bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

Relationship between monomer packing and receptor binding domain pocket status in the spike trimer of SARS-CoV-2 variants

Jim Warwicker[a,*]

[a] School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Institute of Biotechnology, University of Manchester, M1 7DN, UK.

[*] Corresponding author:
Jim Warwicker, email jim.warwicker@manchester.ac.uk, tel. +44 (0)161 306 4490

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

Abstract

Existence of a SARS-CoV-2 spike protein trimer form with closer packing between monomers when receptor binding domains (RBDs) are all down, locked as opposed to closed, has been associated with linoleic acid (LA) binding at neutral pH, or can occur at acidic pH in the absence of LA binding. The relationship between degree of closure of the LA binding pocket of the RBD, and monomer burial in the trimer, is examined for a range of spike protein structures, including those with D614G mutation, and that of the Delta variant (which also carries D614G). Some spike protein structures with this aspartic acid mutation show monomer packing approaching that of the locked form (at neutral pH, without LA binding) for two segments, a third (around the RBD) remains less closely packed. Analysis of other coronavirus RBD structures suggests that mutation of the RBD in spike protein of the Omicron variant could lead to LA binding pocket changes. It is proposed that these changes could lead to one of two consequences for the Omicron variant spike protein (which also has the D614G mutation), at neutral pH and without LA binding, either easier access to a locked form throughout that leads to cooperative transitions between all RBD down and all RBD up, or maintenance of a spike trimer with locked characteristics C-terminal to the RBD at the same time as the RBD is free to transit between down and up states. The situation may also be impacted by spike protein charge mutations in the Omicron lineage that alter pH-dependence around the RBD, in a similar way to the changes induced elsewhere by D614G.

Keywords: pH-Dependence; SARS-CoV-2; Spike protein; Coronaviruses

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

Introduction

Viruses must enter a cell and release their genome for copying, they must also exit the cell. For both entry and exit, pH-dependent processes may play a role, including receptor-mediated endocytosis into low pH endosomes and navigation of acidic pH in the secretory pathway (Robinson et al., 2018). For the membrane enveloped coronavirus family, cell entry may be through fusion directly at the cell membrane or through fusion at an intracellular membrane, subsequent to receptor-mediated endocytosis of the virus (Whittaker et al., 2021). Although there is evidence that altering endosomal pH impedes viral entry to some extent (Prabhakara et al., 2021) for the SARS-CoV-2 coronavirus (causative agent of the covid-19 pandemic), the precise balance of genome release routes (cell surface or interior) may depend on other factors, such as the priming cleavage of S1 and S2 subunits of the spike (S) protein (Papa et al., 2021). For membrane enveloped viruses, mechanisms evolve to protect against mis-timed low pH-induced membrane fusion events in the acidic pH secretory pathways for newly synthesised virus (Fields and Kielian, 2015). It is apparent that questions of pH-dependence in these systems are complex, relate to pH inside and outside of the cell, and couple to other determinants of stability, including receptor and other ligand binding, and proteolytic cleavages. A strong signal in calculations of pH-dependence, relative to the unfolded state, for each of the pre and post-fusion structures of SARS-CoV-2 S protein, are three His sidechains in the S2 subunit, but since their burial and predicted pH-dependence are uniform between pre and post-fusion forms, it is unclear whether they are functionally relevant (Warwicker, 2021). For coronaviruses, conformational variation of pre-fusion S protein is apparent from cryo-EM structures. Indeed, there are major questions to be answered about the pH-dependence of S1/S2 subunit S-protein structure and function, even without considering the large-scale conformational changes of the S2 subunit that follow S1 subunit release, and accompany membrane fusion (Cai et al., 2020).

