bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Structure and mechanism of SARS-CoV-2 Spike N679-V687 deletion variant elucidate cell-type specific evolution of viral fitness.

Kapil Gupta[1,2†], Christine Toelzer[1,2†], Maia Kavanagh Williamson[3†], Deborah K. Shoemark[1,2†], A. Sofia F. Oliveira[1,3], David A. Matthews[4], Abdulaziz Almuqrin[4], Oskar Staufer[5,6,7,8], Sathish K.N. Yadav[1,2], Ufuk Borucu[1,2], Frederic Garzoni[9], Daniel Fitzgerald[10], Joachim Spatz[5,6,7,8], Adrian J. Mulholland[3], Andrew D. Davidson[4], Christiane Schaffitzel[1,2,10*], Imre Berger[1,2,3,8,10*]

[1] School of Biochemistry, University of Bristol, 1 Tankard's Close, Bristol BS8 1TD, UK

[2] Bristol Synthetic Biology Centre BrisSynBio, 24 Tyndall Ave, Bristol BS8 1TQ, UK.

[3] School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

[4] School of Cellular and Molecular Medicine, University of Bristol, University Walk, Bristol, BS8 1TD, UK.

[5] Department for Cellular Biophysics, Max Planck Institute for Medical Research, Jahnstraße 29, 69120 Heidelberg, Germany.

[6] Institute for Physical Chemistry, Department for Biophysical Chemistry, University of Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany.

[7] Max Planck School Matter to Life, Jahnstraße 29, D-69120 Heidelberg, Germany.

[8] Max Planck Bristol Centre for Minimal Biology, Cantock's Close, Bristol BS8 1TS, UK.

[9] Imophoron Ltd, St. Philips Central, Albert Rd, St. Philips, Bristol, BS2 0XJ, UK.

[10] Halo Therapeutics Ltd, St. Philips Central, Albert Rd, St. Philips, Bristol, BS2 0XJ, UK.

† These authors contributed equally to this study.

* Corresponding authors: Christiane Schaffitzel, cb14941@bristol.ac.uk

Imre Berger, imre.berger@bristol.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

As the global burden of SARS-CoV-2 infections escalates, so does the evolution of viral variants which is of particular concern due to their potential for increased transmissibility and pathology. In addition to this entrenched variant diversity in circulation, RNA viruses can also display genetic diversity within single infected hosts with co-existing viral variants evolving differently in distinct cell types. The BriSΔ variant, originally identified as a viral subpopulation by passaging SARS-CoV-2 isolate hCoV-19/England/02/2020, comprises in the spike glycoprotein an eight amino-acid deletion encompassing the furin recognition motif and S1/S2 cleavage site. Here, we analyzed the structure, function and molecular dynamics of this variant spike, providing mechanistic insight into how the deletion correlates to viral cell tropism, ACE2 receptor binding and infectivity, allowing the virus to probe diverse trajectories in distinct cell types to evolve viral fitness.

Teaser

SARS-CoV-2 can exploit different cell types to diversify and evolve virus variants distinct in infectivity and structure.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAIN TEXT

Introduction

SARS-CoV-2 spike (S) glycoprotein prominently differs from other betacoronavirus S proteins in the insertion of a furin cleavage site in the S1/S2 junction site (*1*). The S trimer glycoprotein is responsible for binding to the ACE2 receptor and for viral cell entry after cleavage at the S1/S2 junction and S2' sites (*2*). Critical to this process is proteolytic processing of S by host cell proteases (*3*). After intracellular cleavage at the S1/S2 junction by a furin-like protease to produce the S1 and S2 subunits, S is further primed by cleavage at the S2´ site by host serine proteases on the plasma membrane such as TMPRSS2 (*4,5*) or the endosomal cysteine proteases cathepsin B/L (*6*). S1 comprises the N-terminal domain (NTD), the receptor binding domain (RBD), and the SD1 and SD2 domains (*7,8*). S2 contains the S2´ cleavage site, the fusion peptide, a fusion peptide proximal region (FPPR), a HR1 heptad repeat, a central helix and a connector domain followed by a HR2 heptad repeat, the transmembrane domain and the C-terminal cytoplasmic domain (*7,8*). Receptor binding destabilizes S, allowing S2´ cleavage, leading to shedding of S1 while S2 reorganizes to mediate fusion of viral and cellular membranes, enabling entry of SARS-CoV-2 into the host cells (*9*). The furin cleavage site is a four amino acid motif located on a solvent-exposed flexible loop of S (*7*). Furin-cleaved S was shown to bind more efficiently to human ACE2 than uncleaved S (*10*). The furin cleavage site thus contributes substantially to the high infectivity of SARS-CoV-2, adding to the lethality of the virus.

After growth of a low passage isolate of SARS-CoV-2 from February 2020 in the African Green monkey kidney cell line Vero E6, a cell line routinely used to propagate viruses from clinical isolates, we discovered a virus subpopulation with a S variant (termed here BriSΔ) exhibiting an in-frame 8 amino acid deletion encompassing the furin recognition motif and S1/S2 cleavage site (amino acids 679-687 NSPRRARSV, replaced by I) (*11*). Subsequently, further deletion variants abrogating S1/S2 cleavage were identified after viral passaging in cell culture (*12-14*) and at low frequency in clinical samples, attenuating infection in animal models (*15-18*). Moreover, deleting only PRRA in S by reverse genetics resulted in a recombinant ΔPRRA SARS-CoV-2 which exhibited increased infectivity and viral titer in Vero E6 cells, but a 10-fold reduced viral titer in Calu-3 2B4 lung epithelial carcinoma cells compared to the wildtype (WT) virus (*19*), indicating the acquisition of a furin cleavage site increased SARS-CoV-2 fitness for replication in respiratory cells. BriSΔ, compared to WT virus, shows higher infectivity of Vero E6 cells which have low levels of TMPRSS2 protease and are deficient in the interferon pathway, suggesting that in the heterogenous environment of the human body such deletion variants can emerge in suitable cell types serving as potential niches for SARS-CoV-2 to further evolve or specialize.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, we dissect the structure, dynamics and mechanism of the BriSΔ deletion variant S we identified, to gain insight into how elimination of a loop-region comprising the furin recognition motif and S1/S2 cleavage site impacts viral cell tropism, infectivity and receptor binding, revealing molecular communication between functional regions within the spike glycoprotein.

Results

BriSΔ variant and wildtype SARS-CoV-2 clonal isolation

Direct RNA sequence analysis of a virus stock of SARS-CoV-2 isolate hCoV-19/England/02/2020, produced by a single passage in Vero E6 cells, revealed the presence of the WT SARS-CoV-2 and the BriSΔ variant (Fig. 1A). To obtain homogenous virus populations, the mixed virus stock was subjected to two rounds of limiting dilution in Vero E6 and human Caco-2 cells (fig. S1). Further nanopore direct RNA sequencing confirmed that the limiting dilution yielded WT SARS-CoV-2 from Caco-2 cells. In contrast, BriSΔ was selected for in Vero E6 cells (Fig. 1A) as expected (*11*). The amounts of the BriSΔ and WT viruses was estimated based on genome equivalence by qRT-PCR and used for infection of Vero E6, Vero E6/TMPRSS2, Caco-2, Caco-2-ACE2 and Calu-3 cells (Fig. 1B-F) using a range of virus dilutions, starting at a multiplicity of infection of 10 for the WT virus based on the Vero E6 cell titer. The percentage of virus infected cells was analyzed 18 hours after virus infection, before multiple rounds of virus replication. In Vero E6 cells, half-maximal infection was achieved with an ~ 6-fold higher dilution of BriSΔ as compared to WT virus (Fig. 1B). Overexpression of TMPRSS2 protease in Vero E6 cells (*5*) resulted in a substantially higher infection efficiency for both viruses; close to 100% of cells were infected with an up to 16-fold dilution of WT virus and an up to 64-fold dilution of BriSΔ virus (Fig. 1C). Thus, the lack of the TMPRSS2 protease contributes to, but is not the only reason why the BriSΔ variant infects Vero E6 cells better than WT virus. Differences in the route of cell entry either via fusion at the plasma membrane or receptor mediated endocytosis (*20-22*) may account for this result. Interestingly our results using Vero E6/TMPRSS2 cells are similar to those of Zhu et al (*18*) comparing the replication of WT SARS-CoV-2 and a virus (Sdel) containing a 7 amino acid deletion encompassing the furin cleavage site but differ from those based on a competition assay between the WT and ΔPRRA viruses, which infected Vero E6/TMPRSS2 cells equally well (*19*). Despite the fact that we selected WT SARS-CoV-2 from Caco-2 cells by serial dilution, BriSΔ infected Caco-2 cells better than the WT at high virus titers; 25% versus 10% infected cells were observed with the starting

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dilutions of BriSΔ and WT, respectively (Fig. 1D). Overexpression of the ACE2 receptor in Caco-2 cells led to 70% infection of cells up to 32-fold dilution of WT, whereas only about 35% of cells could be infected by the BriSΔ variant at the same dilution (Fig. 1E). Thus, both the WT and BriSΔ viruses infect Caco-2 cells better when ACE-2 expression was increased, but the improvement for WT virus was substantially higher. Calu-3 lung cells were infected about 2-fold better by WT virus for all dilutions except the starting dilutions (Fig. 1F), corroborating the contribution of the furin site in SARS-CoV-2 S to improved infection of lung cells (*19*).

Next, we tested neutralization of the WT and BriSΔ viruses. No difference was found in neutralization of the two viruses when Vero E6/TMPRSS2 and Vero E6 cells were infected with equal amounts of infectious virus based on cell infectivity, in the presence of a commercial antibody binding the RBD (Fig. 1G) or human serum from a convalescent COVID-19 patient (fig. S2), respectively, indicating that both virus species were neutralized with equal potency by the antibodies tested.

