COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
J O U R N A L

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

Vanessa R. Lobo, Jim Warwicker *

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

A R T I C L E   I N F O

A B S T R A C T

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 side-chains are generally surface located and weakly interacting when not in a locked conformation. According to 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.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

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]. Other, theoretical, work has looked at the wider effects of virus-induced pH change in SARS-CoV-2 infection [13].

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 [14]. These transitions include packing

---

* Corresponding author.
  E-mail address: jim.warwicker@manchester.ac.uk (J. Warwicker).

changes that resemble closed to locked form differences [15]. 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 [15].

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 [16] is important as knowledge of the most effective conformational targets for vaccine production become apparent [17]. 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 [18], including studies of receptor binding [19], drug binding to the linoleic acid binding pocket [20], the opening of cryptic pockets as potential drug targets [21], temperature effects [22], and the effect of glycosylation [23].

Specifically for pH-dependent effects, several computational methods are available. Constant pH molecular dynamics has been applied to coronavirus protease active sites, and reported on the protonation states that may be best suited to inhibitor design [24,25]. 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 [26], a web server for pK_a calculations has been benchmarked against known key sites for the acidic pH-dependence of influenza virus hemagglutinin [27]. 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 [21].

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 [28] in January 2021, from a list of list generated by probing the SARS-CoV-2 spike sequence, UniProt [29] id P0DTC2, with the Basic Local Alignment Search Tool (BLAST) [30] into the RCSB/PDB at the National Center for Biotechnology Infor-

mation, 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 [31], 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. A measure of model fit to cryo-EM map, the segment Manders' overlap coefficient (SMOC) was retrieved for each amino acid in each structure of the sub-groups, from analysis reported by the coronavirus structural task force [32]. An average SMOC value for each structure was calculated.

2.2. Sequence data

Sequences of spike proteins from the seven coronaviruses currently known to infect humans were aligned with the Clustal Omega [33] implementation at the European Bioinformatics Institute [34], 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 SASA) was calculated for S protein monomer burial within a trimer. For each monomer within a trimer (after non-spike components have been removed), SASA was calculated for the monomer alone and for the monomer within the trimer, the difference (d-SASA) giving the extent of burial. The contribution of polar and non-polar atoms to this burial was recorded, following previous methodology [35]. To facilitate comparison between sub-groups, d-SASA 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-SASA) 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 [26]. 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 devi-

ations ($\triangle pK_a$s) from normal values, unless those are stabilising. On the other hand, buried groups, shielded from water-dominated interactions, can have $\triangle pK_a$ values that relate to substantial desta-bilisation as well as stabilisation. The method has been imple-mented in a web tool, following testing with sites known to determine the pH-dependence of influenza virus hemagglutinin stability [27]. 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 [36 37]:

$$\triangle G_{pH-dep} = \int (2.303RT \triangle Q)dpH \qquad (1)$$

where $\triangle G_{pH\text{-}dep}$ is the pH-dependent contribution to conforma-tional stability (over the pH range of the integration, 7.5 to 5), dif-ferenced between two states. The states here are ionisable groups interacting in the FD/DH scheme, and the same set without inter-actions (null state with normal model compound pK_as). $\triangle Q$ 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 identi-fied since $\triangle Q$ separates into component ionisations. For 5 of the 6 structural sub-groups, monomers are either all RBD down or all RBD up, and in many cases will have been refined with imposed trimer symmetry. It is still useful to calculate on all monomers to assess the errors associated with numerical parts of the calculation (use of a finite difference grid and Monte Carlo sampling of proto-nation states). The set of predicted pH-dependent energies for each ionisable group reproduce closely, for example with correlation coefficients of 0.99 between monomers of 6vxx, indicating self-consistency of the electrostatics calculations.

An additional method was used to calculate pK_as, in order to compare with FD/DH results. The empirical calculation tool PROPKA [38] implemented in PDB2PQR at the APBS server [39], was used (with default parameters) for the 6 sub-groups of spike proteins (36 trimers in total). Predicted $\triangle pK_a$s from PROPKA were averaged for various single or groups of amino acids, in each of the 6 spike protein sub-groups.