Based on the response of S protein trimer structure to acidic pH (Zhou et al., 2020; Qu et al., 2021), a scheme has been suggested in which the trimers on a virus within the acidic pH of the secretory pathway would be protected in a more tightly packed (locked) conformation (Qu et al., 2021). This suggestion is consistent with the discovery of a locked and more tightly packed form of SARS-CoV-2 S protein trimer at the normal pH (7.5 to 8) of structure solution, due to binding of a linoleic acid pocket factor in the receptor binding domain (RBD) (Toelzer et al., 2020). These observations prompt questions of which S protein amino acids are sensing pH, and how these might be coupled to pocket factor binding and packing/interface changes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

A study of S protein trimers grouped according to RBD down (with locked and closed forms treated separately), and RBD up, reported differences in the predicted pH-dependence (Lobo and Warwicker, 2021). The largest contributions to these differences arose from interface tightening in the trimer between closed and locked forms, rather than the well-known difference between open (RBD up) and closed (RBD down) forms, and specifically from Asp and Glu sidechains, prominent amongst which is D614. It was proposed that these carboxylate groups have relatively unperturbed pK_as in open and closed forms, but that increased burial in locked forms leads to destabilisation. Depending on the balance of overall interface energetics, the effect could be to favour locked forms at acidic pH (less destabilised carboxylates, pH closer to their normal pK_a), and closed/open forms at neutral pH (since locked form carboxylates would be more destabilised, with pH further from their normal pK_a) (Lobo and Warwicker, 2021), a similar proposal to that of protection in the secretory pathway (Qu et al., 2021).

The picture for SARS-CoV-2 spike protein encountered in the first year of the covid-19 pandemic is therefore of a trimer that can exist in a more tightly packed form (termed locked) either with pocket factor binding at neutral pH, or at acidic pH in the absence of pocket factor. However, the spike protein is changing, leading to the question of what the consequences may be for relative stability of the locked, closed, and open forms.

Materials and Methods

Inspection of the spike protein LA binding pocket (Fig. 1) revealed that its open/closed state could be characterised by a distance between the average of the 4 β-strand C_α atoms on one side of the pocket (amino acids K378, C379 and C432, V433), and the average of the two C_α atoms on the opposite side (amino acids L368, Y369). Whereas the 4 β-strand amino acids on one side, including a disulphide bridge between C379 and C432), are structurally relatively invariant, L368, Y369 are in a turn that is part of the gating mechanism for the pocket (Toelzer et al., 2020).

Burial of spike protein monomer within a trimer for structures of Delta variant S protein was calculated as described previously for other SARS-CoV-2 variant trimers (Lobo and Warwicker, 2021), with d-SASA equal to the change in solvent accessible surface area between free monomer and monomer within a trimer. These d-SASA values are used either as averages over monomers within one trimer structure, or as averages over all monomers within a specified

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

sub-grouping of structures. When plotted as profiles over the spike protein sequence, a further difference is made, between d-SASA profiles, to give a direct indication of where packing varies between systems.

RBD structures for other coronaviruses were examined by searching the SARS-CoV-2 spike sequence, id P0DTC2 from UniProt (UniProt, 2019) against the RCSB/PDB (Berman et al., 2007) with the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) at the National Center for Biotechnology Information. Protein structures were aligned with Swiss PDB Viewer (Guex and Peitsch, 1997) and visualised with Swiss PDB Viewer and PyMol.

Results and Discussion

Previous predictions of potential acid sensor residues, and calculations of the change in solvent accessible surface area (d-SASA) upon spike monomer incorporation in trimer (in various trimer forms) (Lobo and Warwicker, 2021), are supplemented with data for two recent structures of the Delta variant (RCSB/PDB (Berman et al., 2007) identifiers 7v7n (10.2210/pdb7V7N/pdb) and 7sbk (Zhang et al., 2021b)), and consideration of sequence changes around the RBD pocket of the Omicron variant S protein. A new measure is constructed of distance across the RBD pocket in which linoleic acid can bind (Fig. 1). Comparing pocket distance and monomer interface in a scatter plot (Fig. 2), it is apparent that open and closed forms cluster with a distance consistent with a closed pocket, and at low (open) and intermediate (closed) monomer burial in the trimer. Locked forms are quite separate, with a larger (and occupied) pocket, and high monomer burial. The set of pH-locked forms have closed pockets, and monomer burial intermediate between closed and locked forms, but at acidic pH (all other data in the plot is for structures at or slightly above neutral pH). Structures in both the D614Gset-closed and Delta-closed groups carry the D614G mutation, and have been filtered for trimers in which all RBDs are in the down conformation. For both groups, the clustering exhibited by other forms is lost, monomer burial in the trimer can vary substantially, and one structure also has an intermediate pocket distance. Variants with the S protein D614G mutation increase recovery in cryo-EM structure determination of forms that approach a locked degree of monomer interface burial, but at neutral rather than acidic pH, with evidence that they are also able to sample conformations of intermediate pocket closure (Fig. 2). The barrier between locked and closed forms at neutral pH is reduced by the D614G substitution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