Cryo-EM structure of BriSΔ

To understand the structural impact of the deletion of the furin cleavage site on SARS-CoV-2 S architecture, we produced the ectodomain of BriSΔ by MultiBac/insect cell expression as a secreted trimer (*23*). We purified the protein by affinity purification and size exclusion chromatography (fig. S3A,B) We used the peak fraction from SEC for negative stain EM quality control (fig. S3C) and cryogenic electron microscopy (cryo-EM) (fig. S4, table S1). We determined the BriSΔ structure without applying symmetry (C1) at 3.0 Å resolution (fig. S5A). In our analysis, all BriSΔ particles exhibited the locked conformation of the S trimer we had described previously (*23*). After applying 3-fold symmetry (C3) we obtained a 2.8 Å cryo-EM map (Fig. 2, fig. S5B). In this compact locked S conformation, the receptor-binding motif (RBM) is buried inside the RBD trimer obstructing ACE2 receptor binding (fig. S6A). The RBD also comprises an RGD motif implicated in integrin-mediated cell entry (*24*). The three RGD motifs present in BriSΔ are buried at the interfaces between the RBDs (fig. S6B) and thus will be likewise less accessible in the locked form, similar to the RBMs. We previously found a free-fatty acid (FFA) binding pocket in the locked structure of SARS-CoV-2 S, and identified a small molecule tightly bound in the pocket, with the molecular mass of linoleic acid (LA) as determined by electron-spray ionization mass-spectroscopy (ESI-MS) (*23*). In the present locked structure of BriSΔ, all three pockets are again occupied by a small molecule (Fig. 2A,B). Using hydrophilic interaction liquid chromatography followed by tandem mass spectrometry (HILIC-MS-MS) as an orthogonal method and LA as a calibration standard, we confirmed that the BriSΔ S samples

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contain LA (Fig. 2C). LA is bound in a bi-partite binding pocket where one RBD provides a hydrophobic pocket to accommodate the hydrocarbon tail, while residues R408 and Q409 of the adjacent RBD provide a polar lid coordinating the carboxy head group of LA (Fig. 2B). In the BriSΔ C1 structure, we identified the same tube-shaped density in all three RBD domains (fig. S6A), indicating high occupancy of all three pockets. Using masked 3D classification, we scrutinized the data set for potential heterogeneity in LA binding and found that at least 95% of the RBDs were LA-bound in our structure (fig. S7).

We compared our BriSΔ structure with previously solved S structures revealing additional features (Fig. 2). Disulfide bonds play a crucial role in stabilizing the S trimer and individual domains. Five out of 14 annotated disulfide bonds in S (*25*) stabilize the RBD including the disulfide bond linking C336 and C361 (Fig. 2D,E). Three arginine R1031 residues in our construct (R1039 in WT S), one each from the three polypeptide chains in the S trimer, form a hydrogen bond cluster (Fig. 2G). In this cluster, the arginine residues are symmetrically arranged around the central trimer axis with short range contacts of 4.65 Å present between the carbon atoms of the guanidino groups. The guanidinium planes stack parallelly on top of the aromatic plane of the juxtaposed F1034, and a salt bridge is formed to E1023 of the adjacent S polypeptide chain (Fig. 2G). R1031, E1023 and F1034 are conserved in all human coronaviruses, highlighting their central importance. In the vicinity, a disulfide bond is formed by conserved residues C1024 and C1035, arranging E1023 and F1034 at the required distance and in the proper conformation to stabilize the R1031-mediated interaction (not shown). Opening of the RBDs was shown to induce an asymmetry in the trimer structure that breaks this H-bond cluster (*8*,*26*). LA binding in the FFA-pockets induces conformational changes to the residues surrounding the FFA-binding pocket in the RBD and beyond, including the NTD, SD2 and the fusion peptide proximal region (FFPR). Re-organization of SD2 in the locked structure results in a stabilization of the region around R634 (Fig. 2K). This arginine residue is stabilized by π-stacking on Y829 (Y837) and thus connects to the FPPR of the neighboring subunit. Such π-stacking interactions were observed in a previous structure which had an intact furin site, and also (unassigned) density in the FFA-binding pocket (*10*). R634 stacking to Y829 is additionally stabilized by a hydrophobic interaction to L833 (Fig. 2K). Fixed in a rigid position through these interactions, C832 (C840) can form a disulfide bond to C843 (C851) which additionally stabilizes the FPPR (Fig. 2F). This disulfide bond-mediated stabilization of the FPPR has been described in a cryo-EM structure of full-length S protein (*9*). We scrutinized the deposited EM density of this structure and found it contained unassigned density in the FFA-binding pocket, indicating a ligand, presumably a free-fatty acid (*9*). In the meantime, further structures became available which adopt a locked conformation and comprise assigned or unassigned

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

density in the FFA-binding pocket. The sandwiched R634 is present in all; fully modelled in five structures (PDBID 6ZB5, 7JJI, 6ZGI, 6ZGE, 6XR8), not modelled in one structure but evident in the density (PDBID 6ZP2) and in a further structure with the density (but not the model) supporting the sandwiched arginine (PDBID 7DF3). In comparison, S structures in the closed, but not locked, conformation, where the FFA-binding pocket does not contain any ligand density, show no π-stacking interaction of R634. Instead, in these structures (PDBID 6VXX and 6ZP1) residues 620 to 640 as well as parts of the FPPR are disordered, and residues Y636 and R634 adopt different conformations.

Our BriSΔ construct harbors WT residues K978 and V979 (K986 and V987 in WT S), which often are mutated to prolines to stabilize S in the prefusion state. A closeup view of this region in the BriSΔ shows that the valine fits well into the density while the lysine sidechain appears to be somewhat flexible (Fig. 2H). Importantly, BriSΔ lacks 8 amino acids including the furin cleavage site and the S1/S2 cleavage site located on a flexible loop. This loop is now shorter due to the deletion and thus more rigid in BriSΔ, as evidenced by density in the C1 map which allowed to build a poly-alanine chain (Fig. 2I).

We analyzed N-glycosylation of BriSΔ and identified no significant changes compared to previous S structures (table S2). Interestingly, WT residues S673, T676, T678 and S680 close to the furin site are all candidates for O-glycosylation which is dependent on proline P681 (*27*). It was shown that O-glycosylation of these residues negatively affects furin cleavage, contributing to S stability and infectivity (*27*). P681 and S680 are deleted in BriSΔ, and P681 is mutated in other lineages, including the B.1.1.7 variant that emerged in Kent, UK (*28*) and rapidly spread globally. The enzymes responsible for O-glycosylation (GALNTs) are expressed differently in different cell types (*29*), and the absence of P681 in BriSΔ could thus contribute to the observed dominance of this variant in certain cell types.

Functional analysis of BriSΔ

The cryo-EM structure of BriSΔ showed only particles in the locked conformation in which RBM binding to ACE2 is obstructed (Fig. 2; figs S4-S6). In the cell-based assays, however, BriSΔ virus is infectious (Fig. 1). To address this apparent discrepancy, we biochemically analyzed the interaction of a range of S proteins with the ACE2 receptor. We compared BriSΔ binding to ACE2 with S protein lacking the RBM, S protein where the furin site residues are replaced by an alanine ($S^{RRAR->A}$), uncleaved WT S (S^{RRAR}) and furin-cleaved WT S (S^{RRAR*}) (Fig. 3A, table S3). ELISA using a surrogate virus neutralization test kit (sVNT) indicated that all S proteins efficiently bind ACE2, except the S

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein lacking the RBM used as a control (Fig. 3B). For the other S proteins, half-maximal binding was observed between 64 and 128 nM S protein in the assay. We determined the dissociation constant (K_D) of BriSΔ and ACE2 by surface plasmon resonance (SPR) with biotinylated ACE2 immobilized on a streptavidin-coated chip (Fig. 3C, fig. S8). Binding of BriSΔ (K_D=2.5 nM) to ACE2 is only slightly weaker as compared to $S^{RRAR->A}$ (1.4 nM) (*23*). In agreement, the maximal RU values, indicating mass deposited on the ACE2-coated chip, were virtually identical for BriSΔ, $S^{RRAR->A}$ and uncleaved WT S (Fig. 2D, fig. S8A,B). The partially cleaved WT S yielded lower RU_{max} signals, possibly due to partial dissociation of the S trimer and binding of the smaller S1 fragment to ACE2. Our biochemical assays thus establish that the BriSΔ trimer can adopt an open, ACE2-binding competent conformation, which however was not observed on the cryo-EM grids. This finding is in agreement with our results that the BriSΔ variant and WT virus can be neutralized efficiently with similar amounts of a commercial monoclonal antibody recognizing the RBD (Fig. 1G), and with human convalescent sera (fig. S2).

Targeted molecular dynamics simulations indicate that LA stabilizes locked BriSΔ more than wildtype S

We next characterized the conformational changes in WT S and BriSΔ by using targeted molecular dynamics simulations (Fig. 4) to explore the force required to open a single RBD from a closed position in the S trimer for apo and LA-bound systems corresponding to WT and BriSΔ. We performed 50 simulations over 10 ns for each system (Fig. 4A,B,C,D) applying a force constant of 0.2 kJ/mol/nm to raise a single RBD from an equilibrated, closed conformation to a target, open state, corresponding to the energy-minimized open S model built on EMD-1146 (*23*). Starting from their closed conformation, we observed that all BriSΔ and almost all apo WT S opened during the 10 ns MD simulations. In contrast, bound LA in the WT S substantially reduces the number of opening events (the cross-over time-point when the RBD opens more than halfway) compared to the apo WT spike (Fig. 4A,B). Notably, binding of LA to BriSΔ (Fig. 4D) completely prevents RBD opening at 0.2 kJ/mol/nm in these simulations. Increasing the force constant to 0.3 kJ/mol/nm was required to raise the RBD of LA-bound BriSΔ over this time period (Fig. 4E). Example plots illustrate how during the course of the simulation the force applied either resulted in the RBD opening (Fig. 4F, movie S1) or failed to pull the RBD open (no cross-over) as determined by root-mean-square deviation (RMSD) (Fig. 4G) The time at which the RMSD traces cross (cross-over time) occurred was binned (in ns steps) for all 50 replicates for each system, as shown in Fig. 4A-E. In summary, 98% simulations of the apo WT S and 100% of BriSΔ open the RBD with a force constant of 0.2 kJ/mol/nm applied, while only 14% of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

WT and 0% of the BriSΔ systems complexed with LA open with the same force applied (Fig. 4H). Thus, LA binding significantly stabilizes the locked conformation of both, WT S (P=0.0027) and BriSΔ (P=0.0024). Importantly, LA apparently shifts the equilibrium favoring the locked conformation for BriSΔ more than for WT S (P = 0.039) (Fig. 4H). This hypothesis is supported by our cryo-EM structure which shows that virtually all BriSΔ trimers were LA-bound and in the locked conformation (Fig. 2).

