

## **Predicted pH-dependent stability of SARS-CoV-2 spike protein trimer from interfacial acidic groups**

Vanessa R. Lobo<sup>a</sup> and Jim Warwicker<sup>a,\*</sup>

<sup>a</sup> 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](mailto:jim.warwicker@manchester.ac.uk), tel. +44 (0)161 306 4490

## Abstract

Transition between receptor binding domain (RBD) up and down forms of the SARS-CoV-2 spike protein trimer is coupled to receptor binding and is one route by which variants can alter viral properties. It is becoming apparent that key roles in the transition are played by pH and a more compact closed form, termed locked. Calculations of pH-dependence are made for a large set of spike trimers, including locked form trimer structures that have recently become available. Several acidic sidechains become sufficiently buried in the locked form to give a predicted pH-dependence in the mild acidic range, with stabilisation of the locked form as pH reduces from 7.5 to 5, consistent with emerging characterisation by cryo-electron microscopy. The calculated pH effects in pre-fusion spike trimers are modulated mainly by aspartic acid residues, rather than the more familiar histidine role at mild acidic pH. These acidic sidechains are generally surface located and weakly interacting when not in a locked conformation. In this model, their replacement (perhaps with asparagine) would remove the pH-dependent destabilisation of locked spike trimer conformations, and increase their recovery at neutral pH. This would provide an alternative or supplement to the insertion of disulphide linkages for stabilising spike protein trimers, with potential relevance for vaccine design.

*Keywords:* SARS-CoV-2, spike protein, pH-dependence, protein electrostatics, coronaviruses

# 1. Introduction

The spike glycoprotein (S protein) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has received much attention, as the primary point of attachment between virus and ACE2 receptor on a cell surface [1], and as a target in vaccine design [2]. An original focus on the pre-fusion S protein trimer structure in closed and open states, with receptor binding domain (RBD) down and up, respectively, has expanded to include a further RBD down state, termed locked [3], first observed with in a structure with linoleic acid bound [4]. Here, the terms up and down are used to delineate RBD position, so that the open form maps to RBD up, but both closed and locked forms map to RBD down. This 3 state pre-fusion S trimer picture will be expanded with greater characterisation of post-fusion trimers [5], but the transitions between pre-fusion trimers are of key interest for the early steps in cell entry, and also for the export of fusion competent virus. Further, spike protein trimer expressed at the cell surface is able to mediate fusion between cells [6].

The role of pH has been studied for cell entry and exit of coronaviruses. Both plasma membrane fusion and endosomal mediated fusion pathways can be used [7], with fusion primed by S1/S2 spike protein cleavage during the biosynthetic exit pathway [8]. S1/S2 and S2/S2' cleavage sites, and the fusion peptide, are critical locations in considering spike trimer conformational transitions [9]. Most cryo-electron microscopy (cryo-EM) structures of spike trimers have been resolved at pH 7.5 – 8. Studies at mild acidic pH have found structural transitions [3, 10], interpreted as relatively favouring closed over locked forms as pH is increased from mild acidic to neutral [3]. Other cryo-EM analysis shows changes in the fraction of RBD up and down conformations between pH values of 6.5 and 8 [11]. Cell biological investigation of the exit pathway for newly synthesised SARs-CoV-2 virus has revealed a novel mechanism where lysosomes are repurposed for viral transit to the cell surface, with a moderate de-acidification of their mean pH from 4.7 to 5.7 [12].

Biophysical and structural analysis of the D614G S protein mutation that emerged in Spring 2020 indicates that it leads to greater availability of the RBD for ACE2 binding, through altered structural transitions [13]. These transitions include packing changes that resemble closed to locked form differences [14]. A model has been proposed in which the locked form is prevalent in the virus export pathway, giving protection of newly synthesised spike proteins at acidic pH, with the D614G mutation coupling to transition between locked and closed conformations [3]. In a complementary model, conformational change in the D614G structure

has been proposed to suppress release of S1 from cleaved S1/S2 spike trimers, thus enhancing availability for ACE2 binding [14].

Understanding the factors that stabilise locked, closed, and open S protein trimers is important for vaccine design as well as the pathways of viral infection and biosynthesis. Engineering to stabilise spike trimers [15] is important as knowledge of the most effective conformational targets for vaccine production become apparent [16]. Molecular dynamics simulations are a common tool for structure/function analyses, particular those involving structural transitions. They have been extensively applied to SARS-CoV-2 S protein [17], including studies of receptor binding [18], drug binding to the linoleic acid binding pocket [19], the opening of cryptic pockets as potential drug targets [20], temperature effects [21], and the effect of glycosylation [22].

Specifically for pH-dependent effects, several computational methods are available. Constant pH molecular dynamics has been applied to coronavirus protease active sites, reporting on the protonation states that may be best suited to inhibitor design [23]. Atomistic simulations in general are mostly low throughput in terms of the starting coordinate sets that are studied. At the commencement of the current study, about 130 structures were available for SARS-CoV-2 S protein trimers. Since these represent substantial variability, for example locked, closed or open forms, D614 or G614, neutral pH or acidic pH, high throughput computational methods have a role to play, in this case to increase understanding of pH-dependent properties. Based on earlier work [24], a web server for  $pK_a$  calculations has been benchmarked against known key sites for the acidic pH-dependence of influenza virus hemagglutinin [25]. This work was extended with a dataset of 24 spike trimers, collated in late Summer 2020 and mostly lacking what is now recognised as representatives of the locked form. A sparse set of groups predicted to couple with pH-dependence of closed to open form transition was found, focusing on the environment around a single salt-bridge [20].

The locked form is now more extensively characterised, with about 10 S protein trimers at the end of January 2021, as well as structures at mild acidic pH, and with the D614G mutation. Accordingly, the focus shifts to the transition between closed and locked forms in the current report, revealing a wider set of acidic groups (including D614) that are predicted to modulate stability at mild acidic pH. These groups are at interfaces that tighten in the locked form, relative to the closed form. Indeed, the change in burial of monomers within the S protein trimer is greater for the closed to locked transition, than for the closed to open transition. Predicted destabilisation in the locked form at neutral pH can be offset by either transition to

the closed form or reduction to acidic pH. Alternatively, mutation of these acidic groups is predicted to stabilise locked form trimers at neutral pH.

## 2. Methods

### 2.1 Structural data

A set of 129 SARS-CoV-2 spike trimer structures was retrieved from the RCSB/PDB [26] in January 2021, from a list of list generated by probing the SARS-CoV-2 spike sequence, UniProt [27] id P0DTC2, with the Basic Local Alignment Search Tool (BLAST) [28] into the RCSB/PDB at the National Center for Biotechnology Information, and subsequent checking for SARS-CoV-2 amongst the coronavirus spike proteins returned. Non-spike binding partners were removed, leaving the 129 trimers and 387 monomers (spike387). The monomers of spike387 are listed in Supplementary Table 1, along with assignment of RBD up or down according to PDB entry, and confirmed by visual inspection. Subsequent to analysis of spike387, sub-groups of locked, closed, and open S protein trimers were made, each trimer being uniformly RBD down (locked or closed), or RBD up (open). To maintain balance, 8 trimers were used for each sub-group, again with binding partners (ACE2, antibody fragments) removed. Further sub-groups were formed, consisting of: a single trimer with an engineered disulphide link (6zoz [29], termed diSlocked), 3 trimers at acidic pH [10] (pHlocked), and 8 trimers carrying the D614G mutation (D614Gset). Constituent monomers within trimers of the 6 sub-groups, along with RBD up or down designation, are given in Supplementary Table 2.

### 2.2 Sequence data

Sequences of spike proteins from the seven coronaviruses currently known to infect humans were aligned with the Clustal Omega [30] implementation at the European Bioinformatics Institute [31], and amino acid conservation relative to SARS-CoV-2 spike protein noted at particular sites. The seven S protein sequences included (with UniProt identifiers) were SARS-CoV-2 (P0DTC2), SARS (P59594), MERS (K9N5Q8), HCoV-HKU1 (Q5MQD0), HCoV-OC43 (P36334), HCoV-NL63 (Q6Q1S2), and HCoV-229E (P15423).