Molecular viewers used to analyse spike proteins were NGL [40], Swiss PDB Viewer [41], and PyMOL. For representative trimer structures in the locked (6zp2), closed (6zp1) and open (7a98) sub-groups, d-SASA values were transferred to the B factor column (capped at 99 Å^2), 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, SASA loss upon integration of that monomer within the trimer was calculated (d-SASA). The order of monomers in the d-SASA heat map of Supplementary Fig. S1 is the same as that in Fig. 1. Summed d-SASA for each monomer is also shown. Known locked conformations mostly exhibit higher d-SASA than other conformations, including closed. The region around 840 in the d-SASA 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 inter-faces 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 mono-mers) each. Other sub-groups were added: a single locked confor-mation trimer with engineered cysteines creating a disulphide link [31] (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-SASA values show clear dis-tinction 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-SASA intermediate between locked and closed trimers, whilst D614Gset monomer d-SASA values span a much greater range than other sub-groups.

The ratio of non-polar to polar contributions to d-SASA was cal-culated for the S1 and S2 subunits of each monomer (Fig. 2b). For external protein surfaces not involved in protein–protein interac-tions the most common value of this ratio, in a distribution over patches, is about 1.2, increasing for more hydrophobic, interacting surfaces [35]. For the open sub-group, with least monomer burial, non-polarity increases for both S1 and S2 contributions. Con-versely, monomer interface within a trimer becomes successively more polar for sub-groups that have higher d-SASA, with polarity for S1 d-SASA 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_as

Comparison of Fig. 3a and Fig. 3b reinforces the observation that d-SASA differences between S trimer sub-groups are more exten-sively spread over spike monomers than are order/disorder differ-ences. The extent of d-SASA 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
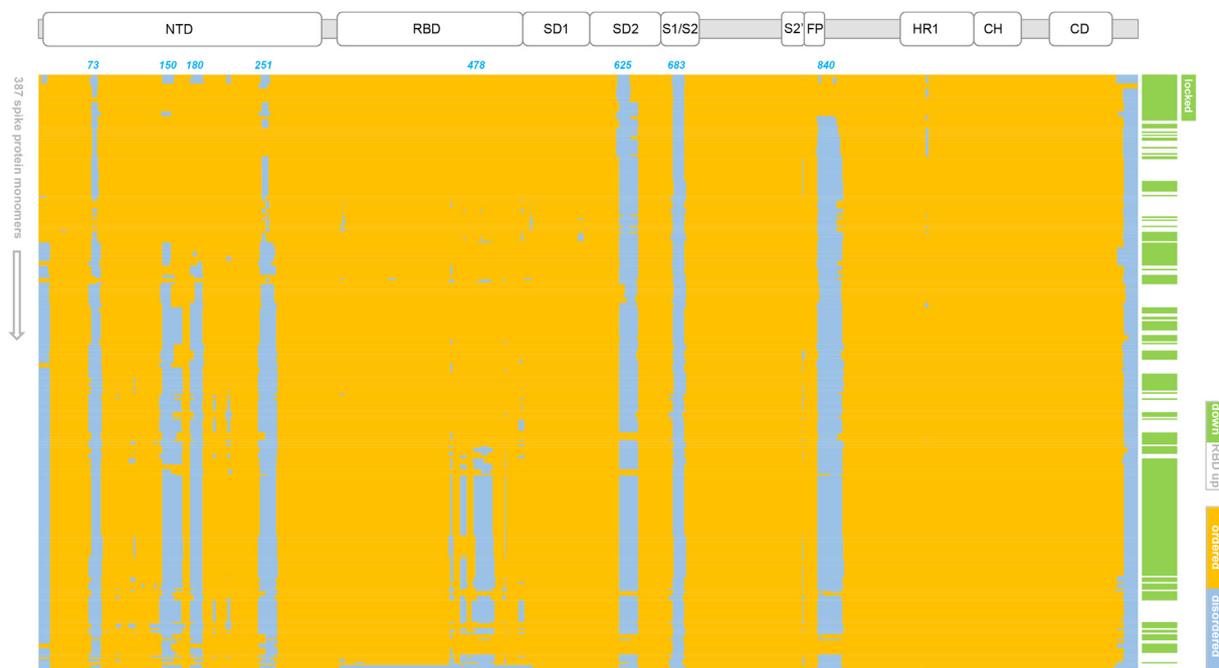
Fig. 1. Spike protein monomers clustered according to similarity of ordered/disordered regions. The spike387 set is listed vertically according to similarity with chain A of 6zb4 (top). Order (yellow) and disorder (blue) are plotted, according to presence or absence of amino acid coordinates in the structure. 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
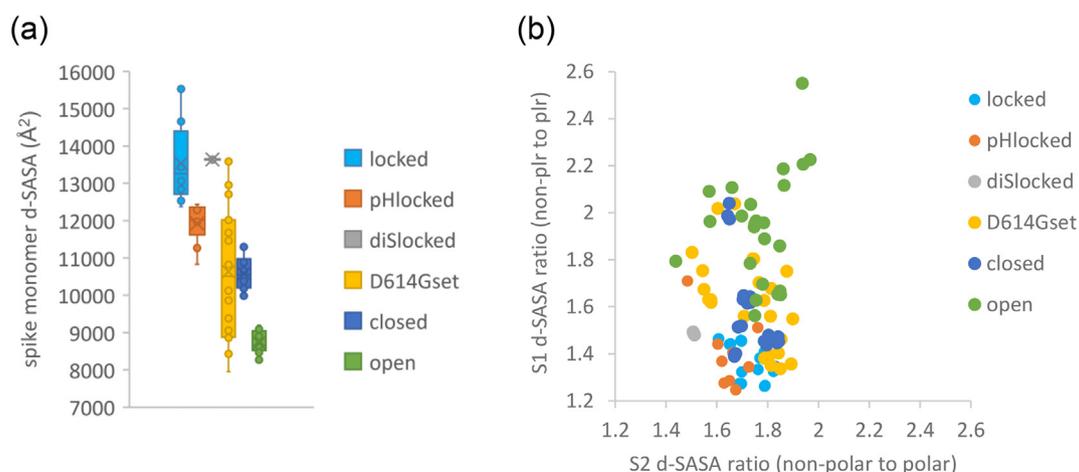
Fig. 2. Buried SASA of monomers within spike trimer, compared between sub-groups. (a) Box and whisker plot for d-SASA distributions within each sub-group. (b) A scatter plot showing the ratio of non-polar to polar buried d-SASA for the S1 and S2 subunits, upon monomer burial in the trimer. Variation 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.