Communication between the furin cleavage site, the FFA-binding pocket and the FPPR

We next performed 180 short dynamical-nonequilibrium simulations to analyze intra-spike communication and identify any structural networks connecting the FFA-binding pocket to functionally important motifs within WT S and BriSΔ (Fig. 5, fig. S9). In these simulations, the LA molecules were removed from the FFA pockets and the response of the systems to this perturbation was determined using the Kubo-Onsager approach (*30-32*). The simulations revealed a cascade of structural changes occurring in response to the LA removal and highlight the route by which such changes are transmitted through the S protein (fig. S10, S11). Strikingly, we find that the furin cleavage site responds swiftly to LA in the WT S, despite being 40 Å away from the FFA pocket (Fig. 5). The nonequilibrium simulations also show allosteric connection between the FFA pocket and V622-L629, D808-S813 and the fusion peptide proximal region (FPPR, residues 833-855) in WT S (Fig. 5, fig. S10). Similar structural responses are observed for the three WT S subunits (fig. S10). The V622-L629 region is part of a larger loop structure spanning residues 617-641, close to the R634-Y829 cation-π interaction site (Fig. 2K) (*33*). The V622-L629 region, close to R634, and the furin-cleavage site respond rapidly to the perturbation in WT S: 0.1 ns after LA removal, conformational rearrangements are observed in both regions (Fig. 5, fig. S10). As the simulations proceed, a gradual increase in deviations is observed for both the furin-cleavage and V622-L629 regions in WT S, with the conformational changes being further propagated to the segments adjacent to the fusion peptide, namely the FPPR and D808-S813. These results show direct coupling between the FFA-binding pocket and important, distant regions of the protein, including the furin-cleavage site. Our simulations also highlight differences in the dynamic response and in the effect of LA between the WT and BriSΔ spikes (Fig. 5, fig. S10, S11): in BriSΔ, only the V622-L629 region responds rapidly to LA removal. The deletion site itself shows smaller conformational rearrangements in BriSΔ when compared to WT S, confirming a more rigid arrangement of the shortened loop which in WT S comprises the furin cleavage and S1/S2 cleavage sites. These results indicate that that BriSΔ has different allosteric and dynamical behaviour from the WT S.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

We investigated here the structure and functional characteristics of BriSΔ. We find important similarities between BriSΔ, an engineered ΔPRRA SARS-CoV-2 in which only the four residues of the furin cleavage site had been removed by reverse genetics (*19*) and Sdel, a variant containing a cell passage acquired deletion encompassing the furin cleavage site (*18*). All deletion variants replicate substantially better in Vero E6 cells but show impaired replication in Calu-3 cells - underscoring that WT SARS-CoV-2 evolved to efficiently infect and replicate in respiratory cells. At the same time, we find differences highlighting the impact of the exact sequence context of the deletion: expression of serine protease TMPRSS2 in Vero E6 cells reduced the replication advantage of ΔPRRA SARS-CoV-2 (*19*), but not of BriSΔ or Sdel (*18*). In our experiments, BriSΔ and WT SARS-CoV-2 both exhibited increased infection and replication in Vero E6/TMPRSS2 cells, but BriSΔ retained higher infection and replication rates as compared to WT virus (Fig. 1C). Importantly, the engineered ΔPRRA virus required more antibody for neutralization of the virus compared to WT SARS-CoV-2, while monoclonal antibody and human sera neutralize BriSΔ and WT SARS-CoV-2 with equal efficacy (Fig. 1G, fig. S2). Interestingly, ACE2 receptor overexpression in Caco-2 cells improved the infection and replication of BriSΔ and WT SARS-CoV-2 as compared to Caco-2 cells with a lower constitutive expression of ACE2 (Fig. 1D,E).

Mechanistically, the different infectivity of cells expressing high levels of TMPRSS2 is probably explained by the fact that WT SARS-CoV-2 S is cleaved at the furin cleavage site (S1/S2 cleavage), which primes S for cleavage with TMPRSS2 (at S2´) resulting in activation of the fusion peptide (*1*). The virus can then fuse at the plasma membrane to enter cells (*20, 22*). It has been shown that the furin-cleaved S can also interact with neuropilin (NRP1) following the CendR rule, potentially enhancing viral entry (*34,35*). Lacking these modalities, BriSΔ is likely taken up by receptor-mediated endocytosis and then potentially cleaved by an intracellular protease before fusing via the endosome. Thus, Calu-3 cells which express TMPRSS2 are better infected by WT virus, and they replicate WT virus more efficiently than BriSΔ virus. In addition, intracellular defence pathways exist such as the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interferon-induced transmembrane proteins (IFITMs) that may be more active against virus entering through the endosome rather than fusing at the plasma membrane (*36*). Vero E6 cells express low very levels of TMPRSS2, therefore WT virus cannot enter efficiently by fusing at the plasma membrane, and it has recently been suggested that WT SARS-CoV-2 is also taken up by receptor-mediated endocytosis in these cells (*21*). Our data suggests this process is more efficient for the BriSΔ virus. Notably, Vero E6 cells are deficient in the interferon pathway, and entry via the endosome may not be inhibited in these cells (*17*). We hypothesize that deletion of the furin cleavage site potentially results in a more stable viral particle secreted from cells, with increased expression of full-length BriSΔ in a prefusion state. At the same time, we cannot exclude further impact of the BriSΔ variant on intracellular replication and immunogenicity.

The cryo-EM structure revealed BriSΔ trimer in an LA-bound, locked conformation (Fig. 2) which obstructs both ACE2 binding and integrin-mediated cell entry. We observe two motifs stabilizing LA-bound locked BriSΔ which appear to be characteristic for this S conformation (PDBID 6ZB5, 7JJI, 6ZGI, 6ZGE, 6XR8, 6ZP2, 7DF3), namely the H-bond cluster formed by R1031 stabilizing the trimer interface (Fig. 2G) and the cation-π interaction between R634 and Y829 stabilizing the interaction of S1 with the S2 of the adjacent trimer subunit (Fig. 2K). Y829 is part of the FPPR motif which is folded in the locked, but not in the closed conformation of S (*33*). R634 is part of an extended loop structure that holds the FPPR in place, and alterations in this region, for instance the D614G S mutation, favor opening of the RBDs and disorder the FPPR (*37*), possibly exposing the S2´ cleavage site upstream of the fusion peptide (*9*).

Biochemical interaction assays show that BriSΔ protein binds ACE2 with comparable affinity as S protein that does not have the deletion (Fig. 3). Thus, BriSΔ can evidently adopt open conformation(s), consistent with the infectivity of the BriSΔ variant (Fig. 1). We conclude that cryo-EM analysis may not fully reflect the dynamics of the S trimer which is subject to a range of stabilizing and destabilizing modalities (*38*). Moreover, WT and other S proteins previously analyzed often contain a variety of alterations such as proline mutations stabilizing the prefusion conformation which we did not include in our construct (table S3). Further, a decreased stability of open BriSΔ may be due to the more rigid loop (T676-Q682) lacking 8 amino acids including two protease cleavage sites. In agreement with this, it has been suggested that furin cleavage facilitates the opening of WT S (*10*). Our study indicates that the locked BriSΔ conformation is stabilized as compared to WT S by the additional interactions we described, and the shortened loop may assist to keep these properly in place. Thus, BriSΔ may have different opening kinetics from WT S and other variants, which however does not

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

noticeably affect the equilibrium binding constant to ACE2. Moreover, it has been shown that the presence of ACE2 receptors enhances the opening of the RBDs (*39*).

Our molecular dynamics simulations clearly support different stability and kinetics of RBD opening of LA-bound WT S and BriSΔ. In our targeted MD simulations, LA-bound BriSΔ requires significantly more exertion of force as compared to WT S, to adopt an open conformation (Fig. 4). Nonequilibrium MD simulations show that in WT S, both the furin site and the V622-L629 region respond rapidly to LA removal; after 0.1 ns conformational changes are observed in both regions (Fig. 5). In contrast, in BriSΔ, a fast conformational response is only observed for the V622-L629 region, indicating effectively no communication between the shortened loop and the LA pocket (Fig. 5). The nonequilibrium simulations confirm allosteric connections between the LA pocket and regions V622-L629, D808-S813 and the FPPR (Fig. 5), which are stabilized by R634 π-stacking interaction with Y829 in the FPPR in the locked conformation (Fig. 2K). LA removal destabilizes these regions.

Appreciation of the scale of intrapatient sequence variability in HIV infection decades ago led to a much better understanding of the HIV infection lifecycle and the realization of the need for combination drug therapy to combat that RNA virus (*40*). Similarly, the identification of viruses with furin cleavage site deletions in clinical samples from COVID-19 patients (16-*17*) and the differential infectivity and replication of the WT and BriSΔ viruses suggests that in the heterogenous environment of the human body, SARS-CoV-2 variants with abrogated S1/S2 cleavage may replicate preferentially in specific cell types. While WT SARS-CoV-2 clearly has a competitive advantage in respiratory tract cells (*19*), the BriSΔ virus could preferentially infect and replicate in other human cell types as a niche and reservoir to delay full clearance of SARS-CoV-2 infection by the immune system. In fact, post-mortem tissue analysis revealed viruses with a furin cleavage site deletion in heart and spleen tissue which could represent such reservoirs (*17*). Previous studies of RNA viruses have empirically demonstrated that the phenotype of a viral population can change measurably without changes in the consensus sequence due to increased diversity in evolved virus subpopulations (*41*). We propose that BriSΔ exemplifies such SARS-CoV-2 evolution, with such variants of the virus exploiting different cells and tissue types as niches for probing evolutionary space in the same infected host to evolve viral fitness, underscoring the need for analyzing individual virus genomes rather than viral populations for better understanding of the pathology, and the development of future treatments to overcome SARS-CoV-2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

SARS-CoV-2 propagation and assay

Cell culture: Human Calu-3 (ATCC® HTB-55™), Caco-2 (ATCC® HTB-37™; a kind gift from Dr Darryl Hill) and African green monkey Vero E6 (ATCC® CRL 1586™) cell lines were obtained from the American Type Culture Collection. A Caco-2 cell line expressing ACE2 (Caco-2-ACE2; a kind gift from Dr Yohei Yamauchi, University of Bristol) and Vero E6 cells modified to constitutively express TMPRSS2 (Vero E6/TMPRSS2 cells (*5*); obtained from NIBSC, UK) were also used in the study. Cells were cultured in Dulbecco's modified Eagle's medium plus GlutaMAX (DMEM, Gibco™, ThermoFisher) supplemented with 10% fetal bovine serum (FBS) and 0.1mM non-essential amino acids (NEAA, Sigma Aldrich) except Calu-3 cells, that were grown in Eagle's minimal essential medium plus GlutaMAX (MEM, GibcoTM, ThermoFisher) supplemented with 10% FBS, 0.1 M NEAA and 1mM sodium pyruvate. All cells were grown at 37 °C in 5% CO2.