### 2.3 Structure-based calculations

Solvent accessible surface area (abbreviated to ASA) was calculated for S protein monomer burial within a trimer. For each monomer within a trimer (after non-spike components have been removed), ASA was calculated for the monomer alone and for the monomer within the trimer, the difference (d-ASA) giving the extent of burial. The contribution of polar and non-polar atoms to this burial was recorded, following previous methodology [32]. To facilitate comparison between sub-groups, d-ASA values were averaged over monomers within a sub-group. A key feature is that all calculations were referenced to an alignment of amino acid sequences for all monomers (spike387 and additional monomers added in subsequent formation of sub-groups). This ensures that all comparisons and heat maps are correctly aligned, allowing for engineered monomers and residues present in a sequence but disordered in a structure. Amino acid numbering throughout is given as the equivalent location in UniProt entry P0DTC2. Missing coordinates are not modelled, structure-based calculations use only the ordered residues reported in each structure. From vectors of 1 (ordered) and 0 (disordered) running along the sequence of each monomer, similarity to chain A of the locked conformation 6zb4 was calculated by vector dot product. Monomers were then listed in heat maps (of order/disorder or d-ASA) according to similarity with 6zb4A.

Electrostatics calculations used  $pK_a$  predictions following the Finite Difference Poisson Boltzmann (FDPB) / Debye-Hückel (DH) hybrid method, termed FD/DH [24]. Ionisable groups that are not buried can sample both FDPB and DH schemes, and these calculations are combined with Boltzmann weighting. Since the latter is a simple water-dominated and highly damped estimate of electrostatic interactions, non-buried groups will not have high  $pK_a$  deviations ( $\Delta pK_a$ s) from normal values, unless those are stabilising. On the other hand, buried groups, shielded from water-dominated interactions, can have  $\Delta pK_a$  values that relate to substantial destabilisation as well as stabilisation. The method has been implemented in a web tool, following testing with sites known to determine the pH-dependence of influenza virus hemagglutinin stability [25]. Here, the focus is on predicted stability change between pH 7.5 (extracellular, cytoplasmic pH) and the lower pH experienced in import and export pathways (pH 5 was used as the lower value). This pH-dependent energy can be derived using the relationship [33]  $\Delta G_{pH-dep} = \int (2.303RT\Delta Q)dpH$ , where  $\Delta G_{pH-dep}$  is the pH-dependent contribution to conformational stability (over the pH range of the integration, 7.5 to 5), differenced between two states. The states here are ionisable groups interacting in the FD/DH scheme, and the same set without interactions (null state with normal model compound  $pK_a$ s).

$\Delta Q$  is the charge difference between those two states,  $R$  is the Universal Gas constant, and  $T$  is set at 298.15 K. Since the null state is uniform between structures, results for each structure can be compared to assess differential responses to pH change in the mild acidic range. Groups responsible for the predicted effects can be identified since  $\Delta Q$  separates into component ionisations.

Molecular viewers used to analyse spike proteins were NGL [34], Swiss PDB Viewer [35], and PyMOL. For representative trimer structures in the locked (6zp2), closed (6zp1) and open (7a98) sub-groups, d-ASA values were transferred to the B factor column (capped at 99 Å<sup>2</sup>), facilitating visualisation of monomer burial within a trimer.

### 3. Results and Discussion

#### 3.1. *Differences in monomer burial between S protein trimer forms are more extensive than are order-disorder transitions*

Monomers within the spike387 set were classed as RBD up or down, noting that down does not distinguish between closed and locked trimers. Clustering of monomers was undertaken according to similarity of vectors representing order/disorder, and referenced to the A chain monomer within the 6zb4 locked trimer [4]. The resulting heat map of order/disorder is shown alongside a record of up or down RBD, and S protein domain structure (Fig. 1). Most clear-cut is ordering around position 840 in a set of structures with down RBDs and known to be in the locked form, at the top of the heat map. Other clusters of monomers, in terms of order/disorder, do not uniformly map to up or down RBD.

A similar procedure was used to map monomer burial within trimers. For each monomer, ASA loss upon integration of that monomer within the trimer was calculated (d-ASA). The order of monomers in the d-ASA heat map of Supplementary Fig. S1 is the same as that in Fig. 1. Summed d-ASA for each monomer is also shown. Known locked conformations mostly exhibit higher d-ASA than other conformations, including closed. The region around 840 in the d-ASA heat map differentiates the locked form again, but there is also a greater richness in differences across the sequence, particular for the RBD, reflecting a tightening of interfaces in the locked form relative to closed form. This analysis gives a view of detailed changes that complement molecular graphics comparison of locked and closed RBD down structures.

### 3.2 *Relatively polar interfacial interactions differentiate locked from closed trimers*

In order to more easily compare different conformational forms of spike trimer, sub-groups from spike387 were made with either all RBD down or all RBD up monomers. Three sub-groups (locked, closed, open) were matched numerically at 8 trimers (24 monomers) each. Other sub-groups were added: a single locked conformation trimer with engineered cysteines creating a disulphide link [29] (diSlocked), 8 structures with the D614G mutation (D614Gset), and 3 structures at acidic pH values (pHlocked). For D614Gset, trimers have various combinations of up and down RBD containing monomers. Summed d-ASA values show clear distinction of locked and diSlocked forms from closed, and open forms (Fig. 2a). The change of monomer burial within a trimer from locked to closed is greater than that between closed and open, emphasising the interfacial tightening of the locked form. These changes are widespread and not solely located in regions that undergo disorder to order transitions (Fig. 1). Monomers of the pHlocked set have d-ASA intermediate between locked and closed trimers, whilst D614Gset monomer d-ASA values span a much greater range than other sub-groups.

The ratio of non-polar to polar contributions to d-ASA was calculated for the S1 and S2 subunits of each monomer (Fig. 2b). For external protein surfaces not involved in protein-protein interactions the most common value of this ratio, in a distribution over patches, is about 1.2, increasing for more hydrophobic, interacting surfaces [32]. For the open sub-group, with least monomer burial, non-polarity increases for both S1 and S2 contributions. Conversely, monomer interface within a trimer becomes successively more polar for sub-groups that have higher d-ASA, with polarity for S1 d-ASA in the locked form close to that seen for non-interfacial surface (Fig. 2b). It can be concluded that transition from RBD up to RBD down, and further transition to the locked form, involves burial of surface that is not especially hydrophobic.

### 3.3 *Acidic sidechains at tightened interfaces are predicted to have elevated $pK_a$ s*

Comparison of Fig. 3a and Fig. 3b reinforces the observation that d-ASA differences between S trimer sub-groups are more extensively spread over spike monomers than are order/disorder differences. The extent of d-ASA change between locked and closed forms, in comparison with that between closed and open forms, is also emphasised. In light of reported pH-dependent transition between S trimer conformations [3, 10],  $pK_a$  predictions were made



for ionisable residues in trimer sub-groups. To directly address changes in the mild acidic pH range, the predicted contribution to change in conformational stability between pH 7.5 (cytoplasm, extracellular) and pH 5 (endosome) was calculated (summed and for each ionisable group). The results are added to Fig. 3 as a heat map of calculated per-residue pH-dependence, averaged over monomers within each sub-group (Fig. 3c). Towards the C-terminal part of S2 are several histidines [25, 36], for which burial and predicted resistance to protonation from pH 7.5 to 5 leads to destabilisation in all sub-groups. Of more interest for differences between sub-groups are amino acids with acidic sidechains that have elevated  $pK_a$ s and are predicted to be stabilising from pH 7.5 to 5, since destabilisation is reduced. Destabilisation derives from a local environment that suppresses ionisation at a pH above the normal  $pK_a$ . Several Asp and Glu sidechains are of interest, particularly those for which the effect is exhibited in locked forms but not others, as a consequence of the increased burial.

### *3.4 Predicted stabilisation of locked forms, and destabilisation of closed and open forms, from pH 7.5 to pH 5*

Summed values of predicted pH-dependence give a stabilisation of locked form sub-groups as pH falls from 7.5 to 5, but a destabilisation of other sub-groups over the same range (Fig. 4). The consequence of these predictions is a model in which the locked form is preferentially stabilised in low pH secretory pathway vesicles, whereas closed (unlocked) and open forms are stabilised at extracellular, neutral pH, as suggested from structural studies [3].

Looking at individual group contributions, 3 His in S2 (H1048, H1064, H1088) are buried and largely inaccessible to solvent throughout the sub-groups, so that they give a uniform destabilisation over the pH 7.5 to 5 range, as each His transits through the  $pK_a$  (6.3) that would normally mark ionisation. These groups are not predicted to contribute differentially to pH-dependence of locked, closed, and open forms. Next, D398 was proposed to couple to pH-dependence of closed (mostly unlocked) to open forms [36]. This result is replicated in the current larger study (Fig. 4), with a predicted D398 contribution to variation of pH-dependence between locked, closed, and open forms, but in the opposite sense to the overall effect. Two histidines have been suggested as potential mediators of locked trimer stabilisation at acidic pH [3], but no clear differential contributions of H49 and H519 are seen between sub-groups of structures in the current study (Fig. 4).