In order to establish the location of interface differences (for monomer burial in trimer), and how they change in a D614G structure with similar burial to locked forms, differences between 7krq (Zhang et al., 2021a) and (averaged) locked form burial are studied (Fig. 3). There is an interaction of equivalent helices in the trimer towards the C-terminus, which is only present for some structures. Other than this, major regions of interaction differences between 7krq and the closed form (555 – 670) and (830 – 860) have, by contrast, similar monomer burial in 7krq (neutral pH, D614G mutation, pocket closed) and the locked form (neutral pH, linoleic acid bound). These two regions contact between monomers in the trimer, and it has been suggested that a salt-bridge lost between D614 and K854 (of neighbouring monomers) in variants carrying the D614G mutation, effectively breaks a latch and leads to a greater fraction of open (RBD up) forms, thereby enhancing interaction with the receptor (Yurkovetskiy et al., 2020). Another interpretation of the effect of D614G mutation is that it stabilises the S protein trimer against dissociation (Zhang et al., 2021a), which is consistent with the hypothesis that once a destabilising burial of D614 is lost, a locked form is more accessible at neutral pH (Lobo and Warwicker, 2021).

A further region of interest is (360 – 470), where there is most differentiation between the D614G 7krq structure and the locked / pocket occupied form (Fig. 3). This segment is of interest for several reasons. Within it lies parts of the linoleic acid pocket, including the gating mechanism (Toelzer et al., 2020) (Fig. 1). It also contains carboxylate groups of D405, D420, and E465 that are proposed (together with D614) to couple burial and pH-dependent stability (Lobo and Warwicker, 2021). Interestingly, in this region, the acid pH-locked structure 6xlu (Zhou et al., 2020) is more similar to the neutral pH locked form than is the D614G 7krq structure (Fig. 3). If there were one or more buried and destabilised Asp/Glu sidechains in this region, then such behaviour would be expected.

Finally, S protein of the Omicron variant (https://www.ecdc.europa.eu/en/covid-19/variants-concern) carries S371L, S373P, S375F mutations in the sequence next to the RBD pocket gate (Fig. 4). There are many further mutations in Omicron S protein, which are being discussed in the context of various features, including the extent to which they could mediate escape of an immune response primed by vaccine or prior infection (Greaney et al., 2021; Zahradník et al., 2021), which are not considered here. The Omicron variant also carries the D614G S protein mutation, and two further aspects are noted. First, considering that a mutation to proline (S373P) might lead to change in mainchain conformation, the RCSB/PDB holds a structure (5gyq) (Huang et al., 2016) for the RBD of coronavirus HKU9, with proline at the equivalent position of S373 in SARS-CoV-2 S protein (Fig. 4). This residue lies in a segment

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

connecting two sides of the LA pocket in the RBD of SARS-CoV-2. On one side is a β-sheet that is relatively invariant in the structural alignment, and on the other side are more structurally variable turns of the gating region. Reduced mainchain conformational freedom at this proline in the Omicron spike protein could bias towards a closed pocket conformation (Fig. 4). Second, D405 (Fig. 3) of one monomer abuts S373 and S375 of a neighbouring monomer, in the locked form. Thus, changes in this region are likely to impact on both the pocket and on predictions of pH-dependence. Other relevant factors are the presence of Y505H in the Omicron S protein variant, which neighbours D405, and in another Omicron lineage (https://github.com/cov-lineages/pango-designation/issues/361), S373P is maintained, but with other mutations in this region, including D405N.