SARS-CoV-2 isolation and sequencing: A mixed virus population containing the SARS-CoV-2 WT isolate hCoV-19/England/02/2020 (GISAID ID: EPI_ISL_407073) and the "Bristol" variant derived from it (BriSΔ) in which spike amino acids 679-687 (NSPRRARSV) were replaced with Ile (*11*) was grown on either Vero E6 cells or Caco-2 cells and a single virus population was isolated after two rounds of limiting dilution in either cell line (fig. S1). In brief, the virus samples were serially diluted and grown on either Vero E6 or Caco-2 cells in order to favor the growth of the BriSΔ variant or the WT virus respectively. After 5 days incubation, the culture supernatants in wells showing cytopathic effect (CPE) at the highest dilution were again diluted and the process repeated. An aliquot of culture supernatant from wells showing CPE at the highest dilution were used for RNA extraction and RT-PCR using a primer set designed to discriminate the wild type and BriSΔ viruses. Stocks of the purified viruses were produced and titered in Vero E6 cells as described previously (*11*). To verify the viral genome sequences, the wildtype and BriSΔ variant virus were grown for 24 h in Vero E6 cells before harvesting the cells and extracting the total RNA with Trizol reagent (ThermoFisher) prior to direct RNA sequencing using an Oxford Nanopore flow cell as previously described (*42*). The sequenced reads were aligned to the SARS-CoV-2 genome using minimap2 and the S gene was assessed visually for deletions as well as by an in-house script designed to look for significant deletions in this region and across the whole genome. All work with infectious SARS-CoV-2 was done inside a class III microbiological safety cabinet in a containment level 3 facility at the University of Bristol.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SARS-CoV-2 infection assays: For infection assays, cells were seeded the day prior to infection in appropriate media in µClear 96-well Microplates (Greiner Bio-one). The culture supernatants were removed, and cells were infected with serial dilutions of SARS-CoV-2 in infection medium (MEM supplemented with 2% FBS and 0.1 M NEAA) for 1 h. The virus was removed and replaced with infection medium. After 18 h incubation at 37 °C in 5% CO_2, the cells were fixed in 4% (v/v) paraformaldehyde for 60 minutes for image analysis. Fixed cells were permeabilized with 0.1% Triton-X100 in PBS and blocked with 1% (w/v) bovine serum albumin before staining with a monoclonal antibody against the SARS-CoV-2 nucleocapsid protein (N) (1:2000 dilution; 200-401-A50, Rockland) followed by an appropriate Alexa Fluor-conjugated secondary antibody (ThermoFisher) and DAPI (Sigma Aldrich). To determine the number of virus infected cells, images were acquired on an ImageXpress Pico Automated Cell Imaging System (Molecular Devices) using the 10X objective. Stitched images of 9 fields covering the central 50% of the well were analyzed for virus infected cells using Cell Reporter Xpress software (Molecular Devices). The cell number was determined by automated counted of DAPI stained nuclei and infected cells were identified as those cells in which positive N staining was detected associated with nuclear DNA staining.

SARS-CoV-2 neutralization assays: For virus neutralization assays Vero E6/TMPRSS2 cells were seeded the day prior to infection in appropriate media in µClear 96-well Microplates. Before infection, a commercial monoclonal antibody (Absolute Antibody; Sb#15) recognizing the S protein receptor-binding domain (RBD) or, alternatively, heat inactivated (30 min at 56 °C) convalescent serum from a SARS-CoV-2 infected individual, were serially diluted, in duplicate, in infection medium over an 8-fold dilution range. Equal amounts of the wild type and BriSΔ viruses (based on cell infectivity) diluted in infection medium, were mixed with the antibody/antisera dilutions and incubated for 60 minutes at 37 °C. Following the incubation, the culture supernatants were removed from the cells and replaced with the virus: antibody/serum dilutions followed by incubation for 18 h at 37°C in 5% CO_2. Cells were then fixed, and the number of infected cells determined by immunofluorescence assay and image analysis as described above. The percentage of infected cells relative to control wells infected with virus and infection media only were calculated and variable slope non-linear fit curves were assigned using GraphPad Prism v8.4.3.

Protein expression and purification.

BriSΔ protein: The construct encoding BriSΔ was synthesized at Genscript (Genscript Inc, New Jersey USA) followed by cloning into pACEBac1 plasmid (Geneva Biotech, Switzerland). The expression construct comprises SARS-CoV-2 spike amino acids 1 to 1208 with the 8 amino acid deletion, followed

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by a linker, a T4-foldon trimerization domain, another linker and finally an octa-histidine affinity purification tag (table S3). BriSΔ was produced with the MultiBac baculovirus expression system (Geneva Biotech) (*43*) in Hi5 cells using ESF921 media (Expression Systems Inc.). Supernatants from transfected cells were harvested 3 days post-transfection by centrifugation of the culture at 1,000g for 10 min followed by another centrifugation of supernatant at 5,000g for 30 min. The final supernatant was incubated with 10 mL HisPur Ni-NTA Superflow Agarose (Thermo Fisher Scientific) per 3 L of culture for 1 h at 4ºC. Subsequently, a gravity flow column was used to collect the resin bound with BriSΔ, followed by washing with 30 column volumes (CV) of wash buffer (65 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 7.5), 30 CV high salt buffer (65 mM NaH_2PO_4, 1000 mM NaCl, 20 mM imidazole, pH 7.5) and again 30 CV wash buffer. BriSΔ protein was eluted using a step gradient of elution buffer (65 mM NaH_2PO_4, 300mM NaCl, 235mM imidazole, pH 7.5). Elution fractions were analyzed by reducing Coomassie-stained SDS-PAGE. Fractions containing BriSΔ were pooled, concentrated using 50 kDa MWCO Amicon centrifugal filter units (EMD Millipore) and buffer-exchanged in SEC buffer (20mM Tris, pH 7.5, 100mM NaCl). Concentrated BriSΔ was subjected to size exclusion chromatography (SEC) using a Superdex 200 increase 10/300 column (GE Healthcare) equilibrated in SEC buffer. Peak fractions were analyzed by reducing SDS-PAGE and negative stain electron microscopy (EM) (fig. S3); fraction 8 was used for cryo-EM.

$S^{\Delta RBM}$ protein. The construct encoding $S^{\Delta RBM}$ was synthesized by Genscript (New Jersey USA) and subcloned into pACEBac1 plasmid (Geneva Biotech). $S^{\Delta RBM}$ comprises, as described above for BriSΔ, a T4-foldon trimerization domain and an octahistidine tag. In $S^{\Delta RBM}$ the ACE2-interacting receptor-binding motif (RBM) in the RBD is deleted and replaced with a glycine-serine rich linker. In addition, $S^{\Delta RBM}$ comprises mutations K986P and V987P (table S3). Protein was produced and purified as described above except that the SEC was performed in 1x PBS at pH 7.5.

S^{RRAR} protein. The construct encoding S^{RRAR} was synthesized and cloned into pACEBac1 (Genscript). S^{RRAR} comprises an intact furin site, mutations K986P and V987P, the T4 foldon trimerization motif and an octa-histidine affinity purification tag (table S3). S^{RRAR} was produced and purified as described above for $S^{\Delta RBM}$.

S^{RRAR*} protein. S^{RRAR*} was obtained by incubating S^{RRAR} with furin protease (New England Biolabs) supplemented with 1mM $CaCl_2$ and incubated at room temperature for 8 hours.

$S^{RRAR->A}$ protein. $S^{RRAR->A}$ protein was produced and purified as described (*23*). In $S^{RRAR->A}$, the furin cleavage site RRAR is replaced with a single alanine residue (*44*) (table S3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Negative stain sample preparation and microscopy

4 µL of 0.05 mg/mL SARS-CoV-2 spike protein was applied onto a freshly glow discharged (1 min at 10 mA) CF300-Cu-50 grid (Electron Microscopy Sciences), incubated for 1 min, and manually blotted. 4 µL of 3% uranyl acetate was applied onto the same grid and incubated for 1 min before the solution was blotted off. The grid was loaded onto a FEI Tecnai12 120 kV BioTwin Spirit TEM. Images were acquired at a nominal magnification of 49,000x.

Cryo-EM sample preparation and data collection

4 µL of 1.25 mg/mL BriSΔ was loaded onto a freshly glow discharged (2 min at 4 mA) C-flat R1.2/1.3 carbon grid (Agar Scientific), blotted using a Vitrobot MarkIV (Thermo Fisher Scientific) at 100% humidity and 4°C for 2 s, and plunge frozen. Data were acquired on a FEI Talos Arctica transmission electron microscope operated at 200 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV using the EPU software. Data were collected in super-resolution at a nominal magnification of 130,000x with a virtual pixel size of 0.525 Å. The dose rate was adjusted to 6.1 counts/physical pixel/s. Each movie was fractionated in 55 frames of 200 ms. 8,639 micrographs were collected in a single session with a defocus range comprised between -0.8 and -2 µm.

Cryo-EM data processing

The dose-fractionated movies were gain-normalized, aligned, and dose-weighted using MotionCor2 (*45*). Defocus values were estimated and corrected using the Gctf program (*46*). 1,168,229 particles were automatically picked using Relion 3.0 software (*47*). The auto-picked particles were extracted with a box size of 110 px (2x binning). Reference-free 2D classification was performed to select well-defined particles. After four rounds of 2D classification a total of 403,252 particles were selected for subsequent 3D classification. The initial 3D model (*23*) was filtered to 60 Å during 3D classification in Relion using 8 classes. 199,758 particles from Class 6 (fig. S4) were re-extracted with a box size of 220 px (1.05 Å/px, unbinned) and used for subsequent 3D refinement. The 3D-refined particles were then subjected to a second round of 3D classification using 3 classes. Class 2 and 3 were combined yielding 196,832 particles. These particles were subjected to 3D refinement without applying any symmetry. The maps were subsequently subjected to local defocus correction and Bayesian particle polishing in Relion 3.1. Global resolution and B factor (-97.6 $Å^2$) of the map were estimated by applying a soft mask around the protein density, using the gold-standard Fourier shell correlation (FSC) = 0.143 criterion, resulting in an overall resolution of 3.03 Å. C3 symmetry was applied to the Bayesian

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

polished C1 map using Relion 3.1, with 590,496 particles yielding a final resolution of 2.80 Å (B factor of -106.7 $Å^2$). Local resolution maps were generated using Relion 3.1 (fig. S5).

Cryo-EM model building and analysis

For model building, UCSF Chimera (*48*) was used to fit an atomic model of the SARS-CoV2 Spike locked conformation (PDB ID 6ZB5, (*23*)) into the C3 symmetrized BriSΔ cryo-EM map. The model was rebuilt using sharpened (*49*) and unsharpened maps in Coot (*50*) and then fitted into the C1 cryo-EM map. Namdinator (*51*) and Coot were used to improve the fit and N-linked glycans were built into the density for both models where visible. Restraints for non-standard ligands were generated with eLBOW (*52*). The model for C1 and C3-symmetrized closed conformation was real space refined with Phenix (*53*), and the quality was additionally analyzed using MolProbity (*54*) and EMRinger (*55*), to validate the stereochemistry of the components. Figures were prepared using UCSF chimera and PyMOL (Schrodinger, Inc).