### 3.5 *D614G mutation and structural preferences*

A feature of predicted pH-dependence for which the picture changes when locked (RBD down) structures are separated from closed (RBD down), is the influence of D614. Contributions of D614 to the calculated pH-dependence are seen for locked and pHlocked sub-groups, in a sense to stabilise those forms at pH 5 relative to pH 7.5 (Fig. 4). It has been reported that the S protein D614G mutation increases virus infectivity [37], and suggested that the mutation leads to a greater population of up relative to down RBDs due to loss of an inter-monomer latch formed by the sidechain of D614 [13]. Recent structural studies with an S protein ectodomain engineered with D614G, report that the increase in infectivity results from stabilisation of the S trimer against dissociation [5]. One of the structures (7krq) from that D614G study has all RBDs down, and d-ASA for monomers in this structure (Supplementary Fig. S2) are similar to those seen for locked forms (Fig. 2a), consistent with trimer stabilisation. Generally, the D614Gset sub-group exhibits a wide distribution of monomer d-ASA values, but these structures mostly have mixed up and down RBD within each trimer (Supplementary Fig. S2).

D614 is predicted to be relatively destabilising at neutral pH in locked and pHlocked forms (Fig. 3c, Fig. 4). Even after incorporating stabilising D614 sidechain interaction with K854, the reduction in solvent exposure due to ordering of the loop around residue 840 leads to a predicted average destabilisation, at pH 7.5 compared to pH 5 and averaged over the locked sub-group, of 7.7 kJ/mol. This implies a mean D614  $pK_a$  shift to about 6.0, from the intrinsic value of 4.0, consistent with wide-spread D614 - K854 salt-bridge formation in locked structures at pH 7.5 or 8. There are examples of this salt-bridge in all sub-groups (other than D614Gset), although at lower incidence outside of the locked sub-group. Our interpretation is that D614 - K854 interaction is solvent exposed and relatively weak outside of locked forms. Then, ordering of the loop around residue 840 in the locked form decreases solvent exposure, putting a greater onus on salt-bridge formation to reduce destabilisation of D614. Summed over interactions though, destabilisation of D614 is predicted to result for the locked form, and will be greater at pH 7.5 than at pH 5. This model is consistent with stabilisation of the D614G S trimer against dissociation at neutral pH [38]. Referring to the 3 structures in the pHlocked sub-group, for two (6xlu, pH 4.0 and 7jwy, pH 4.5) the 840 loop is ordered. In both of these, K854 is outside of salt-bridge range, consistent with elevated  $pK_a$  and protonation of D614 at acidic pH.

### 3.6 *Predicted locations of pH-dependence flank regions of known structural sensitivity*

In common with D614, other predicted sites of pH-dependence are coupled to a tightening of monomer-monomer interfaces in the locked form S trimer (Fig. 5a), so that solvent exposure is reduced and Asp/Glu sidechain  $pK_a$ s are elevated, with predicted destabilisation of the locked form as pH increases through the mild acidic pH range. Presumably, pH-independent features, such as linoleic acid binding [4], also couple to the energetic balance between locked and closed forms. For the pH-dependent component, it is notable that a tendency towards a locked form at acidic pH is still evident for the D614G S protein trimer [3], consistent with the existence of other pH-dependent sites (Fig. 3c, Fig. 5a). Whereas the D614 - K854 interaction is between monomers, neighbouring basic residues for other acidic sites of predicted pH-dependence, where they exist, tend to be from the same monomer. Nevertheless, the predicted effect is the same, it is the solvent exposure of the acidic group rather than the details of salt-bridging that is the key determinant of predicted pH-dependence. Several of the implicated acidic residues are present in only SARS-CoV-2 and SARS, of coronaviruses known to infect humans (Fig. 5b).

Display of d-ASA on S protein structure shows the expected difference between open (Fig. 6a) and closed (Fig. 6b) forms, and emphasises the difference (that is not evident by overall RBD location) between closed (Fig. 6b) and locked (Fig. 6c) forms. General location (Fig. 6d) and detail (Fig. 6e) are shown for residues that are present in the more compact interfaces of the locked form trimer, and which are predicted to contribute to pH-dependence over the mild acidic range. Some residues (D586, D614, D830, D843) flank regions of known importance for structural transitions (S1/S2 and S2/S2' cleavage sites, fusion peptide), and other residues (D405, D420, E465, D979, D994) are adjacent to a site commonly engineered to generate more stable trimers [3]. Additionally, sites of predicted pH-dependence in the RBD are proximal to the linoleic acid binding site [4]. It is therefore suggested that members of this set of acidic residues (Fig. 5b, Fig. 6e), beyond D614, couple pH-dependence to conformational biases of functional relevance.

### 3.7 *pH-dependence generated through acidic groups at interfaces*

In this model for pH-dependence of locked to closed form transition, a surface acidic group, whether or not paired with a basic group, loses substantial ASA on formation or tightening of an interface. The acidic  $pK_a$  increases in the range pH 4 to 7. A bias against

interface formation is introduced, which is relieved as pH reduced. Depending on the conformational details and interactions with neighbouring groups, a buried acidic group could be stabilising, but that is not predicted here.

Although functional ionisation of carboxylate sidechain groups at mild acidic pH is most well-known for acid-base catalytic systems, such as the elevated  $pK_a$  of E35 in hen egg white lysozyme [39], many non-catalytic occurrences have been discussed. In viruses, coiled-coil formation in a variant influenza hemagglutinin is pH-dependent between pH 4.5 and 7.1, possibly mediated by E69 and E74 [40]. A model for pH-dependent association of some antigenic peptides and class II proteins of the major histocompatibility complex suggests that buried interfacial Asp or Glu residues mediate this process, with  $pK_a$ s higher than 7 [41]. Interfacial residues D612 and E613 of PsaB in photosystem I are proposed to be responsible for mild acidic pH-dependence of electron transfer in complexes with plastocyanin or cytochrome  $c_6$  [42]. The periplasmic chaperone HdeA of *Escherichia coli* acts on a shift from neutral to lower pH *via* dimer dissociation and partial unfolding, these changes mediated by two aspartic acid residues [43]. In a further example, aspartic acid residues have been proposed as determining neutral to low pH conformational variation in an antigen-binding region of a therapeutic monoclonal antibody [44].

Given the background of acidic sidechains mediating mild acidic pH-dependent effects, it is reasonable to investigate similar interactions when considering the pH-dependence of spike protein conformation, supplementing the assessment of histidine protonation in regard to pH-induced changes in viruses [45].

## 4. Conclusions

Analysis of spike trimer structures shows the expected large ASA change, for monomer burial, between closed and open forms, and an even larger change and interface tightening between closed and locked forms. A subset of acidic groups was identified in the locked form monomer-monomer interfaces that are predicted to mediate pH-dependent stability at mild acidic pH. These are located in regions that flank sites of known importance for stability engineering and/or function, such as cleavage sites and fusion peptide. D614 is amongst the groups identified and, in common with other acidic groups in the subset, is predicted to destabilise the locked form at neutral pH. This result is consistent with a report that the G614 mutation can exist in a form that resembles locked and stabilises the trimer at neutral pH, leading to suppression of S1 subunit loss and increased RBD – ACE2 binding [38].

The current work is also consistent with a model for pH-dependent conformational masking of RBDs within the S protein trimer, derived from the 3 structures of the pHlocked set [10]. That study highlighted D614 and D586 (in common with our calculations), and also D574. If the prediction of a wider set of acidic groups relevant for pH-dependence is correct, then it might be asked why, from this set, only mutation at D614 has been seen widely in SARS-CoV-2 genomes. Of note though, a study of recurring spike protein mutations highlights several regions that overlap with, or are adjacent to, segments identified in Fig. 6e [46]. These mutations could influence folding and packing around the sites predicted to mediate pH-dependence. Since pH modulation of the D614G spike trimer conformation has been reported [3], there must be groups beyond D614 responsible for some degree of pH-coupling.

The recently discovered lysosomal exit pathway for newly synthesised SARs-CoV-2 virus provides a mild acidic environment, even including the observed moderate de-acidification [12]. Reduced acidification is thought to reduce activity of lysosomal proteases, thereby protecting exposed viral proteins from proteolysis, beyond the S1/S2 priming cleavage of S protein that confers an advantage for SARS-CoV-2 infection [47]. Stabilising the locked form at acidic pH is proposed to contribute to protection of the spike protein along the secretory pathway [3]. The current results are consistent with this model, extending the set of groups that may be responsible for pH-dependent stabilisation (beyond D614). Importantly, the balance between priming cleavage and stabilisation along the biosynthetic pathway is the focus for ongoing viral adaptation, including the observation that P681R mutation in the B.1.617 lineage optimises the furin cleavage site and could enhance transmissibility and pathogenicity [48].