The difference in d-SASA between 7krq (D614G) and locked and LA-bound form is split largely into two regions, omitting the C-terminal helical interaction (Fig. 3). Segments C-terminal to the RBD have similar burial, this D614G structure approaches a locked form here, whereas the RBD of 7krq is less well packed than the locked and LA-bound form. Discussion of the Omicron variant S protein suggests that the RBD pocket is altered, and in terms of d-SASA this could take one of two routes. According to the Asp/Glu burial model for spike protein pH-dependence, either Y505H or (in particular) D405N could ameliorate neutral pH instability, leading to a more closely packed RBD than that seen for 7krq. Equally, changes of ionisable groups, and other mutations, could reflect not just reduction in pH-sensing, but also removal of tighter packing in this region. Common to both models for changes in Omicron S protein relative to D614G S protein, is the maintenance of closer packing (of locked form extent) in regions C-terminal to the RBD, facilitated by the D614G mutation. The divergence in the two models is whether coupling between locked-like packing in the RBD, and C-terminal to the RBD, is broken. If it is, then S trimer protection would still be afforded through locking at interfaces not involving the RBD, with the RBD available for rapid conversion between closed and open (ACE2 binding forms). On the other hand, if locked interfaces act in concert throughout the Omicron S trimer, as appears to be the case for LA-bound S trimer (Toelzer et al., 2020), and with increased locking of the RBD, how would this be consistent with the view that a higher proportion of RBD up correlates directly with ACE2 binding and infectivity? In that case, it is possible that the cooperativity in S protein interaction with ACE2, thought to enhance progression to a fusion competent state (Pak et al., 2021), could play a role. Since locked conformations are more tightly packed between monomers, they could be involved in more cooperative transitions to RBD up (open) forms than would be the more loosely packed

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

closed conformations. In this scheme, protection against S trimer dissociation prior to cell engagement (Zhang et al., 2021a) would be a coupled set of close-knit interfaces throughout the trimer. Thus, the suggested models for Omicron S protein at neutral pH are of a cooperative whole S trimer transition from locked to RBD up, or of locking C-terminal to the RBD, with the RBD more free to transition between down and up conformations. It is not possible to use the pocket distance as guide to one model over the other, since the degree of RBD burial in the S trimer can vary substantially (Fig. 3) with little change in the pocket distance (Fig. 2), and also there are other (non-ionisable group) mutations in this region of the Omicron S protein to consider (Zahradník et al., 2021).

Conclusion

Although structural studies of SARS-CoV-2 spike proteins continue at pace, many mechanisms remain unknown, including that behind the pH-dependence of S trimer conformation, and how it might relate to function. The proposal that Asp/Glu sidechains could be responsible for observed pH-dependent effects is yet to be tested. An alternate hypothesis is that histidines H49 and H519 could be mediating pH-dependence (Qu et al., 2021), but these give small predicted pH-dependence (Lobo and Warwicker, 2021) and are located in regions without large interface change between structural forms (Fig. 3). Here the degree to which the D614G mutation allows access of spike trimers (without LA binding) to conformations that are similar to the LA-bound locked conformations, is assessed. Analysis of mutations around the gate for the LA binding pocket (Toelzer et al., 2020), in the Omicron variant spike protein, suggests that further changes may occur in locked conformation accessibility at neutral pH (and without LA binding), and also that ionisable group changes could contribute to altered pH-dependence, beyond that associated with D614G.

Funding

This work was supported in part by award EP/N024796/1 from the UK Engineering and Physical Sciences Research Council.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215(3)**,** 403-410. doi: 10.1016/S0022-2836(05)80360-2.

Berger, I., and Schaffitzel, C. (2020). The SARS-CoV-2 spike protein: balancing stability and infectivity. *Cell Res* 30(12)**,** 1059-1060. doi: 10.1038/s41422-020-00430-4.