Masked 3D classification in Relion 3.1

The refined C3 particles stack was expanded 3-fold according to C3 symmetry in Relion. The symmetry-expanded particle stack was then used as input for the masked 3D classification with the focus mask corresponding to individual single chain subunits. Masked 3D classification (*56*) was performed without alignment using 5 classes for C3-expanded particles (fig. S7). Visual inspection of the 3D classes showed 95% of the chains clustered in the first 3 classes with LA-bound in the hydrophobic pocket within the RBD.

Mass spectrometry

LA detection was performed by multi reaction monitoring (MRM) experiments using a Sciex QTrap 4500 system coupled to hydrophilic interaction liquid chromatography. The Analyst 1.7.0 software from Sciex was used for instrument control. For calibration, an unprocessed LA (Sigma Aldrich, Germany) sample dissolved in acetonitrile was used. For spike sample preparation, 100 µL of purified protein sample (1.09 mg/mL) was mixed with 400 µL chloroform for 2 hours on a horizontal shaker in a teflon-sealed glass vial at 25°C. Subsequently, the top organic phase was transferred to a new glass vial and placed in a desiccator for evaporation of the chloroform for 30 min. After all chloroform was evaporated, 50 µL acetonitrile was added to dissolve the fatty acids. From this solution, 10 µL were injected for MRM experiments. A flow rate of 0.55 mL/min was used for the binary flow elution program with acetonitrile (solvent B). The measurements were performed in negative ionization mode.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Source ionization, precursor selection and fragmentation parameters for LA monitoring using hydrophilic interaction liquid chromatography were: curtain gas = 35 psi; temperature = 600 °C; nebulizer gas = 65 psi; heater gas = 80 psi; collision gas = 9; ionization voltage = -4500V; Q1 = 279.2 m/z; Q3 = 261.2; dwell time = 250 ms; collision energy = -24 V; declustering voltage = -100 V; cell exit potential = -11 V; entrance potential = -10 V. The recorded data were analyzed using the MultiQuant 3.0.2 software from Sciex.

Surrogate virus neutralization test (sVNT) assay

The SARS-CoV-2 surrogate virus neutralization test (sVNT) kit was obtained from GenScript Inc. (New Jersey, USA). Serial dilution series were prepared (from 0-4096 nM) of the different purified SARS-CoV-2 S proteins and RBD in PBS pH 7.5. This dilution series was treated as described (*23*). Briefly, dilution series were mixed with the same volume of diluted horseradish peroxidase conjugated RBD (HRP-RBD) from the sVNT kit and incubated at 37ºC for 30 min. 100 µL of the mixtures were then added to the ELISA plate wells coated with ACE2, according to the manufacturer's protocol. Triplicates of each sample were made. The plate was then incubated at 37ºC for 15 min sealed with tape to avoid evaporation and subsequently washed 4 times with 1x wash buffer. The signal was developed by adding tetramethyl-benzidine (TMB) solution to each well, incubating 15 minutes in the dark, followed by adding stop solution supplied with the kit. The absorbance at 450 nm was immediately recorded. The data was plotted using Microsoft excel. Standard deviations from three independent replicates were added as error bars.

Surface plasmon resonance (SPR) experiments

Interaction experiments using surface plasmon resonance (SPR) were carried out with a Biacore T200 system (GE Healthcare) according to the manufacturer's protocols and recommendations. Experiments were setup as described before (*23*). Briefly, purified biotinylated ACE2 (*23*) was immobilized on a streptavidin-coated (SA) chip (GE Healthcare) at ~50 RUs. BriSΔ was injected at concentrations of 40nM, 80 nM,120 nM and 160nM. $S^{\Delta RBM}$, $S^{RRAR->A}$, S^{RRAR} and S^{RRAR*} were injected at 40 nM and 160 nM. Running buffer for all measurements was PBS buffer pH 7.5. Sensorgrams were analyzed and K_D, k_{on} and k_{off} values determined with the Biacore Evaluation Software (GE Healthcare), fitting the raw data using a 1:1 binding model. All experiments were performed in triplicates.

Molecular Dynamics Simulations

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The starting structures were comprised of closed WT (WT, P0DTC2) and furin-site deletion (BriSΔ) spike trimers, which had either been equilibrated in the presence of linoleic acid (LA) or not (apo). The target structure was the corresponding S trimer in which a single chain had an RBD in an extended conformation as in EMD-11146(*23*). Exploratory forces were applied to all four systems to find a value that resulted in a distribution of openings when subjected to 10 ns dynamic simulation. Hence starting structures were simulated for 10 ns (each system was replicated 50 times) with restraints of 0.2 kJ/mol/nm force applied to pull the RBD C-alphas to align with those of the corresponding C-alphas of the RBD of the open structure corresponding to (EMD-11146(*23*)). The BriSΔ LA-bound system was subjected to a further 50 replicate simulations at k = 0.3 kJ/mol/nm, because it failed to open at k = 0.2 kJ/mol/nm. A paired T-test was applied to test for significance when comparing the data sets.

Protein coordinates were prepared as described (*23*). Loops for the unstructured regions of the locked (LA-bound) and open (apo) cryo-EM structures were built using Chimera (UCSF)(*57*). Loop deletions were carried out in Chimera for the furin-site deletion (BriSΔ) and the sequences verified by Clustal (*58*) alignment with the published sequence (*11*). Likely disulphide bonds were reconstructed based on experimentally observed distances and each chain sequence was used in an EBI-blast check to verify WT Spike sequence post build. PROCHECK (*59*) was then used to check the quality of the resulting structure prior to simulation. ACPYPE (*60*) was used to prepare the topologies for linoleic acid. For the purposes of comparison for the targeted MD the first 10 residues of the open chain were removed so that both the starting and open target structures had the same number of residues.

Simulation details: All simulations were performed under the Amber99SB-ildn (*61*) (*62*) (*63*) forcefield in NPT ensembles at 310 K using periodic boundary conditions. We have previously shown that these protocols give structures in good agreement with experiment for the WT spike. Hydrogen atoms, consistent with pH7, were added to the complex. Short-range electrostatic and van der Waals interactions were truncated at 1.4 nm while long range electrostatics were treated with the particle-mesh Ewald method and a long-range dispersion correction applied. A simulation box extending 2 nm from the protein was filled with TIP3P water molecules and 150 mM Na^+ and Cl^- ions added to attain a neutral charge overall. Pressure was controlled by the Berendsen barostat and temperature by the V-rescale thermostat. The simulations were integrated with a leapfrog algorithm over a 2 fs time step, constraining bond vibrations with the P-LINCS method. Structures were saved every 0.1 ns for analysis in each of the simulations over 100 ns. Simulations were run on the Bristol supercomputer BlueCrystal4, the BrisSynBio BlueGem, and the UK supercomputer, ARCHER.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Software: The GROMACS-2019.2(*64*) suite of software was used to set up and perform the molecular dynamics simulations and analyses. Molecular graphics manipulations and visualizations were performed using VMD-1.9.1 (*65*) and Chimera-1.10.2 (*57*).

Nonequilibrium simulations

To map the structural changes associated with the LA removal from the FFA-binding pocket in the WT and BriSΔ spike proteins, two sets of 90 dynamical-nonequilibrium simulations were performed. In this approach, the response of a system to a perturbation is directly computed by calculating the difference of a given property between simulations with and without a perturbation. Subtracting the perturbed and unperturbed pairs of simulations at a given time, and averaging the results over many replicates, allows for the identification of the events associated with signal propagation and the determination the statistical significance of the observation (*30-32*). The perturbation was generated by the (instantaneous) removal of LA from the FFA-binding pocket. Note that the perturbation used here is not intended to represent the physical process of unbinding, but rather is designed to force the system out of equilibrium and drive a rapid response within the protein, as it adapts to LA removal. A graphical representation of the procedure is shown in Figure S9. Three equilibrium MD simulations, 200 ns each, were performed for the locked form of the unglycosylated head region of the WT and BriSΔ S proteins with LA bound. The simulation conditions for the equilibrium and dynamical-nonequilibrium simulations were identical to those described in (*23, 66*). The starting conformations for the nonequilibrium simulations were obtained from the equilibrated part of the equilibrium LA-bound simulations. Conformations were taken every 5 ns, and each of the LA molecules were (instantaneously) removed from all the FFA-binding pockets. The resulting apo system was then simulated for 5 ns (fig. S10, S11). 180 apo dynamical-nonequilibrium simulations were performed (90 simulations each for BriSΔ and WT S). Three positive ions were also removed from the solvent to maintain the electroneutrality of the systems (required for the PME method (*67*)).

The Kubo-Onsager approach (*30-32*) was used to extract and characterize the structural changes in the protein associated with LA removal. The response to perturbation is directly measured by averaging a given property (in this case, the position of the Cα atoms) in the perturbed (apo) and unperturbed (LA-bound) simulations at a given time, for multiple pairs of trajectories (fig. S9). For each pair of LA-bound equilibrium and perturbed apo dynamical-nonequilibrium simulations, the difference in positions of each Cα was determined at equivalent points in time, namely after 0, 0.1, 1, 3 and 5 ns of simulation. The pairwise comparison between the positions of Cα atoms allows for the direct identification of the conformational rearrangements while reducing the noise coming from side-chains

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fluctuation. The Cα-positional deviations at each point in time were averaged over all 90 replicates. The statistical significance of the structural changes identified here is demonstrated by a low standard error of the averages (data not shown).