An implication of the current work is that it should be possible to engineer more stable locked forms at neutral pH through mutation of the acidic groups, (e.g. Asp to Asn), at sites indicated in Fig. 6e. In line with the observation of close packing around G614, Asp to Gly changes could be tested, although the D614N mutation also leads to a more compact spike trimer [49]. Such analysis would probe the balance of stabilising and destabilising contributions to spike trimer structure during trafficking, and it could provide further insight into production of specific spike protein conformational forms for vaccines.

## Funding

Support from the UK Engineering and Physical Sciences Research Council (award EP/N024796/1) is gratefully acknowledged.

## Acknowledgements

The authors thank Lorena Zuzic for discussion, and the computational shared facility at the University of Manchester for support and resources.

## References

- [1] Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, NY)* 2020;367(6485):1444-8.
- [2] Kumar A, Dowling WE, Roman RG, Chaudhari A, Gurry C, Le TT, et al. Status Report on COVID-19 Vaccines Development. *Curr Infect Dis Rep* 2021;23(6):9.
- [3] Qu K, Xiong X, Ciazynska KA, Carter AP, Briggs JAG. Structures and function of locked conformations of SARS-CoV-2 spike. *bioRxiv* 2021.
- [4] Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, Kavanagh Williamson M, et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science (New York, NY)* 2020.
- [5] Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, Jr., et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science (New York, NY)* 2020;369(6511):1586-92.
- [6] Buchrieser J, Dufloo J, Hubert M, Monel B, Planas D, Rajah MM, et al. Syncytia formation by SARS-CoV-2-infected cells. *EMBO J* 2020;39(23):e106267.
- [7] V'Kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 2021;19(3):155-70.
- [8] Whittaker GR, Daniel S, Millet JK. Coronavirus entry: how we arrived at SARS-CoV-2. *Curr Opin Virol* 2021;47:113-20.
- [9] Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 2020.
- [10] Zhou T, Tsybovsky Y, Gorman J, Rapp M, Cerutti G, Chuang GY, et al. 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* 2020;28(6):867-79 e5.
- [11] Pramanick I, Sengupta N, Mishra S, Pandey S, Girish N, Das A, et al. Conformational flexibility and structural variability of SARS-CoV2 S protein. *Structure* 2021.
- [12] Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, Fisher M, et al. beta-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* 2020;183(6):1520-35 e14.
- [13] Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 2020.
- [14] Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science (New York, NY)* 2021;372(6541):525-30.
- [15] Riley TP, Chou HT, Hu R, Bzymek KP, Correia AR, Partin AC, et al. Enhancing the Prefusion Conformational Stability of SARS-CoV-2 Spike Protein Through Structure-Guided Design. *Frontiers in immunology* 2021;12:660198.
- [16] Gavor E, Choong YK, Er SY, Sivaraman H, Sivaraman J. Structural Basis of SARS-CoV-2 and SARS-CoV Antibody Interactions. *Trends Immunol* 2020;41(11):1006-22.
- [17] Amaro RE, Mulholland AJ. Biomolecular Simulations in the Time of COVID19, and After. *Comput Sci Eng* 2020;22(6):30-6.



- [18] Casalino L, Dommer A, Gaieb Z, Barros EP, Sztain T, Ahn SH, et al. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *bioRxiv* 2020.
- [19] Shoemark DK, Colenso CK, Toelzer C, Gupta K, Sessions RB, Davidson AD, 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* 2021;60(13):7098-110.
- [20] Zuzic L, Samsudin F, Shivgan AT, Raghuvamsi PV, Marzinek JK, Boags A, et al. Uncovering cryptic pockets in the SARS-CoV-2 spike glycoprotein. *bioRxiv* 2021.
- [21] Marti D, Torras J, Bertran O, Turon P, Aleman C. Temperature effect on the SARS-CoV-2: A molecular dynamics study of the spike homotrimeric glycoprotein. *Comput Struct Biotechnol J* 2021;19:1848-62.
- [22] Allen JD, Chawla H, Samsudin F, Zuzic L, Shivgan AT, Watanabe Y, et al. Site-specific steric control of SARS-CoV-2 spike glycosylation. *bioRxiv* 2021.
- [23] Verma N, Henderson JA, Shen J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *Journal of the American Chemical Society* 2020;142(52):21883-90.
- [24] Warwicker J. Improved pKa calculations through flexibility based sampling of a water-dominated interaction scheme. *Protein Sci* 2004;13(10):2793-805.
- [25] Hebditch M, Warwicker J. protein-sol pKa: prediction of electrostatic frustration, with application to coronaviruses. *Bioinformatics* 2020.
- [26] Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 2007;35(Database issue):D301-3.
- [27] UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47(D1):D506-D15.
- [28] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-10.
- [29] Xiong X, Qu K, Ciazynska KA, Hosmillo M, Carter AP, Ebrahimi S, et al. A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat Struct Mol Biol* 2020;27(10):934-41.
- [30] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 2011;7:539.
- [31] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 2019;47(W1):W636-W41.
- [32] Hebditch M, Warwicker J. Web-based display of protein surface and pH-dependent properties for assessing the developability of biotherapeutics. *Sci Rep* 2019;9(1):1969.
- [33] Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. *J Mol Biol* 1994;238(3):415-36.
- [34] Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL Viewer: Web-based molecular graphics for large complexes. *Bioinformatics* 2018.
- [35] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;18(15):2714-23.
- [36] Warwicker J. A model for pH coupling of the SARS-CoV-2 spike protein open/closed equilibrium. *Brief Bioinform* 2021;22(2):1499-507.
- [37] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 2020;182(4):812-27 e19.
- [38] Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *bioRxiv* 2020.
- [39] Bartik K, Redfield C, Dobson CM. Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional <sup>1</sup>H NMR. *Biophys J* 1994;66(4):1180-4.

- [40] Higgins CD, Malashkevich VN, Almo SC, Lai JR. Influence of a heptad repeat stutter on the pH-dependent conformational behavior of the central coiled-coil from influenza hemagglutinin HA2. *Proteins* 2014;82(9):2220-8.
- [41] Belmares MP, Rabinowitz JD, Liu W, Mellins ED, McConnell HM. pH stability of HLA-DR4 complexes with antigenic peptides. *Biochemistry* 2000;39(47):14558-66.
- [42] Kuhlert S, Drepper F, Fufezan C, Sommer F, Hippler M. Residues PsaB Asp612 and PsaB Glu613 of photosystem I confer pH-dependent binding of plastocyanin and cytochrome c(6). *Biochemistry* 2012;51(37):7297-303.
- [43] Foit L, George JS, Zhang BW, Brooks CL, 3rd, Bardwell JC. Chaperone activation by unfolding. *Proc Natl Acad Sci U S A* 2013;110(14):E1254-62.
- [44] Lan W, Valente JJ, Iltott A, Chennamsetty N, Liu Z, Rizzo JM, et al. Investigation of anomalous charge variant profile reveals discrete pH-dependent conformations and conformation-dependent charge states within the CDR3 loop of a therapeutic mAb. *MAbs* 2020;12(1):1763138.
- [45] Kampmann T, Mueller DS, Mark AE, Young PR, Kobe B. The Role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* 2006;14(10):1481-7.
- [46] Zahradník J, Nunvar J, Schreiber G. SARS-CoV-2 convergent evolution as a guide to explore adaptive advantage. *bioRxiv* 2021:2021.05.24.445534.
- [47] Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* 2021.
- [48] Peacock TP, Sheppard CM, Brown JC, Goonawardane N, Zhou J, Whiteley M, et al. The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. *bioRxiv* 2021:2021.05.28.446163.
- [49] Juraszek J, Rutten L, Blokland S, Bouchier P, Voorzaat R, Ritschel T, et al. Stabilizing the closed SARS-CoV-2 spike trimer. *Nat Commun* 2021;12(1):244.



## Figure Legends

**Fig. 1.** Spike protein monomers clustered according to similarity of ordered/disordered regions. The spike387 set is listed vertically according to similarity with chain A of 6zb4 (top). Order (yellow) and disorder (blue) are plotted. A smaller vertical heat map indicates monomers with down (green) or up (white) RBD positions. A group of locked conformations that cluster with 6zb4 monomers is indicated. Amino acid sequence (horizontal) covers the extent of ordered regions in spike ectodomain structures, with specific regions of disorder noted along the top of the heat map. Domain structure of the spike protein is also shown, in register with the heat map.

**Fig. 2.** Buried ASA of monomers within spike trimer, compared between sub-groups. (a) A scatter plot showing the ratio of non-polar to polar buried d-ASA for the S1 and S2 subunits, upon monomer burial in the trimer. Variation of this ratio is much greater for S1 than for S2 (seen with equal scaling of the axes), with higher values for the open form indicating that residual buried surface with RBD up is relatively non-polar. (b) Box and whisker plot for d-ASA distributions within each sub-group.