Berman, H., Henrick, K., Nakamura, H., and Markley, J.L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35(Database issue)**,** D301-303. doi: 10.1093/nar/gkl971.

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Jr., et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* 369(6511)**,** 1586-1592. doi: 10.1126/science.abd4251.

Fields, W., and Kielian, M. (2015). Interactions involved in pH protection of the alphavirus fusion protein. *Virology* 486**,** 173-179. doi: 10.1016/j.virol.2015.08.028.

Greaney, A.J., Starr, T.N., and Bloom, J.D. (2021). An antibody-escape calculator for mutations to the SARS-CoV-2 receptor-binding domain. *bioRxiv***,** 2021.2012.2004.471236. doi: 10.1101/2021.12.04.471236.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18(15)**,** 2714-2723. doi: 10.1002/elps.1150181505 [doi].

Huang, C., Qi, J., Lu, G., Wang, Q., Yuan, Y., Wu, Y., et al. (2016). Putative Receptor Binding Domain of Bat-Derived Coronavirus HKU9 Spike Protein: Evolution of Betacoronavirus Receptor Binding Motifs. *Biochemistry* 55(43)**,** 5977-5988. doi: 10.1021/acs.biochem.6b00790.

Lobo, V.R., and Warwicker, J. (2021). Predicted pH-dependent stability of SARS-CoV-2 spike protein trimer from interfacial acidic groups. *Comput Struct Biotechnol J* 19**,** 5140-5148. doi: 10.1016/j.csbj.2021.08.049.

Pak, A.J., Yu, A., Ke, Z., Briggs, J.A.G., and Voth, G.A. (2021). Cooperative multivalent receptor binding promotes exposure of the SARS-CoV-2 fusion machinery core. *bioRxiv*. doi: 10.1101/2021.05.24.445443.

Papa, G., Mallery, D.L., Albecka, A., Welch, L.G., Cattin-Ortola, J., Luptak, J., et al. (2021). Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLoS Pathog* 17(1)**,** e1009246. doi: 10.1371/journal.ppat.1009246.

Prabhakara, C., Godbole, R., Sil, P., Jahnavi, S., Gulzar, S.E., van Zanten, T.S., et al. (2021). Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors. *PLoS Pathog* 17(7)**,** e1009706. doi: 10.1371/journal.ppat.1009706.

Qu, K., Xiong, X., Ciazynska, K.A., Carter, A.P., and Briggs, J.A.G. (2021). Structures and function of locked conformations of SARS-CoV-2 spike. *bioRxiv*.

Robinson, M., Schor, S., Barouch-Bentov, R., and Einav, S. (2018). Viral journeys on the intracellular highways. *Cell Mol Life Sci* 75(20)**,** 3693-3714. doi: 10.1007/s00018-018-2882-0.

Toelzer, C., Gupta, K., Yadav, S.K.N., Borucu, U., Davidson, A.D., Kavanagh Williamson, M., et al. (2020). Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science*. doi: 10.1126/science.abd3255.

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47(D1)**,** D506-D515. doi: 10.1093/nar/gky1049.

Warwicker, J. (2021). A model for pH coupling of the SARS-CoV-2 spike protein open/closed equilibrium. *Brief Bioinform* 22(2)**,** 1499-1507. doi: 10.1093/bib/bbab056.

Whittaker, G.R., Daniel, S., and Millet, J.K. (2021). Coronavirus entry: how we arrived at SARS-CoV-2. *Curr Opin Virol* 47**,** 113-120. doi: 10.1016/j.coviro.2021.02.006.

Xiong, X., Qu, K., Ciazynska, K.A., Hosmillo, M., Carter, A.P., Ebrahimi, S., et al. (2020). A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat Struct Mol Biol* 27(10)**,** 934-941. doi: 10.1038/s41594-020-0478-5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., et al. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*. doi: 10.1016/j.cell.2020.09.032.