monomers within each sub-group (Fig. 3c). Towards the C-terminal part of S2 are several histidines [27,42], 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_as 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].
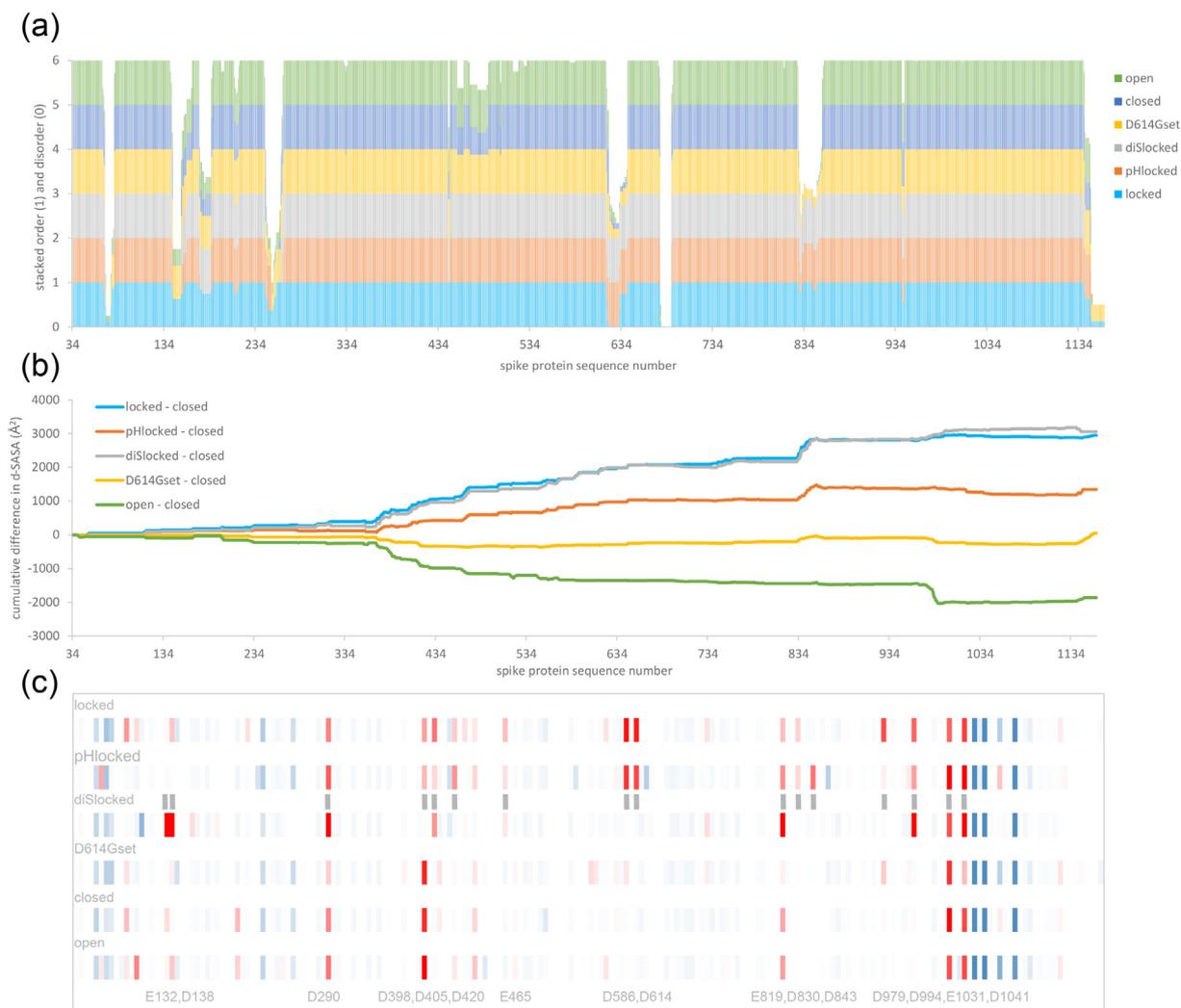
Fig. 3. Acidic groups with increased burial in the locked form are predicted to give a mild acidic pH-dependence. (a) Order/disorder is illustrated with stacked histograms (1/0) for the 6 sub-groups of spike trimers. Differences of d-SASA between sub-groups are more extensive than are those of order/disorder, 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. All sub-groups other than D614Gset consist of either all RBD down or all RBD up S trimers. The small difference between D614Gset and closed sub-groups reflects an average over substantial variation in D614Gset (Fig. 2a). (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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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.