Statistics. Statistical significance was determined by calculating standard deviations following standard mathematical formulae *(31)***.** Standard deviations were calculated from independent triplicates unless indicated otherwise.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References and Notes

1. B. Coutard *et al.*, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* **176**, 104742 (2020).
2. J. Shang *et al.*, Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020).
3. J. K. Millet, G. R. Whittaker, Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120-134 (2015).
4. M. Hoffmann *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 e278 (2020).
5. S. Matsuyama *et al.*, Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl. Acad. Sci. USA* **117**, 7001-7003 (2020).
6. X. Ou *et al.*, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, 1620 (2020).
7. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
8. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 (2020).
9. Y. Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592 (2020).
10. A. G. Wrobel *et al.*, SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763-767 (2020).
11. A. D. Davidson *et al.*, Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med.* **12**, 68 (2020).
12. N. S. Ogando *et al.*, SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925-940 (2020).
13. M. Sasaki *et al.*, SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS Pathog.* **17**, e1009233 (2021).
14. W. B. Klimstra *et al.*, SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. *J. Gen. Virol.* **101**, 1156-1169 (2020).
15. S. Y. Lau *et al.*, Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg. Microbes Infect.* **9**, 837-842 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Y. C. Wong *et al.*, Natural transmission of bat-like SARS-CoV-2PRRA variants in COVID-19 patients. *Clin. Infect. Dis.* (2020).
17. T. P. Peacock *et al.*, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* (2021).
18. Y. Zhu *et al.*, A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat Commun* 12, 961 (2021).
19. B. A. Johnson *et al.*, Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293-299 (2021).
20. J. Shang *et al.*, Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **117**, 11727-11734 (2020).
21. M. Dittmar *et al.*, Drug repurposing screens reveal cell-type-specific entry pathways and FDA-approved drugs active against SARS-Cov-2. *Cell Rep.* **35**, 108959 (2021).
22. N. Murgolo *et al.*, SARS-CoV-2 tropism, entry, replication, and propagation: Considerations for drug discovery and development. *PLoS Pathog.* **17**, e1009225 (2021).
23. C. Toelzer *et al.*, Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725-730 (2020).
24. L. Makowski, W. Olson-Sidford, W. W. J, Biological and Clinical Consequences of Integrin Binding via a Rogue RGD Motif in the SARS CoV-2 Spike Protein. *Viruses* **13**, (2021).
25. C. UniProt, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480-D489 (2021).
26. K. Karathanou *et al.*, A graph-based approach identifies dynamic H-bond communication networks in spike protein S of SARS-CoV-2. *J. Struct. Biol.* 212, 107617 (2020).
27. L. Zhang *et al.*, Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation. *bioRxiv*, (2021).
28. A. Rambaut *et al.*, 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 (2020).
29. E. P. Bennett *et al.*, Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736-756 (2012).
30. G. Ciccotti, G. Jacucci, I. R. McDonald, Thought-experiments by molecular dynamics. *J. Stat. Phys.* **21**, 1-12 (1979).
31. A. S. F. Oliveira, Identification of the Initial Steps in Signal Transduction in the α4β2 Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* **27**, 1171–1183 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

32. G. Ciccotti, M. Ferrario, Non-equilibrium by molecular dynamics: a dynamical approach. *Mol. Simul.* **42**, 1385-1400 (2016).
33. K. Kun Qu, X. Xiong, K. A. Ciazynska, A. P. Carter, J. A. G. Briggs, Structures and function of locked conformations of SARS-CoV-2 spike. *bioRxiv* 2021.03.10.434733, (2021).
34. J. L. Daly *et al.*, Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861-865 (2020).
35. L. Cantuti-Castelvetri *et al.*, Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860 (2020).
36. H. Winstone *et al.*, The Polybasic Cleavage Site in SARS-CoV-2 Spike Modulates Viral Sensitivity to Type I Interferon and IFITM2. *J. Virol.* **95**, e02422-20 (2021).
37. L. Yurkovetskiy *et al.*, Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739-751 (2020).
38. I. Berger, C. Schaffitzel, The SARS-CoV-2 spike protein: balancing stability and infectivity. *Cell Res* **30**, 1059-1060 (2020).
39. D. J. Benton *et al.*, Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327-330 (2020).
40. A. S. Perelson *et al.*, Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188-191 (1997).
41. V. J. Morley, M. Sistrom, J. A. Usme-Ciro, S. K. Remold, P. E. Turner, Evolution in spatially mixed host environments increases divergence for evolved fitness and intrapopulation genetic diversity in RNA viruses. *Virus Evol.* **2**, vev022 (2016).
42. I. Donovan-Banfield, A. S. Turnell, J. A. Hiscox, K. N. Leppard, D. A. Matthews, Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. *Commun. Biol.* **3**, 124 (2020).
43. D. J. Fitzgerald *et al.*, Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021-1032 (2006).
44. F. Amanat *et al.*, A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* **26**, 1033-1036 (2020).
45. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331-332 (2017).
46. K. Zhang, Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
47. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-530 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48. T. D. Goddard, C. C. Huang, T. E. Ferrin, Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281-287 (2007).
49. T. C. Terwilliger, O. V. Sobolev, P. V. Afonine, P. D. Adams, Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. D Struct. Biol.* **74**, 545-559 (2018).
50. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
51. R. T. Kidmose *et al.*, Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526-531 (2019).
52. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074-1080 (2009).
53. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861-877 (2019).
54. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2010).
55. B. A. Barad *et al.*, EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943-946 (2015).
56. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, e36861 (2018).
57. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
58. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
59. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. P. Thornton, PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* 26, 283-291 (1993).
60. A. W. Sousa da Silva, W. F. Vranken, ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5**, 367 (2012).
61. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157-1174 (2004).
62. D. A. Case *et al.*, The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668-1688 (2005).
63. K. Lindorff-Larsen *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

64. D. Van Der Spoel *et al.*, GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701-1718 (2005).
65. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38 (1996).
66. D. K. Shoemark *et al.*, Molecular Simulations suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the SARS-CoV-2 Spike Protein*. *Angew. Chem. Int. Ed. Engl.* **60**, 7098-7110 (2021).
67. U. Essmann, L. Perera, M. L. Berkowitz, A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577-8593 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements: We thank all members of the Berger and Schaffitzel teams as well as Robin Shattock (Imperial College, UK) and Adam Finn (Bristol UNCOVER Group and Children's Vaccine Centre, Bristol Medical School) for their assistance and advice. We thank Simon Burbidge, Thomas Batstone and Matt Williams for computation infrastructure support. We would like to thank the Advanced Computing Research Centre (ACRC) at the University of Bristol for access to BlueCryo, BlueCrystal Phase 4 and BlueGEM, and the UK HECBioSim for access to the UK supercomputer, ARCHER. We are particularly grateful to Thiru Thangarajah (Genscript) for early access to Genscript's cPass™ SARS-CoV-2 Neutralization Antibody Detection/Surrogate Virus Neutralization Test Kit (L00847). We thank Sebastian Fabritz and the Core Facility for Mass Spectrometry at the Max Planck Institute for Medical Research for their support on MS measurements. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Funding: This research received support from the Elizabeth Blackwell Institute for Health Research and the EPSRC Impact Acceleration Account EP/R511663/1, University of Bristol, from BrisSynBio a BBSRC/EPSRC Research Centre for synthetic biology at the University of Bristol (BB/L01386X/1) (to I.B., C.S., A.J.M., D.K.S., A.S.F.O.) and from the BBSRC (BB/P000940/1) (to C.S. and I.B.). This work received generous support from the Oracle Higher Education and Research program to enable cryo-EM data processing using Oracle's high-performance public cloud infrastructure (https://cloud.oracle.com/en_US/cloud-infrastructure) and the EPSRC through a COVID-19 project award via HECBioSim to access ARCHER (A.J.M.). We acknowledge support and assistance by the Wolfson Bioimaging Facility and the GW4 Facility for High-Resolution Electron Cryo-Microscopy funded by the Wellcome Trust (202904/Z/16/Z and 206181/Z/17/Z) and BBSRC (BB/R000484/1). The authors are grateful to University of Bristol's Alumni and Friends, which funded the ImageXpress Pico Imaging System. O.S. acknowledges support from the Elisabeth Muerer Foundation, the Max Planck School Matter to Life and the Heidelberg Biosciences International Graduate School. J.S. is the Weston Visiting Professor at the Weizmann Institute of Science, part of the excellence cluster CellNetworks at Heidelberg University and acknowledges funding from the European Research Council (ERC, contract no. 294852), SynAd and the MaxSynBio Consortium, funded by the Federal Ministry of Education and Research of Germany and the Max Planck Society, from the SFB 1129 and Project 240245660-SFB1129 P15 of the German Research Foundation (DFG) and from the Volkswagen Stiftung (priority call "Life?"). A.D.D. and D.A.M. are supported by the United States Food and Drug Administration (HHSF223201510104C) and UK Research and Innovation/Medical

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Research Council (MRC) (MR/V027506/1). M.K.W is supported by MRC grants MR/R020566/1 and MR/V027506/1 (awarded to A.D.D). A.J.M. is supported the British Society for Antimicrobial Chemotherapy (BSAC-COVID-30) and the EPSRC (EP/M022609/1, CCP-BioSim). I.B. acknowledges support from the EPSRC Future Vaccine Manufacturing and Research Hub (EP/R013764/1). C.S. and I.B. are Investigators of the Wellcome Trust (210701/Z/18/Z; 106115/Z/14/Z).

Author contributions: C.S. and I.B. conceived and guided the study. K.G. and F.G. produced proteins, K.G. purified and analyzed all protein samples and carried out biochemical and biophysical experiments, S.K.N.Y. and U.B. prepared grids and collected EM data, S.K.N.Y. carried out image processing, C.T. carried out model building and structural analysis. A.D.D. and M.K.W. performed all live virus CL3 work and analyzed data. D.A.M and A.A. performed direct RNA sequencing and analysis of sequence data. D.K.S., A.S.F.O. and A.J.M. performed all MD simulations. O.S. and J.S. performed and interpreted mass spectrometry. I.B., D.F. and C.S. wrote the manuscript with input from all authors.