**Fig. 3.** Acidic groups with increased burial in the locked form are predicted to give a mild acidic pH-dependence. (a) Order/disorder is illustrated with stacked histograms (1/0) for the 6 sub-groups of spike trimers. Differences of d-ASA between sub-groups are more extensive than are those of order/disorder, illustrated with cumulative plots of the closed form subtracted from the 5 other sub-groups (b). The magnitude of locked to closed difference is larger than that of open to closed. (c) A heat map of predicted conformational stability change from pH 7.5 to 5, for each sub-group (labelled), with stabilising (red) or destabilising (blue) indicated. Amino acids that differ the most between sub-groups are indicated at the bottom of the plot, together with additional grey vertical bars aligned with the heat map. Panels (a) and (b) are in register. Panel (c), with only ionisable groups, is approximately in register with the full sequence plots.

**Fig. 4.** Predicted pH-dependent stabilisation for different conformational forms. Calculated stabilisation or destabilisation (pH 7.5 to 5) is shown for the average over each sub-group of: all ionisable groups, D398, H1048 + H1064 + H1088 of S2, H49 + H519, D614.

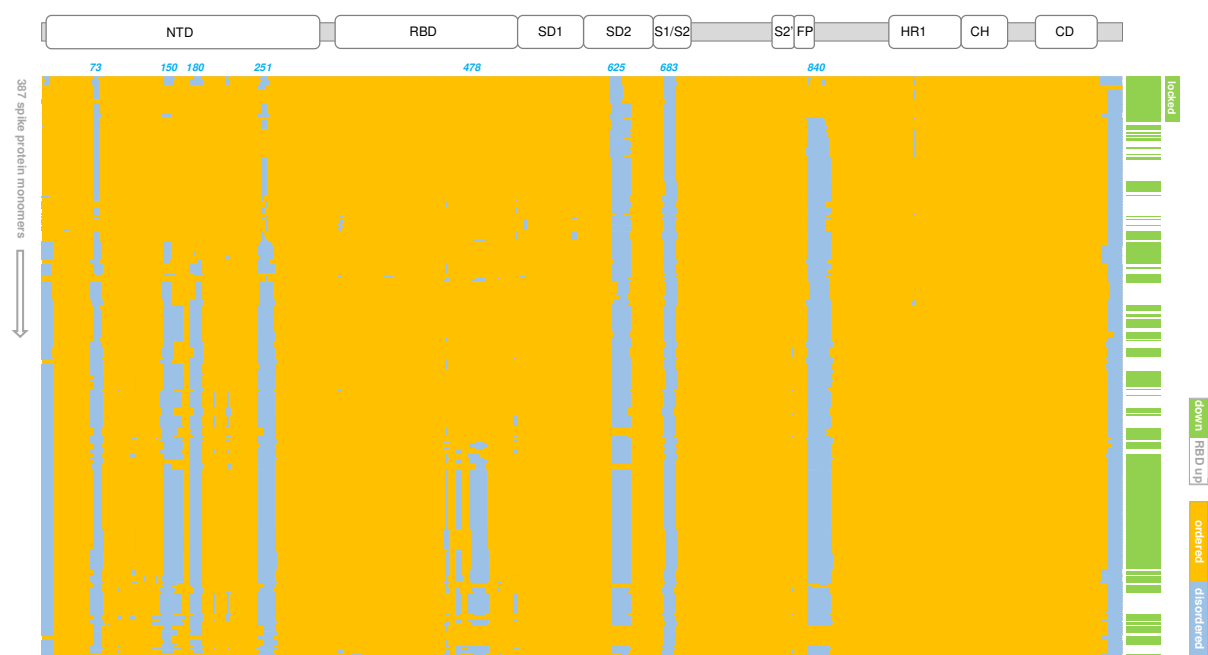
**Fig. 5.** Individual residue contributions to locked / closed form differences. (a) For amino acids highlighted in Fig. 3c, both d-ASA and predicted pH 7.5 to 5 stability change are plotted as differences (locked – closed). (b) The presence of these residues in a set of coronaviruses that infect humans (other than SARS-CoV-2) is shown with a stacked histogram. For example, E132 is present only SARS-CoV-2, whilst D994 is also present in the other 6 viruses.

**Fig. 6.** Location of sites predicted to differentiate the mild acidic pH-dependence of locked and closed forms. Tube plots of spike protein backbone are shown with colour-coding from d-ASA of 0 Å<sup>2</sup> (yellow) to 80 Å<sup>2</sup> (red) for open/7a98 [9] (a), closed/6zp1 [29] (b), and locked/6zp2 [29] (c) forms. Within the regions of increased d-ASA in locked relative to closed forms are amino acids with acidic sidechains, shown in a coarse view with magenta tube plot (one of 3 symmetry-related representatives) on a background of spike trimer S1 (grey) and S2 (orange) subunits (d). This region is drawn in greater detail (e) with backbone tube and amino acid stick plots, noting that this one of the 3 representatives consists of regions from all 3 monomers (colour-coded tube). Sites of interest with regard to conformational change and stability within the trimer are also indicated.

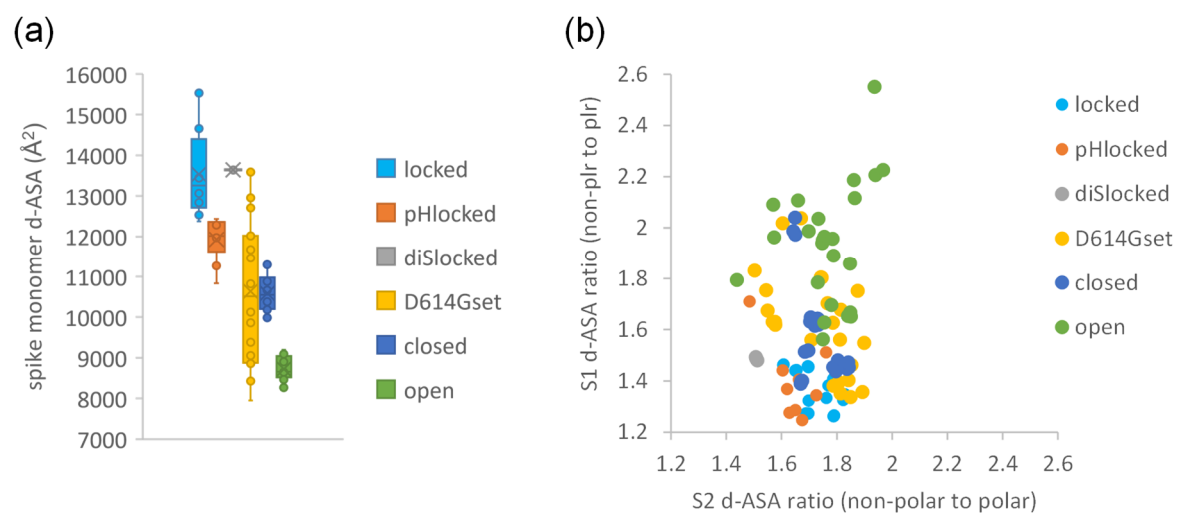
**Supp. Fig. 1.** Spike protein monomer burial within a trimer. Monomers are listed vertically according to the clustering derived for Fig. 1 (6zb4A and a group of locked forms at the top), and with sequence variation horizontally, referenced to the domain structure. A heat map of burial ASA (d-ASA) for monomer within trimer is shown, from zero (white) with increasing (brown) colour, capped at d-ASA = 80 Å<sup>2</sup>. A smaller vertical heat map gives d-ASA per monomer (white, smaller to purple, greater).

**Supp. Fig. 2.** Monomer burial within trimer for members of D614Gset. Variation in both S1 and S2 subunit contributions to d-ASA contribute to the large overall range of d-ASA values exhibited within D614Gset. RBD up or down monomer conformations are indicated.