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 [42]. 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, with the FD/DH calculation of pH-dependent stability (Fig. 4).

It is of interest to examine whether structure resolution is a factor in the results shown in Fig. 4, in particular for the key predicted differences in pH-dependence between locked and closed sub-groups. A comparison of reported resolution and segment Manders' overlap coefficients for the 36 sub-group trimers

(Supplementary Fig. S2a) shows the general decline of SMOC values for better resolution structures that has been observed previously [43], and interpreted as the influence of map resolution on the level of detail required for a given fit of model to map. Given this rather complex relationship, a SMOC criterion was excluded from filtering for structure quality, leaving the reported resolution. Filtering at 3.0 Å resolution gives 5 of 8 locked form sub-group trimers and 4 of 8 closed form sub-group trimers. Predicted pH-dependence for these resolution-filtered structures (Supplementary Fig. S2b) matches those given for the full locked and closed form sub-groups (Fig. 4), thus giving confidence in those calculations.

In order to provide context for this particular calculation method, the PROPKA empirical algorithm for pK_a prediction [38] was employed, as implemented at the APBS/PDB2PQR server [39]. Here the pH-dependent stability contributions of individual groups are not available, so $\triangle pK_a$ values were used, which are positive for aspartic acid and glutamic acid destabilisation and negative for histidine destabilisation (Supplementary Fig. S3). Over different selections of groups, similar predicted effects are seen to those for the FD/DH method (Fig. 4). The 3 S2 subunit His (H1048, H1064, H1088) are predicted to be substantially destabilising in all S protein sub-groups. D398 and D614 are predicted to be destabilising, with similar trends between sub-groups to those in Fig. 4. The subset of groups labelled 14-DE refers to those Asp/Glu labelled in Fig. 3c, but excluding D398 and D614 since they are shown separately. Locked, pHlocked and diSlocked all show $\triangle pK_a$s greater than average Asp/Glu values, for 14-DE, whilst $\triangle pK_a$s for closed and open forms are more comparable to average, with D614Gset intermediate. This is consistent with highlighting these acidic amino acids as being of interest in differentiating the predicted pH-dependent stabilities of locked and other S trimer forms. There is a difference between predictions of FD/DH (Fig. 4) and PROPKA (Supplementary Fig. S3) in the greater contribution of H49 and H519 according to the PROPKA method. Interestingly, H519 is in the proximity of D979, one of the highlighted acidic groups in Fig. 3c, and the mutation H49Y increases cell entry (relative to wild type), in an S-pseudotyped lentiviral system (although much less than D614G) [44]. Given these observations, it would be reasonable to keep H49 and H519 in mind when considering potential sites of pH-sensing in the S protein trimer.

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 [45], 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 [14]. 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-SASA for monomers in this structure (Supplementary Fig. S4) 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-SASA values, but these structures mostly have mixed up and down RBD within each trimer (Supplementary Fig. S4).

