexes were based on the previously determined model for the non-uniform map refinement of the monomeric Wuhan spike in complex with ACE2 (PDB 7A91)⁵. The low-resolution model of the

monomeric UK S-ACE2 complex was built from the non-uniform-refined, high resolution model of the same protein from this study combined with the monomeric Wuhan S-ACE2 (PDB 7A92)⁵. Models for the unbound trimeric spikes were based on our model of D614G spike (PDB IDs 7BNM, 7BNN, 7BNO)¹². The model of the three-ACE2-bound trimer of the 2P UK spike was based on the three-ACE2 bound Wuhan spike we determined before (PDB ID 7A98)⁵, in which the RBD (spike residues 333-527) and ACE2 were replaced by those from the monomeric, non-uniformed-refined UK S-ACE2 determined in this study. All structures were manually adjusted in COOT³⁵, refined with PHENIX Real Space Refine and validated in PHENIX³⁶. Measurements were performed with CCP4mg³⁷ and Chimera³⁸.

319 **Author Contributions**

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

322 **Conflict Statement**

323 We have no conflicts of interest to declare.

324 **Data Availability**

325 Maps and models have been deposited in the Electron Microscopy Data Bank,
326 <http://www.ebi.ac.uk/pdbe/emdb/> (Accession numbers XXX). Models have been deposited in
327 the Protein Data Bank, <https://www.ebi.ac.uk/pdbe/> (PDB ID codes XXX). [Accession numbers
328 will be available before publication].

329 **Acknowledgements**

330 We would like to acknowledge Andrea Nans of the Crick Structural Biology Science
331 Technology Platform for assistance with data collection, Phil Walker and Andrew Purkiss of
332 the Crick Structural Biology Science Technology Platform and the Crick Scientific Computing
333 Science Technology Platform for computational support. We thank Peter Cherepanov, George
334 Kassiotis, and Svend Kjaer for discussions. This work was funded by the Francis Crick Institute
335 which receives its core funding from Cancer Research UK (FC001078 and FC001143), the
336 UK Medical Research Council (FC001078 and FC001143), and the Wellcome Trust
337 (FC001078 and FC001143). The work done at the Crick Worldwide Influenza Centre, a WHO
338 Collaborating Centre for Reference and Research on Influenza, was supported by the Francis
339 Crick Institute receiving core funding from Cancer Research UK (FC001030), the Medical
340 Research Council (FC001030) and the Wellcome Trust (FC001030).

341

342 References

- 343
- 344 1. Walls, A. C. *et al.* Cryo-electron microscopy structure of a coronavirus spike
- 345 glycoprotein trimer. *Nature* (2016) doi:10.1038/nature16988.
- 346 2. Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike
- 347 glycoprotein promote membrane fusion. *PNAS* (2017) doi:10.1073/pnas.1708727114.
- 348 3. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike
- 349 Glycoprotein. (2020) doi:10.1016/j.cell.2020.02.058.
- 350 4. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion
- 351 conformation. *Science* (80-.). (2020) doi:10.1126/science.aax0902.
- 352 5. Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2
- 353 for membrane fusion. *Nature* (2020) doi:10.1038/s41586-020-2772-0.
- 354 6. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science*
- 355 (80-.). eabd4251 (2020) doi:10.1126/science.abd4251.
- 356 7. Ke, Z. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact
- 357 virions. *Nature* (2020) doi:10.1038/s41586-020-2665-2.
- 358 8. Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures
- 359 inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* 1–5 (2020)
- 360 doi:10.1038/s41594-020-0468-7.
- 361 9. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G
- 362 Increases Infectivity of the COVID-19 Virus. *Cell* (2020)
- 363 doi:10.1016/j.cell.2020.06.043.
- 364 10. Challen, R. *et al.* Risk of mortality in patients infected with SARS-CoV-2 variant of
- 365 concern 202012/1: Matched cohort study. *BMJ* **372**, (2021).
- 366 11. Davies, N. G. *et al.* Increased mortality in community-tested cases of SARS-CoV-2
- 367 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).
- 368 12. Benton, D. J. *et al.* The effect of the D614G substitution on the structure of the spike
- 369 glycoprotein of SARS-CoV-2. *PNAS* **118**, (2021).
- 370 13. Zhang, J. *et al.* Structural impact on SARS-CoV-2 spike protein by D614G
- 371 substitution. *Science* (80-.). **372**, eabf2303 (2021).
- 372 14. Gobeil, S. M. C. *et al.* D614G Mutation Alters SARS-CoV-2 Spike Conformation and
- 373 Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep.* **34**, 108630 (2021).
- 374 15. Peter Horby, Catherine Huntley, Nick Davies, John Edmunds, N. F. & Graham
- 375 Medley, C. S. NERVTAG - Presented to SAGE on 21/1/21.
- 376 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961037/NERVTAG_note_on_B.1.1.7_severity_for_SAGE_77__1_.pdf)
- 377 [ent_data/file/961037/NERVTAG_note_on_B.1.1.7_severity_for_SAGE_77__1_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961037/NERVTAG_note_on_B.1.1.7_severity_for_SAGE_77__1_.pdf).
- 378 16. Public Health England. *Investigation of novel SARS-CoV-2 variant: Variant of*
- 379 *Concern 202012/01*.
- 380 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf)
- 381 [ent_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf)
- 382 .
- 383 17. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage
- 384 B.1.1.7 in England. *Science* (80-.). **372**, eabg3055 (2021).
- 385 18. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2
- 386 lineage in the UK defined by a novel set of spike mutations.
- 387 [https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563)
- 388 [2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563).
- 389 19. Tegally, H. *et al.* Emergence of a SARS-CoV-2 variant of concern with mutations in
- 390 spike glycoprotein. *Nature* **592**, 438 (2021).
- 391 20. Tegally, H. *et al.* Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.*
- 392 **27**, 440–446 (2021).
- 393 21. Xiong, X. *et al.* A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat. Struct.*
- 394 *Mol. Biol.* (2020) doi:10.1038/s41594-020-0478-5.