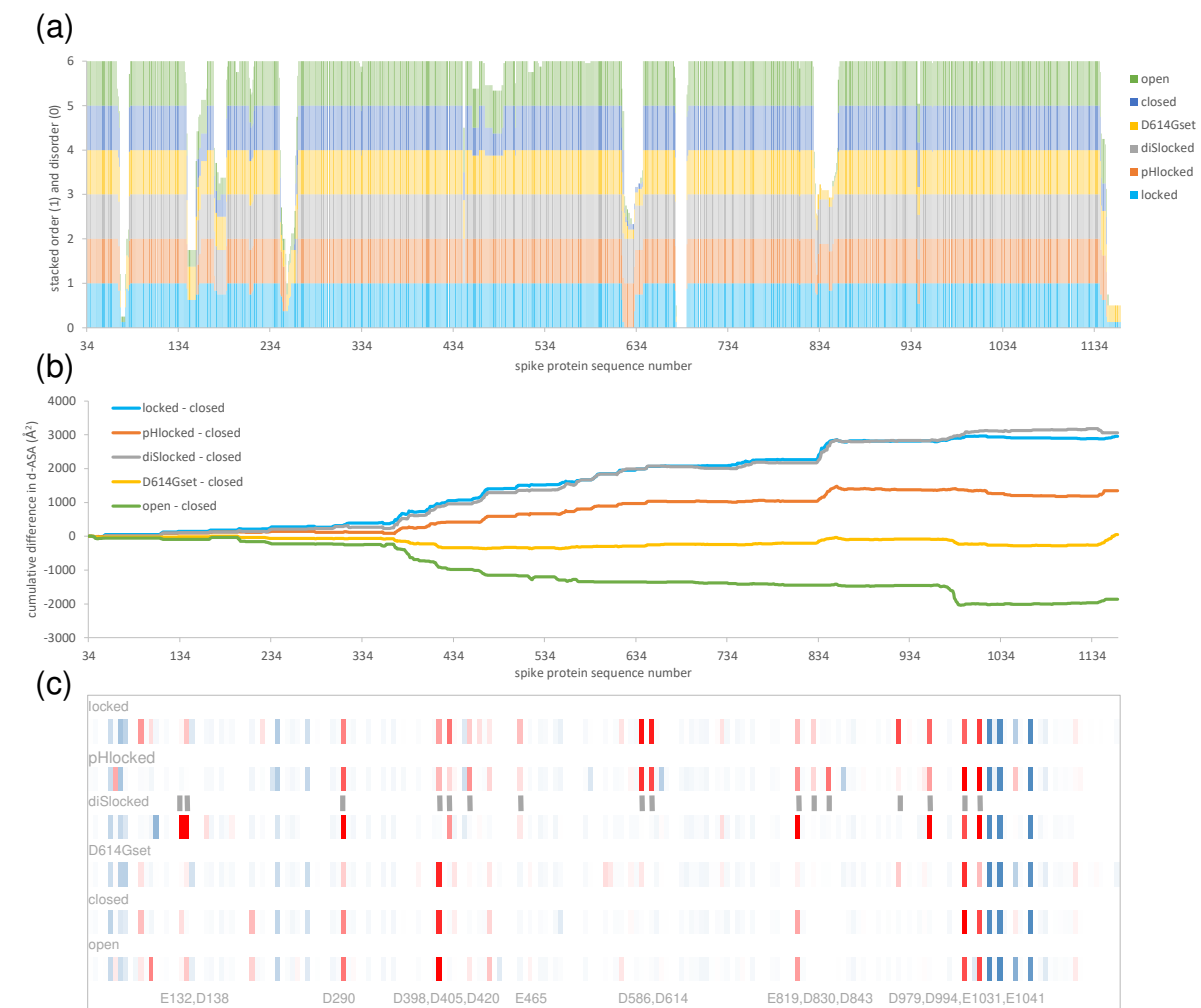
**Fig. 1**



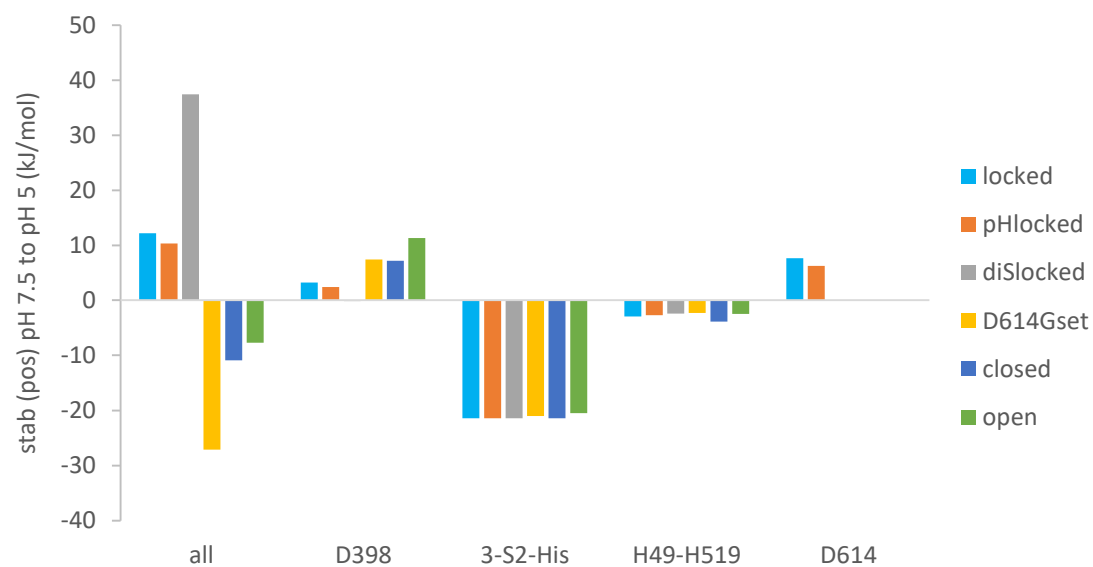
**Fig. 2**



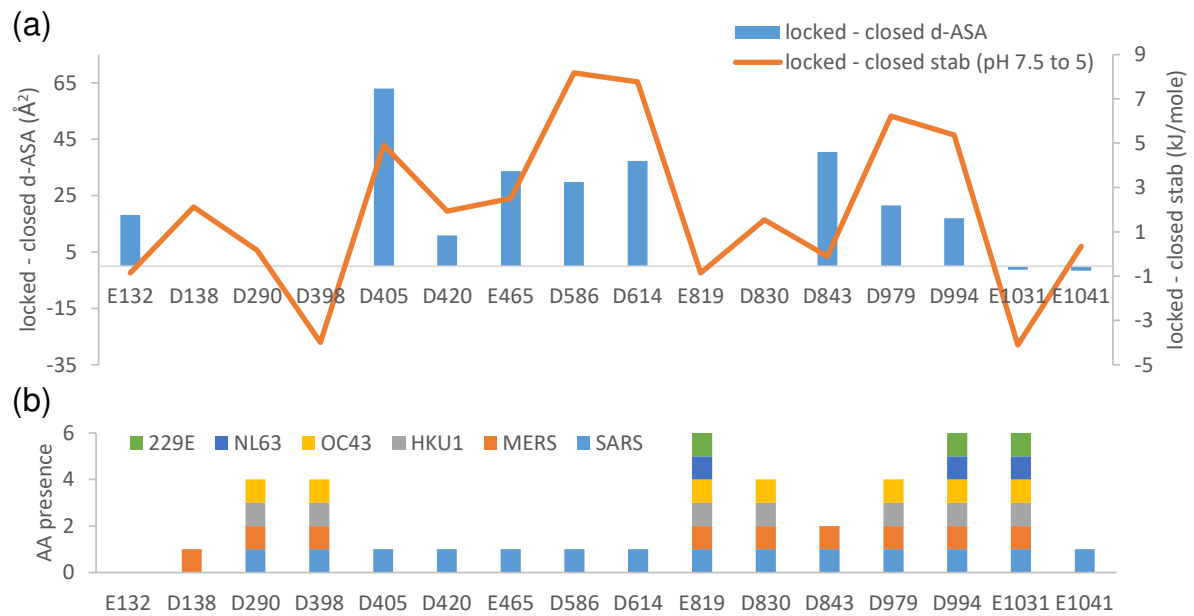
**Fig. 3**



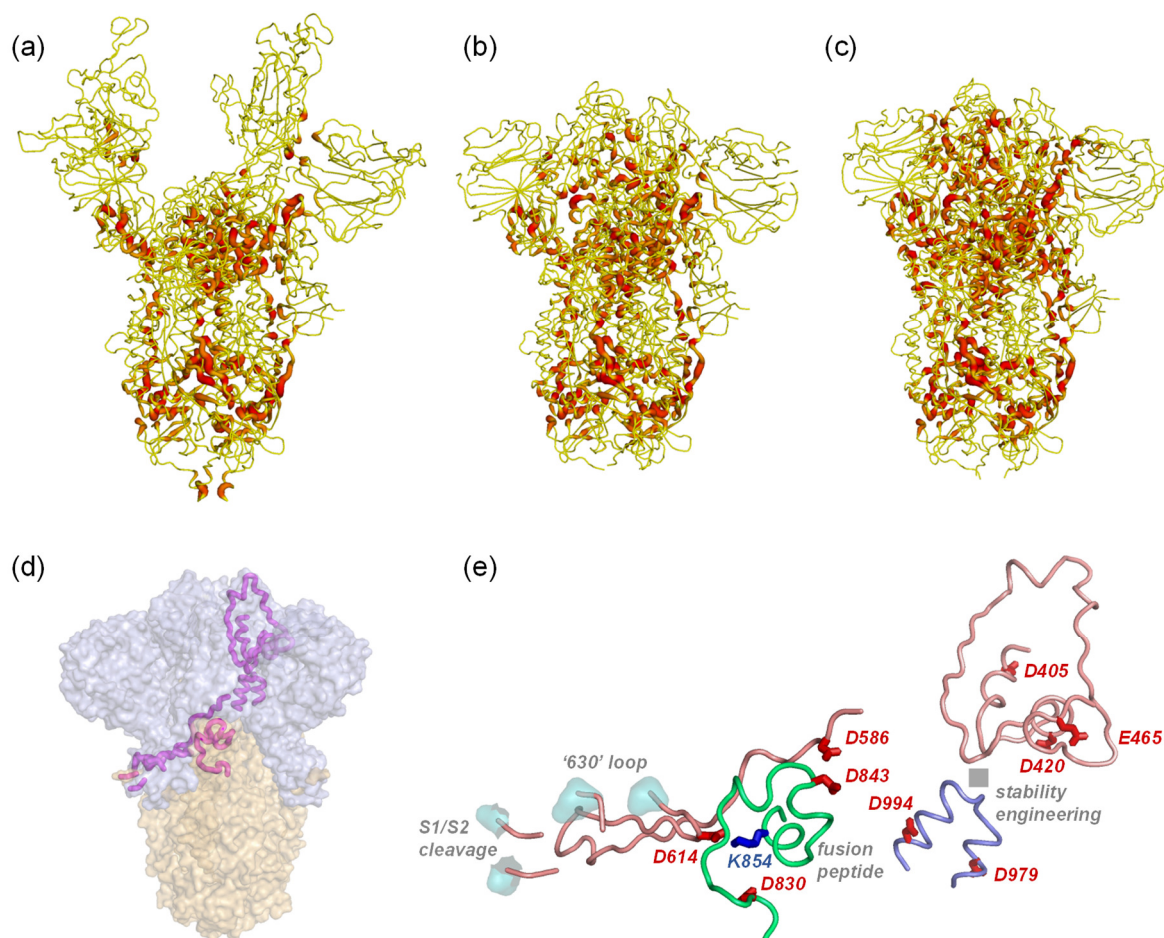
**Fig. 4**



**Fig. 5**

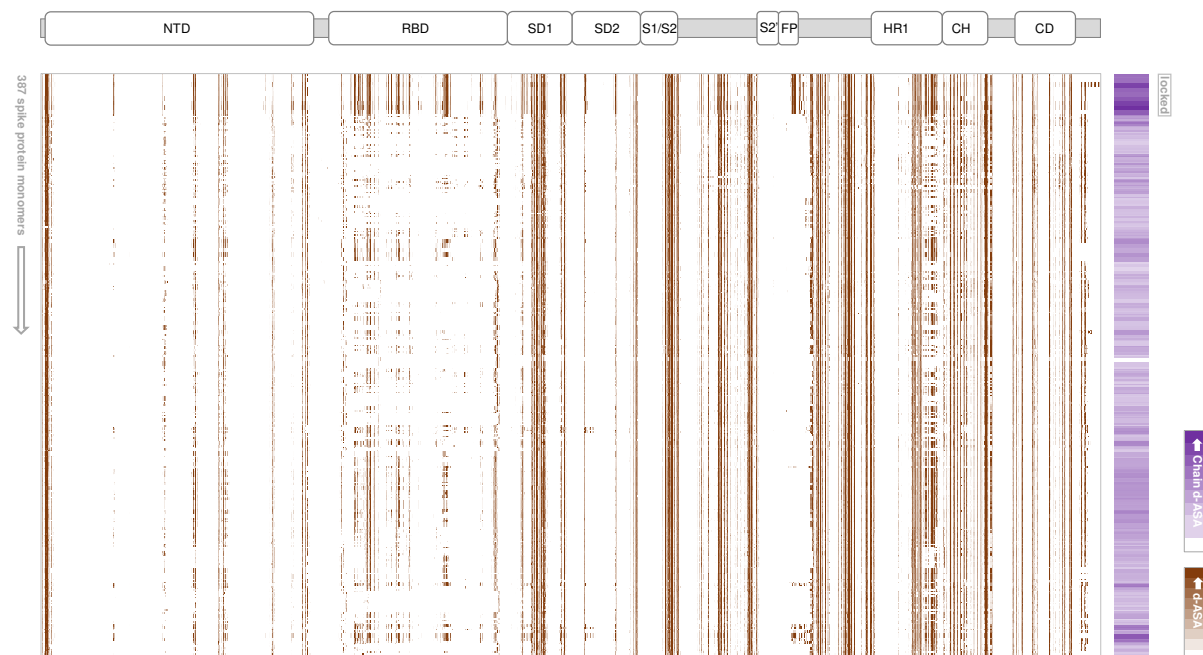


**Fig. 6**

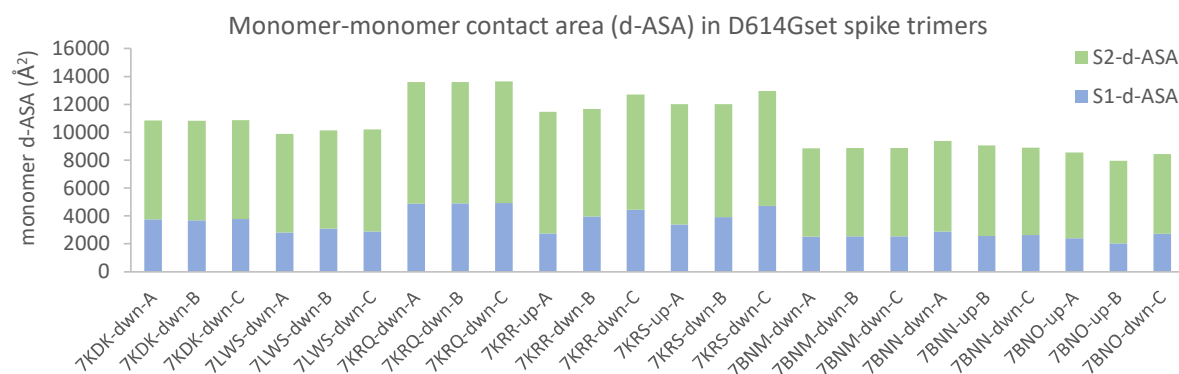




## Supplementary Fig. S1



## Supplementary Fig. S2



**Supplementary Table 1.** Monomer PDB/chain identifiers are given in pairs with down or up states for the RBD. These 387 monomers from 129 spike protein trimers form the spike387 set. The order of monomers (left to right, top to bottom) is that given in the clustered heat maps of Fig. 1 and Supplementary Fig. 1.