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 foms. 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 [46]. 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_as 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-SASA 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 SASA on formation or tightening of an interface. The acidic pK_a increases in the range pH 4 to 7. A bias
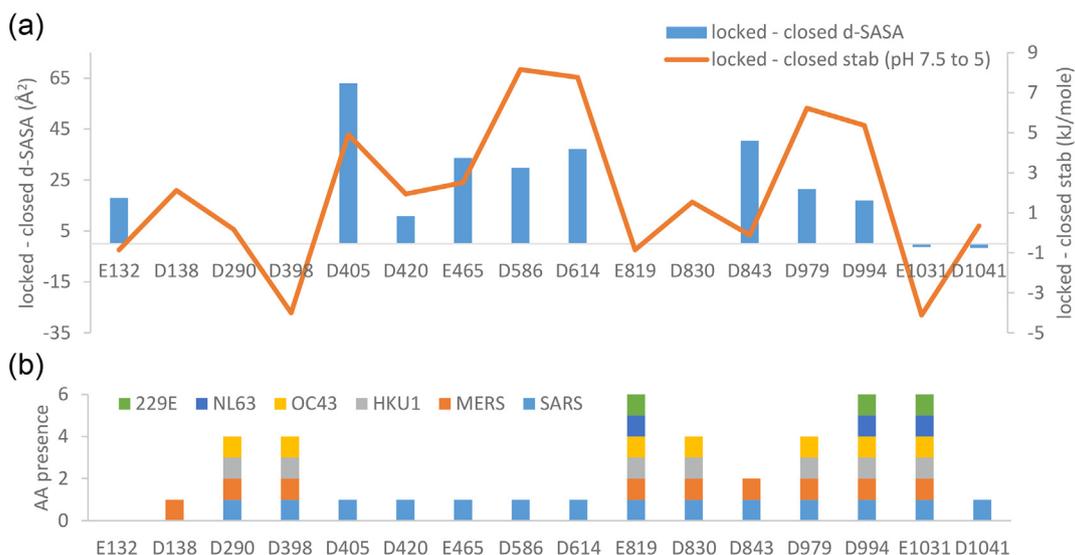
Fig. 5. Individual residue contributions to locked / closed form differences. (a) For amino acids highlighted in Fig. 3c, both d-SASA and predicted pH 7.5 to 5 stability change are plotted as differences (locked – closed). (b) The presence of these residues in a set of coronaviruses that infect humans (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.
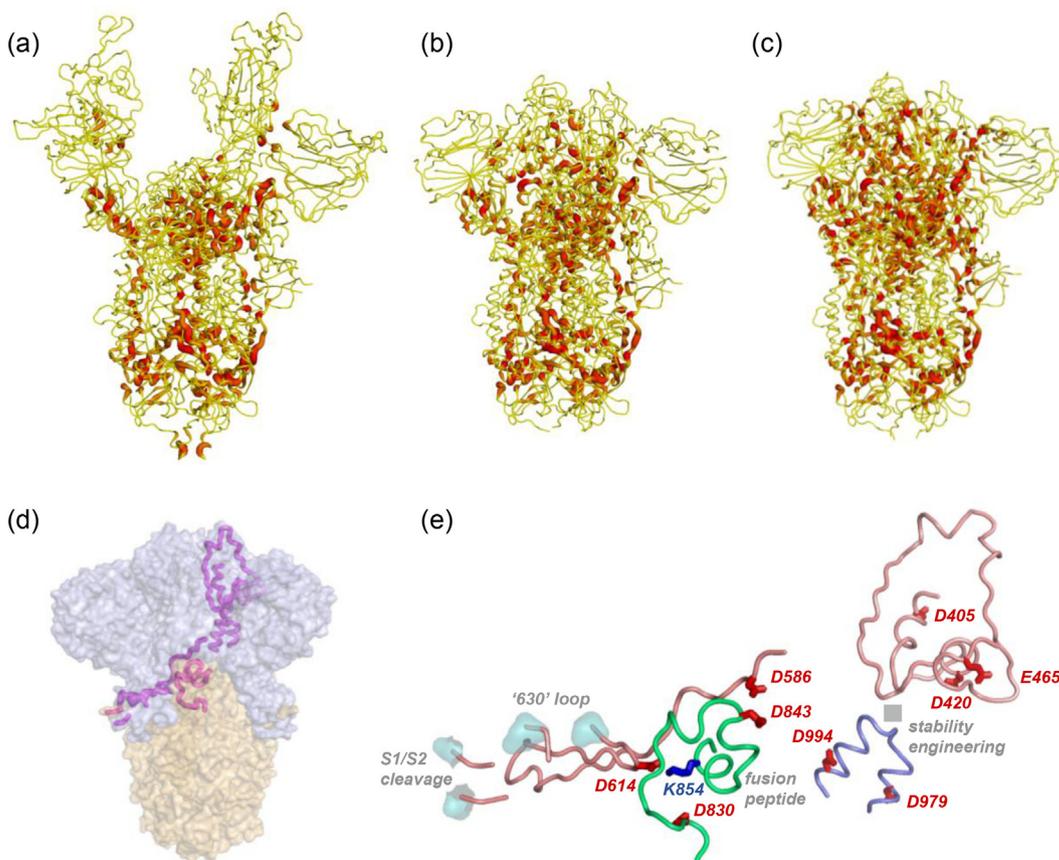
Fig. 6. Location of sites predicted to differentiate the mild acidic pH-dependence of locked and closed forms. Tube plots of spike protein backbone are shown with colour-coding from d-SASA of 0 Å² (yellow) to 80 Å² (red) for open/7a98 [9] (a), closed/6zp1 [31] (b), and locked/6zp2 [31] (c) forms. Within the regions of increased d-SASA 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 [47], 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 [48]. 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_as higher than 7 [49]. 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 [50]. 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 [51]. 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 [52]. An acidic pH-dependent stability has been engineered into staphylococcal nuclease with the mutation V66E, burying a glutamic acid (measured pK_a 8.8) in the protein interior [53]. Here, the unfavourable stability consequence of glutamic acid burial is absorbed by the stability of the enzyme fold, with the addition of water molecules that penetrate to form a small channel around V66E [53]. For spike protein, it is hypothesised that unfavourable burial of acidic groups could also be coupled to other conformational factors (that are energetically favourable), including formation of the tighter locked form interfaces and linoleic acid binding. In general terms, there are some parallels with conformational transitions and pocket factor binding in picornaviruses [54].

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 [55].

4. Conclusions

Analysis of spike trimer structures shows the expected large SASA 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 [46].

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 seen 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 [56]. 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 [57]. 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 [58].

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 [59]. 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.

CRediT authorship contribution statement

Vanessa R. Lobo: Data curation, Investigation, Methodology. **Jim Warwicker:** Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.08.049.