22. Threat Assessment Brief: Emergence of SARS-CoV-2 B.1.617 variants in India and situation in the EU/EEA. <https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-sars-cov-2-b1617-variants>.
23. Nervtag. SARS-CoV-2 variants of concern and variants under investigation in England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984274/Variants_of_Concern_VOC_Technical_Briefing_10_England.pdf (2021).
24. Wrobel, A. G. *et al.* Structure and binding properties of Pangolin-CoV spike glycoprotein inform the evolution of SARS-CoV-2. *Nat. Comm.* **12**, 1–6 (2021).
25. Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *PNAS* **114**, E7348–E7357 (2017).
26. Ecdc. *Detection of new SARS-CoV-2 variants related to mink*. <https://www.ecdc.europa.eu/sites/default/files/documents/RRA-SARS-CoV-2-in-mink-12-nov-2020.pdf> (2020).
27. Esposito, D. *et al.* Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* **174**, 105686 (2020).
28. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
29. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
30. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
31. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, (2019).
32. 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).
33. 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.* (2013) doi:10.1016/j.jsb.2013.08.002.
34. 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).
35. Emsley, P., Lohkamp, B., Scott, W. G., Cowtan, K. & IUCr. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
36. 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).
37. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394 (2011).
38. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Table 1:

Cryo-EM data collection, refinement and validation statistics

	UK Monomer +ACE2 (EMDB- xxxx) (PDB xxxx)	UK Monomer +ACE2 (with accessories) (EMDB- xxxx) (PDB xxxx)	SA Monomer +ACE2 (EMDB- xxxx) (PDB xxxx)	Mink Monomer +ACE2 (EMDB- xxxx) (PDB xxxx)	UK Trimer Closed (EMDB- xxxx) (PDB xxxx)	UK Trimer 1 Erect RBD (EMDB- xxxx) (PDB xxxx)	UK Trimer 2 Erect RBD (EMDB- xxxx) (PDB xxxx)	SA Trimer 1 Erect RBD (EMDB- xxxx) (PDB xxxx)	SA Trimer 2 Erect RBD (EMDB- xxxx) (PDB xxxx)	Mink Trimer Closed (EMDB- xxxx) (PDB xxxx)	Mink Trimer 1 Erect RBD (EMDB- xxxx) (PDB xxxx)	Mink Trimer 2 Erect RBD (EMDB- xxxx) (PDB xxxx)	UK Trimer + 3 ACE2 (EMDB-xxxx) (PDB xxxx)
Data collection and processing													
Voltage (kV)	300	300	300	300	300	300	300	300	300	300	300	300	300
Electron exposure (e ⁻ /Å ²)	49	49	49	49	35	35	35	35	35	49	49	49	35
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	-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	-1.5 to -3.0	-1.5 to -3.0	-1.5 to -3.0
Pixel size (Å)	1.08	1.08	1.08	1.08	1.09	1.09	1.09	1.09	1.09	1.08	1.08	1.08	1.09
Symmetry imposed	C1	C1	C1	C1	C3	C1	C1	C1	C1	C3	C1	C1	C3
Final particle images (no.)	187k	187k	121k	393k	50k	316k	46k	196k	111k	84k	602k	106k	45k
Map resolution (Å) FSC threshold = 0.143	3.5	4.0	3.5	3.3	3.6	3.4	4.1	3.5	3.7	3.0	2.8	3.3	3.9
Refinement													
Initial model used (PDB code)	7A91	7A92	7A91	7A91	7BNM	7BNN	7BNO	7BNN	7BNO	7BNM	7BNN	7BNO	7A98
Model resolution (Å) FSC threshold = 0.5	3.5	4.0	3.7	3.3	3.6	3.4	4.3	3.6	3.9	3.2	2.9	3.5	4.0