Zahradník, J., Tuekprakhon, A., Ginn, H.M., Duyvesteyn, H.M.E., Bahar, M., Khan, S., et al. (2021). Receptor binding and escape from Beta antibody responses drive Omicron-B.1.1.529 evolution. *bioRxiv*, 2021.2012.2003.471045. doi: 10.1101/2021.12.03.471045.

Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S.M., et al. (2021a). Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 372(6541), 525-530. doi: 10.1126/science.abf2303.

Zhang, J., Xiao, T., Cai, Y., Lavine, C.L., Peng, H., Zhu, H., et al. (2021b). Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science*, eabl9463. doi: 10.1126/science.abl9463.

Zhou, T., Tsybovsky, Y., Gorman, J., Rapp, M., Cerutti, G., Chuang, G.Y., et al. (2020). Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* 28(6), 867-879 e865. doi: 10.1016/j.chom.2020.11.004.

Figure Legends

Figure 1. A measure of pocket opening in the RBD. Receptor binding domains are aligned for representative locked (6zp2) and closed (6zp1) (Xiong et al., 2020) S protein trimers, and also for an S trimer structure carrying the D614G mutant, 7krq (Zhang et al., 2021a), with all RBDs down and monomer burial approaching the locked form. Distance across the pocket is shown schematically for 6zp2, with linoleic acid (LA) bound. The distance is calculated between the average of the 4 β-strand C_α atoms displayed (left of LA, which align well structurally between these RBDs), and the average of the two C_α atoms shown to the right of LA, present on a turn within a sub-structure that gates the LA binding pocket (Toelzer et al., 2020).

Figure 2. Variants carrying the D614G mutation do not cluster on a plot of pocket distance against S monomer burial within the S trimer. Clustering of locked, pH-locked, closed, and open forms in terms of monomer burial (Lobo and Warwicker, 2021) extends also to pocket distance. However, D614G set S proteins show greater variation, and the two Delta variant S proteins displayed are also well separated in monomer burial. Values are 3 monomer averages for each of the S protein trimers. Other than the open set, only trimers with all RBDs down are included (and thus sets are named D614Gset-closed and Delta-closed).

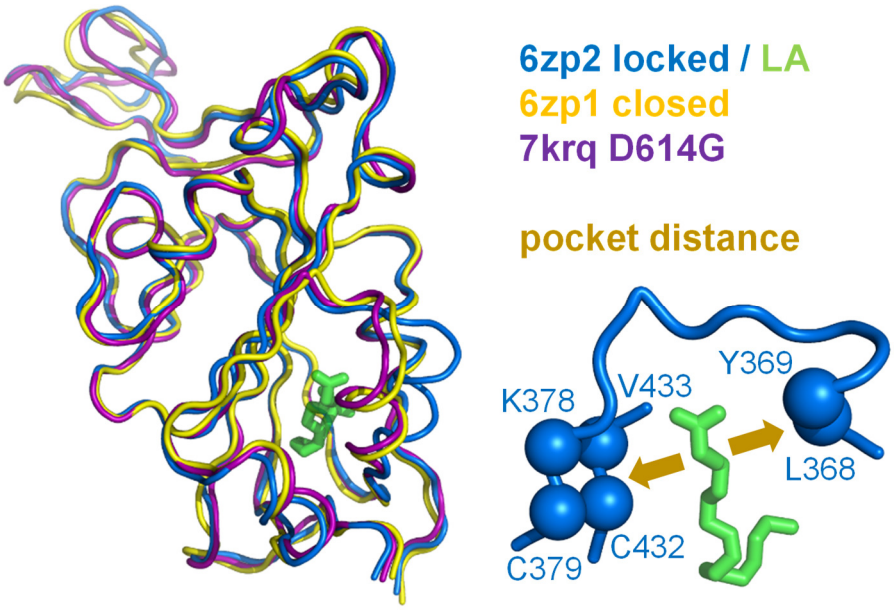
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3. Comparison of acidic pH and D614G mutation effects on packing within the S trimer. The d-SASA value (monomer burial within a trimer) is further differenced between the systems indicated, between 7krq (D614G, closed pocket but burial approaching the locked form) and the average over the locked set, between 7krq and the average over the closed set, and between 6xlu (pH-locked) and the average over the locked set. These double difference quantities are plotted cumulatively over the sequence of the S trimer, for which structure is available, with sub-domains (Berger and Schaffitzel, 2020) indicated (NTD N-terminal domain, SD1/SD2 subdomains 1 and 2, S1/S2 proteolytic cleavage site between subunits, S2' cleavage site within S2 subunit, FP fusion peptide, HR1 heptad repeat 1, CH central helix, CD connector domain). Regions and amino acids of particular interest are displayed with residue numbers and in relation to the changes in burial. A C-terminal helical coil interaction between monomers (present only in some structures) is labelled (hel).