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, Román RG, Chaudhari A, Gurry C, Le TT, et al. Status Report on COVID-19 Vaccines Development. Curr Infect Dis Rep 2021;23(6). https://doi.org/10.1007/s11908-021-00752-3.
[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. NY: Science (New York); 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). https://doi.org/10.15252/embj.2020106267.

[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;588(7837):327–30.

[10] Zhou T, Tsybovsky Y, Gorman J, Rapp M, Cerutti G, Chuang G-Y, 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):867–879.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] Lucia U, Grisolia J, Deisboeck TS. Thermodynamics and SARS-CoV-2: neurological effects in post-Covid 19 syndrome. arXiv:210712006 [physics] 2021.

[14] 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.

[15] 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.

[16] Riley TP, Chou H-T, 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. Front Immunol 2021;12:660198.

[17] 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.

[18] Amaro RE, Mulholland AJ. Biomolecular Simulations in the Time of COVID19, and After. Comput Sci Eng 2020;22(6):30–6.

[19] 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.

[20] 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.

[21] 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.

[22] Martí D, Torras J, Bertran O, Turon P, Alemán 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.

[23] 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.

[24] Henderson JA, Verma N, Harris RC, Liu R, Shen J. Assessment of proton-coupled conformational dynamics of SARS and MERS coronavirus papain-like proteases: Implication for designing broad-spectrum antiviral inhibitors. J Chem Phys 2020;153(11):115101. https://doi.org/10.1063/5.0020458.

[25] 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. J Am Chem Soc 2020;142(52):21883–90.

[26] Warwicker J. Improved pKa calculations through flexibility based sampling of a water-dominated interaction scheme. Protein Sci 2004;13(10):2793–805.

[27] Hebditch M, Warwicker J. protein-sol pKa: prediction of electrostatic frustration, with application to coronaviruses. Bioinformatics 2020.

[28] 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.

[29] UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 2019;47(D1):D506-D15.

[30] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.

[31] 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.

[32] Croll TI, Diederichs K, Fischer F, Fyfe CD, Gao Y, Horrell S, et al. Making the invisible enemy visible. Nat Struct Mol Biol 2021;28(5):404–8.

[33] 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. Mol Syst Biol 2011;7(1):539. https://doi.org/10.1038/msb:2011.75.

[34] 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):W636–41.

[35] 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.

[36] Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238(3):415–36.

[37] Gilson MK. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. Proteins 1993;15(3):266–82.

[38] Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput 2011;7(7):2284–95.

[39] Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci 2018;27(1):112–28.

[40] Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL Viewer: Web-based molecular graphics for large complexes. Bioinformatics 2018.

[41] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18(15):2714–23.

[42] Warwicker J. A model for pH coupling of the SARS-CoV-2 spike protein open/closed equilibrium. Brief Bioinform 2021;22(2):1499–507.

[43] Lawson CL, Kryshtafovych A, Adams PD, Afonine PV, Baker ML, Barad BA, et al. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. Nat Methods 2021;18(2):156–64.

[44] Ozono S, Zhang Y, Ode H, Sano K, Tan TS, Imai K, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat Commun 2021;12(1). https://doi.org/10.1038/s41467-021-21118-2.

[45] 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.

[46] 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.

[47] Bartik K, Redfield C, Dobson CM. Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional 1H NMR. Biophys J 1994;66(4):1180–4.

[48] 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.

[49] 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.

[50] Kuhlgert 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.

[51] Foit L, George JS, Zhang BW, Brooks 3rd CL, Bardwell JC. Chaperone activation by unfolding. Proc Natl Acad Sci U S A 2013;110(14):E1254–62.

[52] Lan W, Valente JJ, Ilott 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. https://doi.org/10.1080/19420862.2020.1763138.

[53] Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, et al. High apparent dielectric constants in the interior of a protein reflect water penetration. Biophys J 2000;79(3):1610–20.

[54] Rossmann MG. Viral cell recognition and entry. Protein Sci 1994;3(10):1712–25.

[55] 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.

[56] Zahradník J, Nunvar J, Schreiber G. SARS-CoV-2 convergent evolution as a guide to explore adaptive advantage. bioRxiv 2021:2021.05.24.445534.

[57] 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;6(7):899–909.

[58] 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.

[59] 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). https://doi.org/10.1038/s41467-020-20321-x.