Map sharpening <i>B</i> factor (Å ²)	-130.2	-108.6	-132.7	-134.2	-140.0	-151.5	-123.0	-141.7	-132.8	-98.7	-103.0	-88.4	-119.0
Model composition													
Non-hydrogen atoms	6891	9823	6870	6858	24303	23655	24109	24067	23641	25773	25562	25400	38670
Protein residues	839	1210	837	839	3003	2939	3000	3005	2981	3189	3193	3178	4758
Ligands	9	8	8	7	57	50	43	40	22	60	42	38	60
R.m.s. deviations													
Bond lengths (Å)	0.005	0.002	0.003	0.003	0.003	0.005	0.004	0.004	0.003	0.005	0.004	0.003	0.003
Bond angles (°)	0.621	0.457	0.529	0.529	0.547	0.607	0.526	0.571	0.504	0.673	0.651	0.556	0.511
Validation													
MolProbity score	1.27	1.53	1.42	1.33	1.67	1.78	1.97	1.60	1.46	1.32	1.34	1.21	1.76
Clashscore	3.78	6.34	4.17	4.03	6.22	5.29	10.95	4.65	5.15	2.68	3.31	3.99	8.24
Poor rotamers (%)	0.00	0.00	0.54	0.00	0.00	0.04	0.00	0.19	0.50	0.32	0.54	0.18	0.00
Ramachandran plot													
Favored (%)	97.47	96.96	96.62	97.23	95.32	91.81	93.85	94.80	96.88	96.07	96.64	97.90	95.50
Allowed (%)	2.53	3.04	3.38	2.77	4.68	8.19	6.15	5.20	3.12	3.93	3.36	2.10	4.50
Disallowed (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure S1: Location of changes in spike protein of different SARS-CoV-2 variants investigated in this study. Changes are highlighted on our previous structure of SARS-CoV-2 spike in closed conformation (PDB ID 6ZGE⁸) (left). Substitutions of particular interest—located either on the intra-trimer interfaces or the RBD—are highlighted in red. Changes in the NTD—mainly restricted to surface interfaces—are shown in either dark (substitutions) or light (deletions) green. Other changes are highlighted in blue. A full list of changes is shown for each variant on the right.

Figure S2: Cleavage state of variant spikes. (A, B) SDS PAGE analysis of purified spikes used in this study. (A) From left, lanes contain: molecular weight markers (Precision Blue Protein Standards, All Blue, Bio-rad), uncleavable ('FUR2P': R682S+R685S) Wuhan spike, and 2P-only spikes: Wuhan, SA, UK, P681H-only (otherwise identical to the Wuhan 2P). UK and P681H-only spikes are almost fully cleaved into S1 and S2 while SA and Wuhan spikes are cleaved only partially; this indicates that the P681H substitution in the cleavage site is sufficient to cause the extent of cleavage observed in the UK variant. (B) Cleavage state of P681H-only (left two lanes) and UK (right two lanes) spikes expressed in presence of furin inhibitor I (lanes "I") compared to non-inhibited (n-I). The inhibitor almost fully prevents spike cleavage into S1 and S2. (C) Sequence alignment of the S1/S2 cleavage site from several SARS-CoV-2 strains and related sarbecoviruses: SARS-CoV and the bat-CoV most closely related to SARS-CoV-2, RaTG13. SARS-CoV-2 acquired furin-cleavage site RxxR during its evolution and this site has become more polybasic (blue) in the recent UK and Indian variants.

UK spike, furin uninhibited (cleaved)

UK spike, furin inhibited (uncleaved)

Figure S3: cryoEM image processing scheme for UK variant in furin uninhibited (cleaved) and furin inhibited (uncleaved) forms binding to ACE2.

ACE2 + UK G614D

ACE2 + SA

ACE2 + Mink G614D

Figure S4: cryoEM image processing scheme for UK (G614D), SA and Mink (G614D) variants in complex with ACE2.