Figure 4. Proline in the RBD of coronavirus HKU9, a site equivalent to S373 in the SARS-CoV-2 RBD, exhibits a kink in the mainchain. Regions around the LA binding site of SARS-CoV-2 S protein RBD are shown for locked (6zp2) and closed (6zp1) structures, with LA from 6zp2. The β-strand elements align well structurally, whereas the difference in pocket gating structures between pocket open/occupied (6zp2) and closed/unoccupied (6zp1) is clear in the turn region at the right hand side. A proline in the RBD of bat-derived coronavirus HKU9 (Huang et al., 2016), at the equivalent location to S373 of SARS-CoV-2 S protein, is indicated with a sphere at the C_α added to its tube mainchain. Amino acid sequences are shown for the contiguous turn-strand, with changes in the SARS-CoV-2 Omicron variant added.

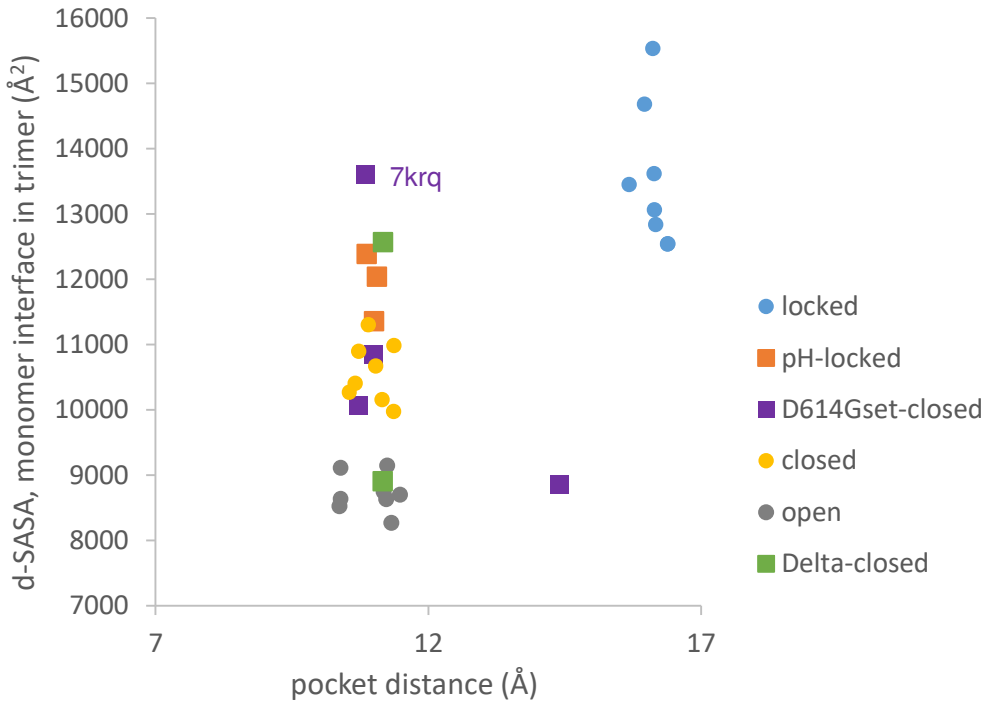
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1