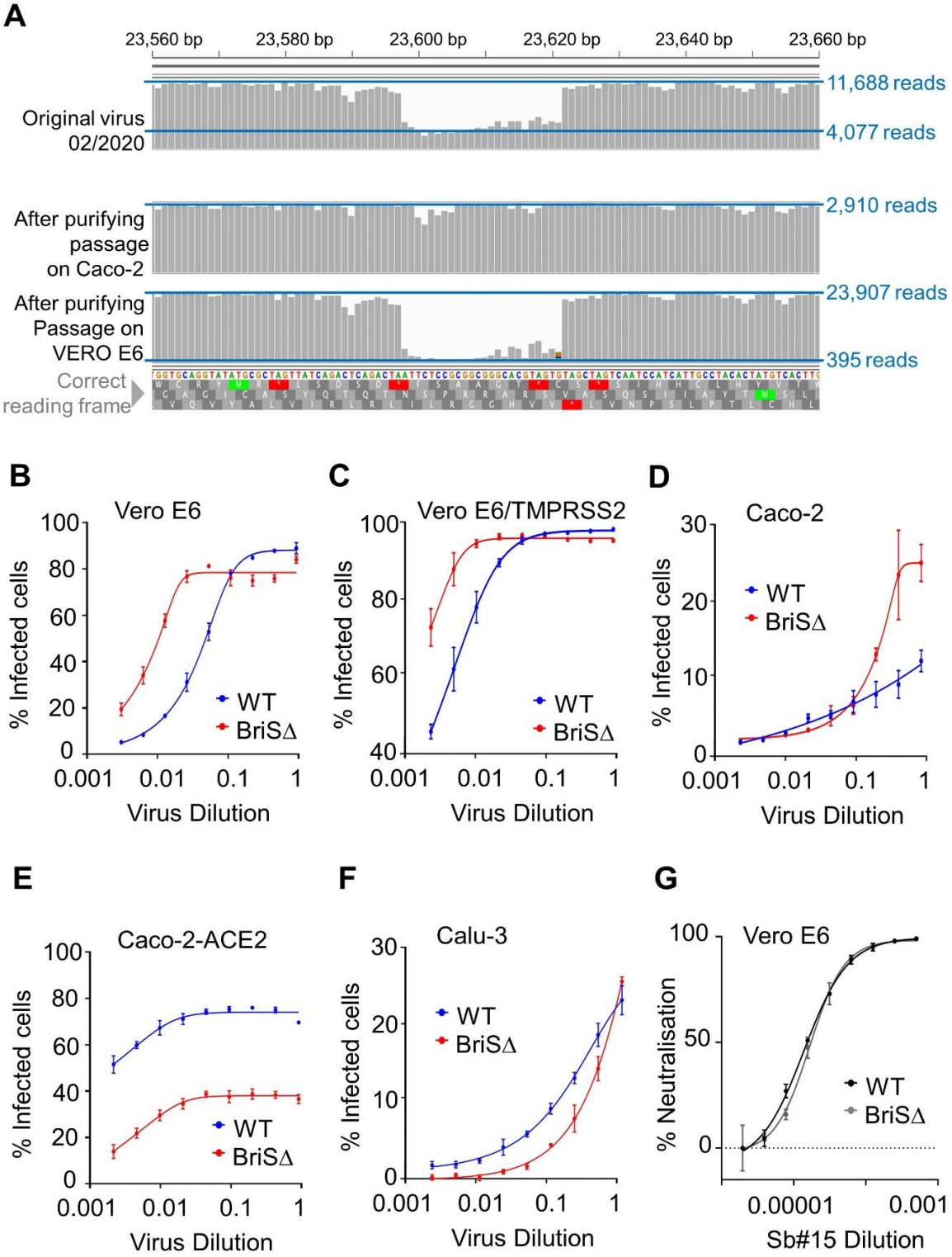
Competing interests: The authors declare competing interests. F.G. and I.B. report shareholding in Imophoron Ltd, unrelated to this Correspondence. D.F. and I.B. report shareholding in Geneva Biotech SARL, unrelated to this Correspondence. C.S., D.F. and I.B. report shareholding in Halo Therapeutics Ltd related to this Correspondence. Patent applications describing methods, material compositions and formulations based on the present observations have been filed.

Data and materials availability: Datasets and coordinates generated during the current study have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-12818 (C3 structure) and EMD-12842 (C1 structure) and in the Protein Data Bank (PDB) under accession numbers: 7OD3 (C3 structure) and 7ODL (C1 structure). Reagents are available from C.S. and I.B. under a material transfer agreement with the University of Bristol.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

Figure 1. "Bristol" SARS-CoV-2 BriSΔ variant.



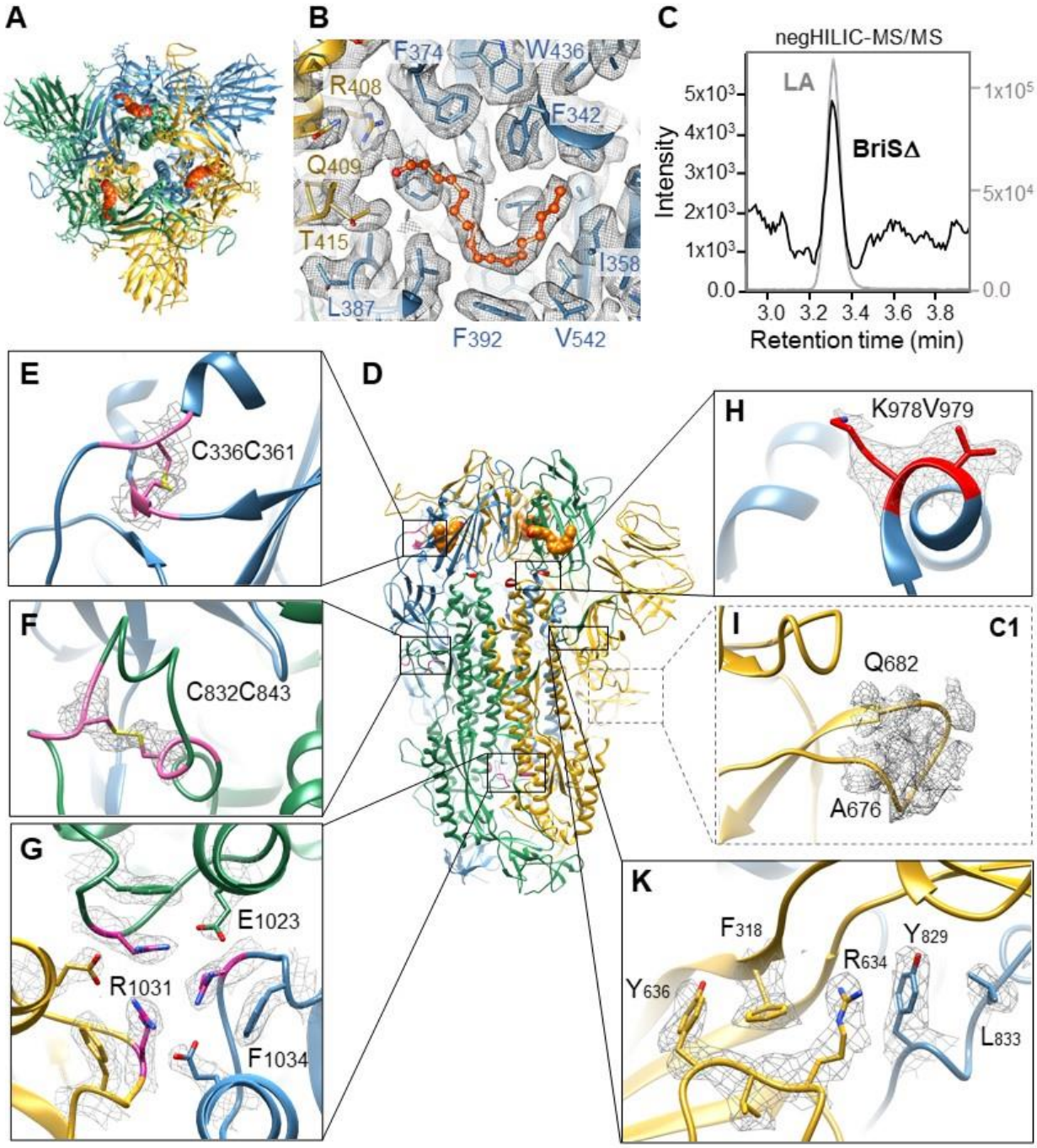
(A) Depth of read across S glycoprotein gene at furin cleavage site is shown for three different stocks of SARS-CoV-2 isolate hCoV-19/England/02/2020. Horizontal blue lines indicate read depth. Original stock of virus (top) evidences sharp decline in read depth corresponding to in-frame deletion of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

furin cleavage site and indicative of a mixed population of viruses. The middle panel shows the sequencing depth at same region for a virus stock that has been isolated by growth on human Caco-2 cells and purified by limiting dilution. The bottom panel shows the equivalent sequencing data for a stock of the SARS-CoV-2 BriSΔ variant grown on Vero E6 cells and purified by limiting dilution. **(B-F)** Approximately equal amounts of the WT virus and BriSΔ virus based on genome amounts (estimated by qRT-PCR) were diluted (2-fold dilution series starting with neat virus) and used to infect Vero E6, Vero E6/TMPRSS2, Caco-2, Caco-2-ACE2 and Calu-3 cells. At 18 h after infection, cells were fixed, stained with an anti-N antibody and the % of cells infected was determined by immunofluorescence microscopy. **(G)** WT virus and BriSΔ virus were used to infect Vero E6/TMPRSS2 cells in the presence of a range of dilutions of a commercial antibody against SARS-CoV-2 RBD. Cells were infected with equal amounts of infectious virus (based on cell infectivity). At 18 h after infection, cells were fixed and stained with an anti-N antibody and the % of cells infected was determined by immunofluorescence microscopy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. Cryo-EM structure of BriSΔ.



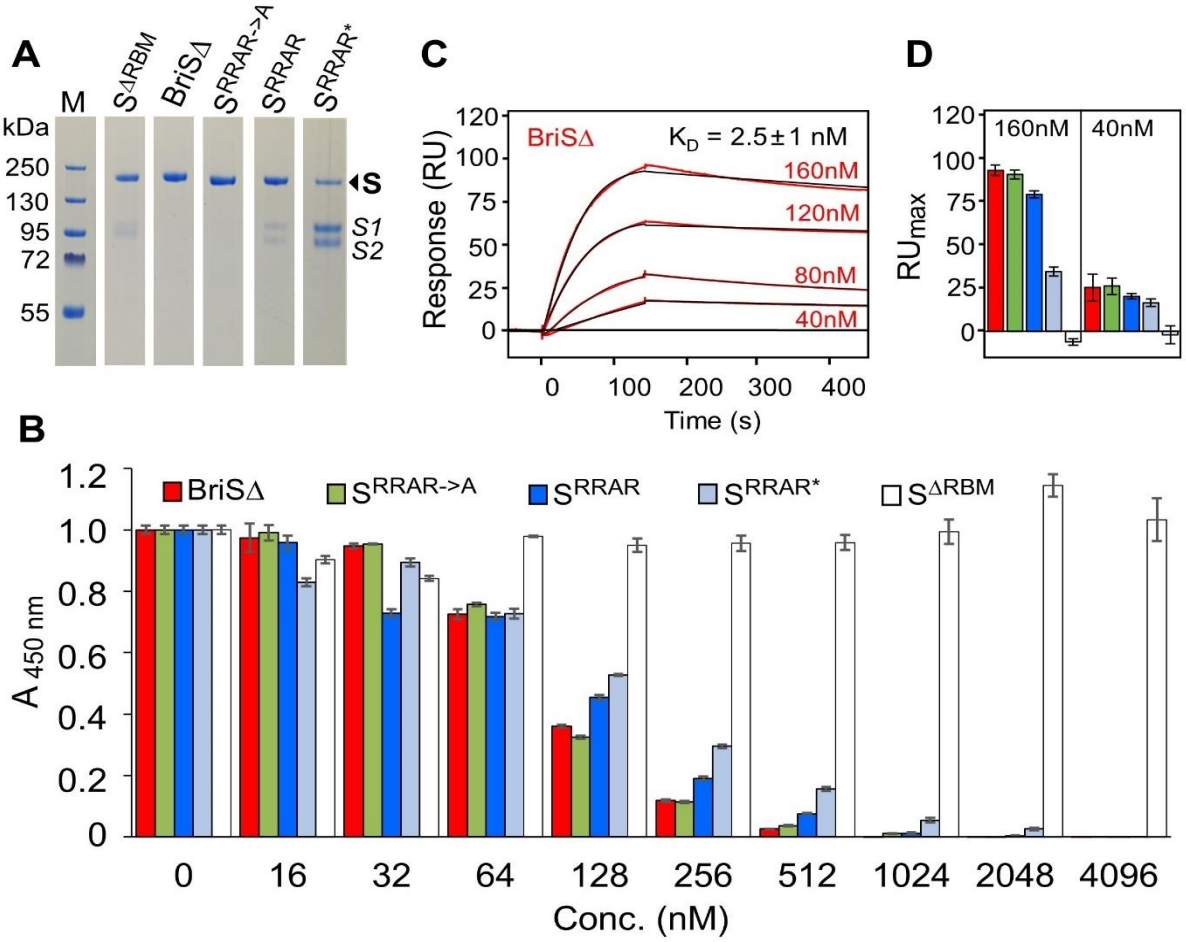
(**A**) Top view cartoon representation with trimer subunits colored yellow, green and blue, LA shown as orange spheres. (**B**) Composite LA-binding pocket formed by adjacent RBDs (yellow and blue). EM density is shown as grey-colored mesh; LA ligand (orange) in sticks and balls representation (**C**) Selected-reaction monitoring mass chromatogram of hydrophilic interaction liquid chromatography (HILIC) coupled tandem mass spectrometry analysis for 10 ng/mL LA analytical standard (grey) and BriSΔ protein preparation (black). (**D**) Side view of BriSΔ trimer with boxes for the close-up views in panels E to K. (**E**) Disulfide bond between Cys336 and Cys361 in the RBD. (**F**) Cys832 forms a