6ZB4A	down	7A96B	up	7CWUC	up	7CT5B	up	7K8WG	down	7JV6B	down	7K8ZA	down	7KECB	up
6ZB4B	down	7A96C	down	7CWL	up	7K8SA	down	6ZHDC	down	7JV6E	down	7KJ2A	down	7KEAB	up
6ZB4C	down	7A93B	up	7A94C	down	7K8SB	down	7K8UA	down	7K8ZC	down	7KJ5A	down	7KDLB	up
6ZB5A	down	7DK3A	down	6ZXNB	down	7K8SC	down	7JWBA	down	7LCNA	down	7KDJ	down	7JWBC	up
6ZB5B	down	7DK3B	down	6ZXNC	down	7K8YB	down	7K8WA	down	7LCNC	down	7KEBA	down	6X2AB	up
6ZB5C	down	7DCCE	up	7A25B	down	7K8YD	up	7JZNC	up	7LCNK	down	7KE9A	down	6VSB	down
6XR8A	down	7DCCI	up	7A25C	down	7K8YE	up	6Z97A	down	6X6PA	down	6VYBA	down	7A4NA	down
6XR8B	down	7DCC	up	7A29C	down	7K8XA	down	7CHHA	up	6X6PB	down	6X2AA	down	7A4NB	down
6XR8C	down	7DD2C	up	7DK3C	up	7K8XB	down	6ZHDA	up	6X6PC	down	7KECA	down	7A4NC	down
6ZGEA	down	7DD2K	up	6ZP0A	down	7K8XC	up	6Z43A	up	6XM3C	down	7KEAA	down	6WPTB	up
6ZGEB	down	7DD8C	up	6ZP0B	down	7K8YA	down	7CHHC	down	7KDIA	down	6XCMA	down	7AD1C	down
6ZGEC	down	7DCXC	up	6ZP0C	down	7KKKC	down	7K8UB	up	7KDIB	down	7KDLA	down	7AD1A	down
6ZGIA	down	7DCXK	up	6ZOYA	down	7KKKE	down	7JWBB	up	7KDIC	down	7KDHC	down	6XF6B	up
6ZGIB	down	7DK6A	up	6ZOYB	down	7KKLA	down	6XCNA	up	7KDBA	down	6X2BA	down	7JWYC	down
6ZGIC	down	7DK6B	up	6ZOYC	down	7KKLC	down	6XCNC	up	7KDBG	down	6X2AC	down	6XLUC	down
7JJIA	down	7DK7A	up	6ZP1A	down	7KKLD	down	6XCNE	up	7KJ3C	down	6XM5A	down		
7JJIB	down	7DK7B	up	6ZP1B	down	6ZDHA	up	6XCMB	up	7L06A	down	7KJ2C	down	7CN9C	down
7JJIC	down	7DK7C	up	6ZP1C	down	6ZDHB	up	6XCNC	up	7L06B	down	7KJ5C	down	6XM3B	up
7DF3A	down	7DK5A	up	6ZOXA	down	6ZDHC	up	7JZLC	up	7L06C	down	7KDJC	down	6XM4B	up
7DF3B	down	7DD2D	down	6ZOXB	down	6XF5A	down	7K90A	down	7KE6A	down	7KEBC	down	6ZOZA	down
7DF3C	down	7DD8D	down	6ZOXC	down	6XF5B	down	7K90B	down	7KE6B	down	7KE9C	down	6ZOZB	down
7DDDA	down	7DD8E	down	7K43A	down	6XF5C	down	7K90C	down	7KE6C	down	6VYBC	down	6ZOZC	down
7DDDB	down	7DCXD	down	7K43B	down	7K8TA	down	6Z97B	up	6X2CA	down	7KECC	down	7JWYB	down
7DDDC	down	7DK6C	down	7K43E	down	7K8TB	up	7BYRB	down	6X2CB	down	7KEAC	down	6XLUB	down
6ZP2A	down	7DK5B	down	7CAIA	up	7K8TC	down	7KJ3A	up	6X2CC	down	7KDLC	down	6XM5B	down
6ZP2B	down	7DK5C	down	7CAIB	up	7KL9B	up	7KJ3B	up	7KDKA	down	6WPTC	up	6XM0B	up
6ZP2C	down	7DK4A	up	7CAIC	down	7K4NA	up	7KJ2B	up	7KDKB	down	6WPSA	down	6XKLA	up
6ZGHC	down	7DK4B	up	7CAKA	up	7K4NB	up	7KJ4A	up	7KDKC	down	6WPSB	down	7AD1B	up
6ZGHB	down	7DK4C	down	7CAKB	up	7K4NE	up	7KJ4B	up	7KE4A	down	6WPSE	down	7C2LA	up
7DDNB	down	7KSGA	up	7CAKC	up	7KNEA	down	7KJ4C	up	7KE4B	down	7KDB	up	7BYRA	up
7DDNC	up	7KSGB	up	6XEYC	down	7KNEC	down	7KJ5B	up	7KE4C	down	6X79A	down	6VSB	up
7DF4B	up	6ZXNA	up	6XEYA	down	7KNHC	down	7JZNB	up	7L02A	down	6X79B	down	7CN9B	down
7DF4D	down	7CWUB	up	7CWMA	down	7KNBA	down	6XM0C	down	7L02B	down	6X79C	down	7CN9A	up
7DF4C	down	7CWSO	up	7C2LB	down	7KNBC	down	6XM0A	down	7L02C	down	6XKLC	down	6XS6C	up
7DDNA	down	7CWSQ	up	7C2LC	down	7KMZC	down	6XM4A	down	7L09A	down	6WPTA	up	6XS6A	up
6ZGGB	up	7CWSR	up	6XEYB	down	7KL9A	up	6XM5C	down	7L09B	down	6XM4C	down	6XS6B	up
7A95C	up	7CWUA	up	7KNIA	up	7KL9C	up	6XM3A	down	7L09C	down	7JWYA	down	6ZGHA	up
6ZGGA	down	7A25A	up	7KNIB	up	7JW0A	up	7CHHB	down	7KE8A	down	6XLUA	down		
7A95A	up	7A29A	up	7KNIC	up	7JW0B	up	7JZNA	down	7KE8B	down	7BYRC	down		
6ZGGC	down	7A29B	up	7KNEB	up	7JW0E	up	7JZLA	down	7KE8C	down	6VSB	down		
7A93A	up	7CWLC	up	7KNHA	up	7JVC	up	7JZLB	up	6X29A	down	6XKLB	down		
7A93C	down	7A94A	up	7KNBB	up	7JVCB	up	7K8VC	down	6X29B	down	7K8WB	up		
7A95B	down	7CWL	down	7KMZA	up	7JVC	up	6ZWVA	down	6X29C	down	7LD1C	up		
7A97B	up	7KSGC	up	7KNHB	up	7K8UC	up	6ZWVB	down	7KE7A	down	6X2BB	up		
7A98A	up	7CWMC	down	7KMZB	up	6Z43C	down	6ZWVC	down	7KE7B	down	6X2BC	up		
7A98B	up	7CWNA	up	7KMSB	up	6XF6C	down	7K8VA	down	7KE7C	down	7KDJ	up		
7A98C	up	7CWNC	up	7KMSA	up	6Z97C	down	7K8VB	up	6VXXA	down	7KEBB	up		
7A97C	down	7CWMB	up	7KMSC	up	6ZHB	down	7LD1A	up	6VXXB	down	7KE9B	up		
7A97A	up	7A94B	down	7CT5A	up	6Z43B	down	7LD1B	up	6VXXC	down	7K8ZB	up		
7A96A	up	7CWNB	up	7CT5C	up	6XF6A	down	7JV6A	down	7KDHA	down	6VYBB	up		

**Supplementary Table 2.** Monomer PDB/chain identifiers are given in pairs with down or up states for the RBD, S protein trimers in each of 6 sub-groups (labelled by the column headers).

locked	locked	pHlocked	pHlocked	diSlocked	diSlocked	D614Gset	D614Gset	closed	closed	open	open
6ZB4A	Down	6XLUA	Down	6ZOZA	Down	7KDKA	Down	6VXXA	Down	7A98A	Up
6ZB4B	Down	6XLUB	Down	6ZOZB	Down	7KDKB	Down	6VXXB	Down	7A98B	Up
6ZB4C	Down	6XLUC	Down	6ZOZC	Down	7KDKC	Down	6VXXC	Down	7A98C	Up
6ZB5A	Down	6XM5A	Down			7LWSA	Down	6X6PA	Down	7KJ4A	Up
6ZB5B	Down	6XM5B	Down			7LWSB	Down	6X6PB	Down	7KJ4B	Up
6ZB5C	Down	6XM5C	Down			7LWSC	Down	6X6PC	Down	7KJ4C	Up
6XR8A	Down	7JWYA	Down			7KRQA	Down	6X29A	Down	7KNIA	Up
6XR8B	Down	7JWYB	Down			7KRQB	Down	6X29B	Down	7KNIB	Up
6XR8C	Down	7JWYC	Down			7KRQC	Down	6X29C	Down	7KNIC	Up
6ZGEA	Down					7KRRA	Up	6X2CA	Down	7K4NA	Up
6ZGEB	Down					7KRRB	Down	6X2CB	Down	7K4NB	Up
6ZGEC	Down					7KRRC	Down	6X2CC	Down	7K4NE	Up
6ZGIA	Down					7KRSA	Up	6ZP0A	Down	7KSGA	Up
6ZGIB	Down					7KRSB	Down	6ZP0B	Down	7KSGB	Up
6ZGIC	Down					7KRSC	Down	6ZP0C	Down	7KSGC	Up
7JJIA	Down					7BNMA	Down	6ZOYA	Down	7CWNA	Up
7JJIB	Down					7BNMB	Down	6ZOYB	Down	7CWNB	Up
7JJIC	Down					7BNMC	Down	6ZOYC	Down	7CWNC	Up
7DDA	Down					7BNNA	Down	6ZP1A	Down	7KMSA	Up
7DDDB	Down					7BNNB	Up	6ZP1B	Down	7KMSE	Up
7DDDC	Down					7BNNC	Down	6ZP1C	Down	7KMSE	Up
6ZP2A	Down					7BNOA	Up	6ZOXB	Down	7LD1A	Up
6ZP2B	Down					7BNOB	Up	6ZOXB	Down	7LD1B	Up
6ZP2C	Down					7BNOB	Down	6ZOXC	Down	7LD1C	Up