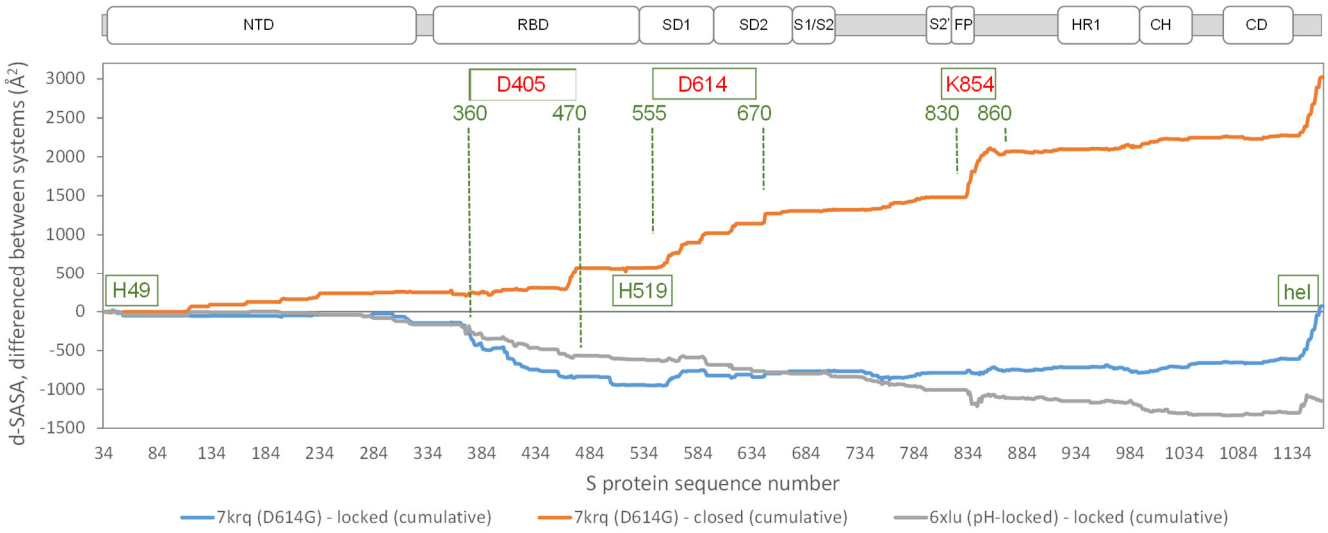
bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472554; this version posted December 16, 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 4.0 International license.

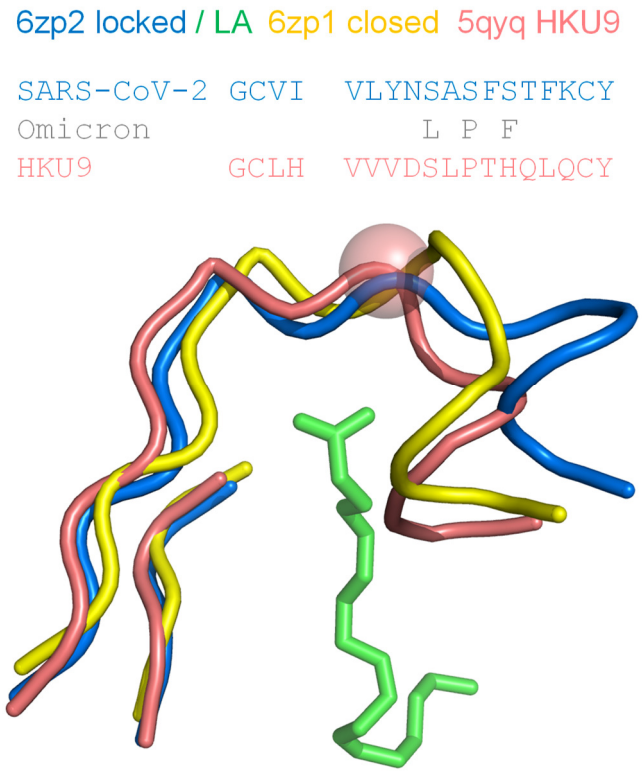



Figure 4