A

Mink spike trimer

B

Mink monomer S/ACE2 complex

C

D

UK spike trimer

E

F

Figure S5: Structural insights into mink and UK variant spikes. (A) Surface representations of mink spike structures determined in this study: in a dilated, closed conformation different to that of the Wuhan spike but resembling D614G spike¹²; and one- and two-RBD erect conformations, the latter of which is assumed by D614G but not by Wuhan spike¹². (B) Cryo-EM density of monomeric complex of D614 mink spike and ACE2. (C) Detail of the mink spike / ACE2 interface (left) compared to Wuhan (right). The His-34, which adopts two alternative conformations in Wuhan spike, is present only as one rotamer in the structure of the mink spike/ACE2 complex (Fig. S4). (D) Surface representations of furin-uncleavable UK spike in three conformations which resemble those of the mink and D614G-only spike we described previously¹²: a dilated-closed conformation different to that of the Wuhan variant and two open conformations, with one or two RBDs erect. (E, F) Detail of substitutions in UK spike (left panels) compared to Wuhan (right panels) that may further contribute to enhanced trimeric state of the open and receptor-bound forms of UK spike. (E) A570 in Wuhan strain lies close to the interface between RBD-associated subdomain (RBDG, blue) and the fragment of S2 core of the neighbouring chain (residues 815-855, red), which undergoes significant rearrangement upon spike opening (the unfolded region 824-853 is structured in the closed Wuhan spike). In (D570) UK spike the whole RBDG domain shifts slightly closer towards S2 and the same region of the neighbouring-chain S2, especially F855, adopts different conformation upon spike opening. This alternative conformation might be attributed to a formation of a salt bridge between K854, which incidentally makes a salt bridge with D614 in the original Wuhan strain but not in the later G614 variants, and D570 in the UK variant spike. (F) The S2 residue 1118 lies at the membrane proximal end of the UK spike trimer, close to the trimer axis. It is an aspartic acid in Wuhan (right) but a histidine in UK. It appears that a cluster of trimer-related aspartic acid side chains at this position forms a less stable arrangement than is adopted by a corresponding cluster of neutral histidine residues in the UK spike, which also coordinate a smaller molecule or an ion-here modelled as water. These two substitutions could explain greater stability of the open form of UK spike. Indeed, a comparison of the overall surface areas of the monomer-monomer interfaces in the 1-RBD-erect conformations of the UK and G614-only trimers shows that the former is 700 Å² larger providing more evidence to why the UK spike is less likely to undergo disassembly upon receptor binding than other variants.

Figure S6: cryoEM image processing scheme for spike trimers of UK, SA and Mink variants.

A

Spike	k_{on} ($\mu\text{M}^{-1}\text{s}^{-1}$)	k_{off} (s^{-1})	K_d (Kin) (nM)	K_d (Amp) (nM)
Wuhan *	0.119 ± 0.005	0.0107 ± 0.0015	90.7 ± 14.3	114.2 ± 17.5
UK G614D	0.130 ± 0.006	0.0124 ± 0.0017	93.0 ± 14.0	111.4 ± 21.9
Mink G614D	0.103 ± 0.004	0.0112 ± 0.0015	113.0 ± 15.3	90.5 ± 22.4
Wuhan D614G	0.075 ± 0.003	0.0108 ± 0.0014	144.0 ± 19.0	154.2 ± 29.7
UK	0.138 ± 0.003	0.0023 ± 0.0012	16.7 ± 7.6	25.3 ± 4.2
Mink	0.088 ± 0.006	0.0033 ± 0.0014	36.6 ± 10.5	46.2 ± 5.8
South Africa	0.180 ± 0.004	0.0075 ± 0.0019	44.7 ± 10.5	31.2 ± 4.3

B

C

Figure S7: Biolayer interferometry binding measurements of ACE2 binding to immobilised variant spikes. (A) Thermodynamic parameters for ACE2 binding to different spikes. Association rate constants (k_{on}) were determined from the slopes of plots of the observed rate constant against ACE2 concentration shown in panel C. Dissociation rate constants (k_{off}) were determined from the intercepts of these plots and through independent analysis of the dissociation phase. The K_d values were calculated from the kinetic data as k_{off}/k_{on} (K_d (Kin)) and from analysis of the dependence of fractional saturation on ACE2 concentration shown in main Fig. 2b and panel B below (K_d (Amp)). (B) Variation of fractional saturation with ACE2 concentration for different spikes. The solid lines are the computed best fits. * The data for Wuhan (D614) shown here are adapted from our previous work (Pangolin ref). (C) Dependence of the observed rate constant (k_{obs}) on ACE2 concentration for different spikes.

Figure S8: Fourier Shell Correlation (FSC) curves for each of the deposited structures.