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

disulfide bond with Cys843 and stabilizes the fusion peptide proximal region (FPPR). (**G**) H-bond cluster involving R1031 cation-π interaction on F1034 and forming a salt bridge to E1023. (**H**) BriSΔ K978 and V979. K978 sidechain EM density indicates flexibility. (**I**) BriSΔ shortened loop devoid of furin and S1/S2 proteolytic sites modelled as a poly-alanine chain in the C1 structure. (**K**) R634 cation-π interaction to Y829 in the FPPR of neighboring polypeptide chain.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

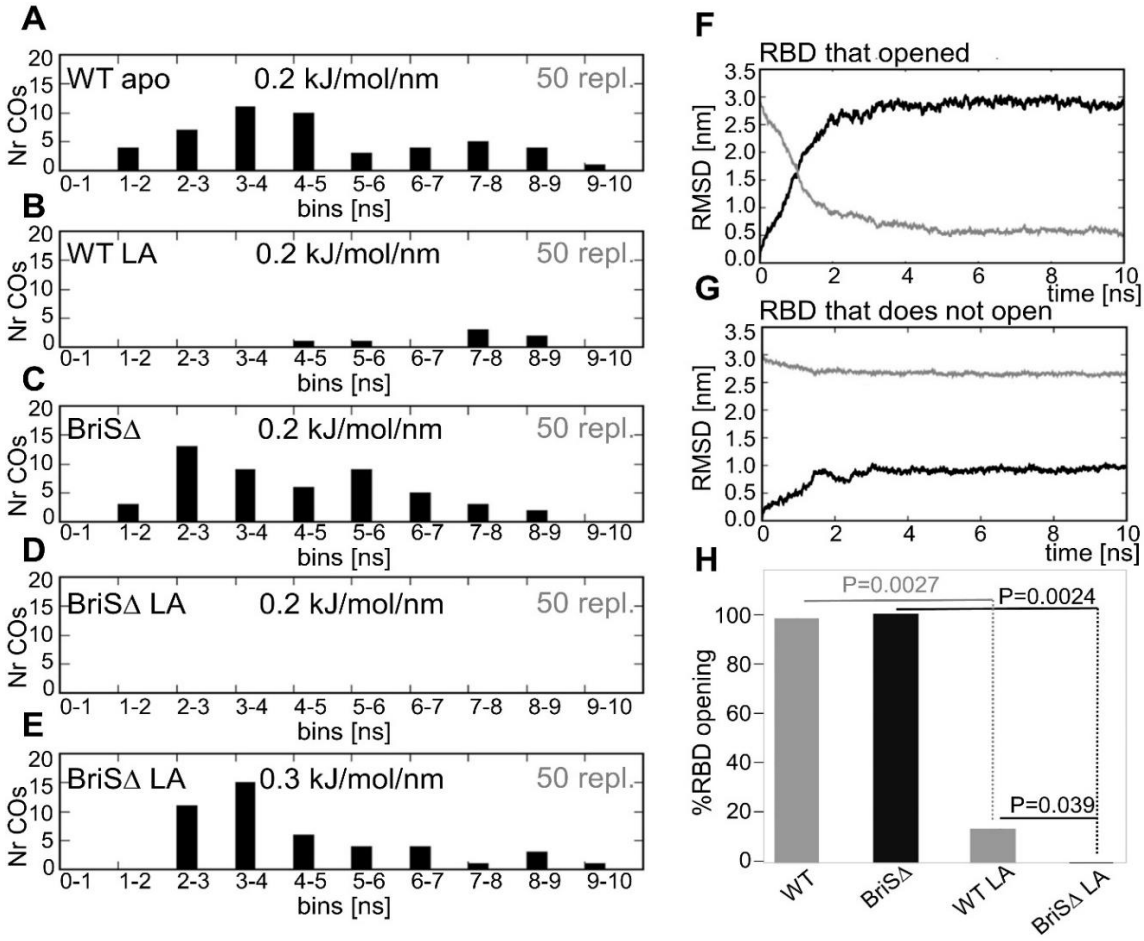
Figure 3. Functional analysis of SARS-CoV-2 S proteins.



(**A**) Coomassie-stained SDS-PAGE sections of S protein variants used for biochemical characterization. (**B**) Competition ELISAs utilizing immobilized ACE2, HRP-labelled RBD and S proteins (shown in different colors) at the concentrations indicated. Error bars: standard deviations, three replicates. (**C**) Surface plasmon resonance (SPR) of BriSΔ. Concentrations between 40 nM and 160 nM were flowed over 50 RU of biotinylated ACE2 immobilized on a streptavidin-coated sensor chip. Black lines correspond to a global fit of the data using a 1:1 binding model. Each experiment was repeated independently three times. All protein concentrations were used to calculate the K_D value. (**D**) SPR analysis comparing maximal binding (RUmax) of the S protein variants at two representative concentrations (160 nM, 40 nM). Color coding as in panel B.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

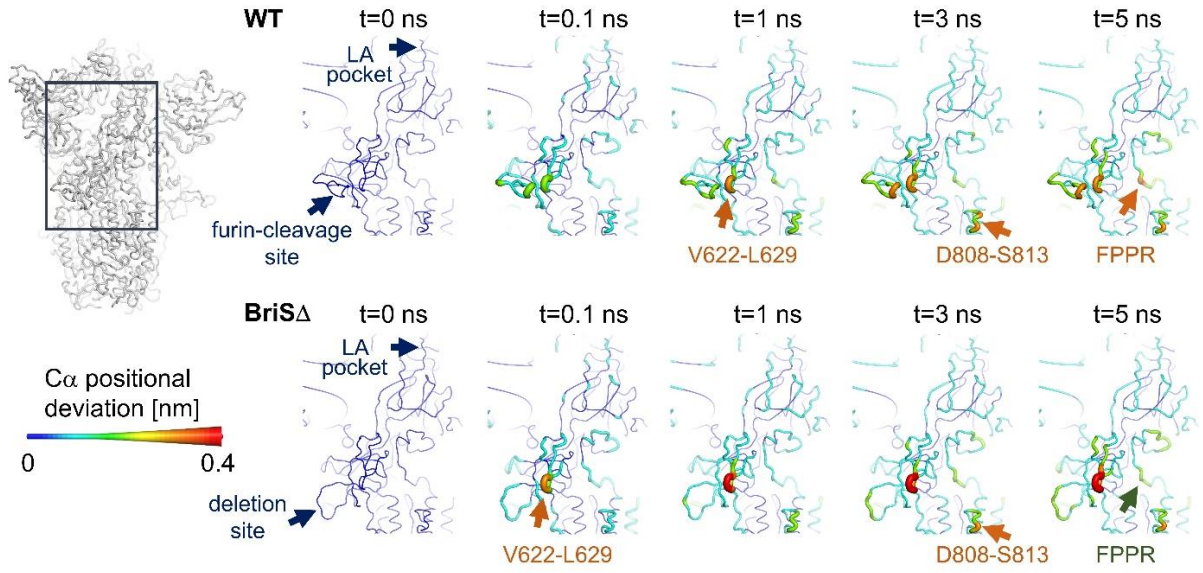
Figure 4. Impact of furin-site loop deletion and LA binding on RBD opening.



(**A**-**E**) Distribution of the cross-overs (COs) for 50 x 10 ns replicates carried out for each of the four spike systems under a force constant of 0.2 kJ/mol/nm for (**A**) WT apo, (**B**) WT with LA bound (**C**) BriSΔ apo and (**D**) BriSΔ with LA bound. (**E**) Additional force (0.3 kJ/mol/nm) is required to open BriSΔ with LA. (**F**) Example of a cross-over plot of a system in which the RBD opens. The black trace shows the RMSD from the closed spike (the starting structure). The grey trace shows the RMSD from the open target at time points along the trajectory over which the force was applied. The point at which the traces cross (the cross-over time) is measured in nanoseconds. (**G**) Example of a cross-over plot in which the RBD fails to cross-over as the RBD does not open under the force constant applied. (**H**) Comparison of RBD opening events. Significance of the differences between WT apo form (WT-apo) and WT LA-bound form (WT-LA) (P=0.0027), between BriSΔ apo form (BriSΔ-apo) and BriSΔ LA-bound form (BriSΔ-LA) (P=0.0027) and WT-LA and BriSΔ-LA (0.039) was determined by paired T-test.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.11.443384; this version posted May 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5. Dynamical nonequilibrium simulations elucidate impact of LA on S structure.



Average Cα-positional deviations are shown at 0, 0.1, 1, 3 and 5 ns following LA removal from the FFA-binding pockets of WT S (above) and BriSΔ (below). The Cα deviations between the simulations with and without LA were determined for each residue and averaged over the 90 pairs of simulations for each system. The Cα average deviations are mapped onto the structure used as the starting point for the LA-bound equilibrium simulations. Both colors and cartoon thickness indicate the average Cα-positional deviation values. Important regions of changes are highlighted.