



# The Physical Basis for pH Sensitivity in Biomolecular Structure and Function, With Application to the Spike Protein of SARS-CoV-2

Jim Warwicker\*

School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom

## OPEN ACCESS

### Edited by:

Jan M Antosiewicz,  
University of Warsaw, Poland

### Reviewed by:

Miguel Machuqueiro,  
University of Lisbon, Portugal  
Son Tung Ngo,  
Ton Duc Thang University, Vietnam  
Aihua Zheng,  
Chinese Academy of Sciences (CAS),  
China

### \*Correspondence:

Jim Warwicker  
jim.warwicker@manchester.ac.uk

### Specialty section:

This article was submitted to  
Biophysics,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 12 December 2021

**Accepted:** 19 January 2022

**Published:** 18 February 2022

### Citation:

Warwicker J (2022) The Physical Basis for pH Sensitivity in Biomolecular Structure and Function, With Application to the Spike Protein of SARS-CoV-2. *Front. Mol. Biosci.* 9:834011. doi: 10.3389/fmolb.2022.834011

Since pH sensitivity has a fundamental role in biology, much effort has been committed to establishing physical models to rationalize and predict pH dependence from molecular structures. Two of the key challenges are to accurately calculate ionizable group solvation and hydration and then to apply this modeling to all conformations relevant to the process in question. Explicit solvent methods coupled to molecular dynamics simulation are increasingly complementing lower resolution implicit solvent techniques, but equally, the scale of biological data acquisition leaves a role for high-throughput modeling. Additionally, determination of ranges of structures for a system allows sampling of key stages in solvation. In a review of the area, it is emphasized that pH sensors in biology beyond the most obvious candidate (histidine side chain, with an unshifted  $pK_a$  near neutral pH) should be considered; that modeling can benefit from other concepts in bioinformatics, in particular modulation of interactions and function in families of homologs; and that it can also be beneficial to incorporate as many experimental structures as possible, to mitigate against small variations in conformation and to analyze larger, functional, conformational changes. These aspects are then demonstrated with new work on the spike protein of SARS-CoV-2, looking at the pH dependence of variants, including prediction of a change in the balance of locked, closed, and open forms at neutral pH for the Omicron variant spike protein.

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

## INTRODUCTION

Examples of pH dependence occur throughout biological systems. The narrow control of key properties, such as blood pH, mediated by various sensors, transporters, and channels (Rajkumar and Pluznick, 2018), or cytosolic pH (Chiche et al., 2010) points to the importance of pH in underlying molecular processes. Gradients of pH at a cell's outer membrane or within mitochondria underlie some of the most basic processes of metabolic energy transduction, mechanistically with structures demonstrating a probable proton pathway for mitochondrial ATP synthase (Klusck et al., 2017). Further, more complex cell types display an array of membrane-enclosed subcellular organelles that compartmentalize sets of macromolecules and small molecules, including protons and pH (Demaurex, 2002). Global properties of organelle sub-proteomes, including pI distributions (Kurotani et al., 2019) and predicted maximal pH-dependent stability (Chan and Warwicker, 2009)

correlate with their environmental pH. Differences in acidification of the import and export pathways and organelles of eukaryotes are linked to processes that include altered protease activation and receptor/ligand cycling under pH control (Mellman, 1992).

In some cases, sufficient information is available to allow redesign of systems toward a targeted endpoint. An important example is lengthening the circulation half-life of protein therapeutics (biologics), making use of (and extending) pH-dependent recycling of IgGs and serum albumin by the neonatal Fc receptor (FcRn) (Pyzik et al., 2019). Other efforts are underway to engineer systems for altered performance with regard to pH, often screening for successful designs from high-throughput mutagenesis. The pH dependence of antibody-antigen binding can be altered, for example, with increased binding at the acidic pH of the (extracellular) tumor environment (Sulea et al., 2020), or with design aimed specifically at the generation of a pH switch (Strauch et al., 2014). These types of studies are enhanced by advances in fluorescent protein technology that allow increasingly sensitive analysis of extracellular and intracellular pH values and which themselves benefit from molecular engineering (Bencina, 2013). Enzyme pH dependence can be engineered through directed evolution for adaptation to particular tasks in biotechnology or synthetic biology (Scheiblbrandner et al., 2017). Computational design and experimental validation studies have been combined in a study that adjusts buried histidine environments using the Rosetta suite of design tools to control conformation and oligomerization through pH switching (Boyken et al., 2019). Even this study, though, starts with the rather simple observation that histidine side chains are the most notable candidates for pH sensing (owing to their normal  $pK_a$  of 6.3) and that ionization of buried groups leads (in the absence of very precise complementary interactions in the immediate vicinity) to instability.

This work will examine the computational methods used to predict pH dependence in proteins, comparing their maturity with other techniques of structural bioinformatics, and will discuss examples of naturally occurring pH dependence, emphasizing the lessons that can be learned when further developing prediction tools. Then, a system that is currently challenging the scientific community on multiple fronts is studied, SARS-CoV-2 and specifically the spike protein. Emerging evidence of a role for pH switching between conformational forms, potential molecular sources of that pH dependence, and how these might be changing in variants are discussed.

## Computational Methods for Predicting pH Dependence

Buried histidine side chain involvement in various biological systems has been experimentally characterized, including conformational stability and disease susceptibility in prion protein diseases (Malevanets et al., 2017), structural core stabilization in a K-homology module RNA binding motif (Fraternali et al., 1999), pH sensing in the Cnu transcription

factor (Narayan and Naganathan, 2018), and probing transmembrane protein helical structure (Afrose et al., 2021). In some studies, groups such as buried histidine side chains are described as electrostatically frustrated (depending on the pH) (Gopi et al., 2018; Narayan et al., 2020), a concept also used in application to viral protein structure (Hebditch and Warwicker, 2020). It seems that burial from solvent of an ionizable group with  $pK_a$  close to physiological pH, which is normally around neutral in the cytoplasm, goes a long way in identifying regions of pH sensitivity. Indeed, it may be that there exists more scope for bioinformatics searches for pH sensors purely on this basis, perhaps with extension to incorporate non-buried histidine side chains, as has been suggested for arginine-to-histidine somatic mutations in cancer adaptation to metabolic-induced pH changes (White et al., 2017). This philosophy has been successfully employed to study several systems, including pH sensors in Ras-specific guanine nucleotide exchange factor (Vercoulen et al., 2017) and in  $\beta$ -catenin (White et al., 2018).

All aspects of sequence and structural bioinformatics are now informed by the plethora of data available, which is why finding a model for searching those data is so valuable. Sequence analysis has benefited enormously from Kyte-Doolittle (hydrophobicity scale) prediction of transmembrane domains (Kyte and Doolittle, 1982) and the incorporation of hydrophobicity and charge for identification of protein regions likely to be intrinsically disordered (Uversky et al., 2000). Prediction of pH dependence requires an estimate of the balance of interactions between ionized and non-ionized forms of pH-titratable groups, in each of the states being considered (Antosiewicz et al., 1994). Typically, these states might be folded and denatured protein (folding), or complex and non-complex (binding), or two different folded conformations. Net electrostatic interactions between ionizable groups are often approximated as zero in the unfolded state, although this can be a problem (Tan et al., 1995). Methods for predicting pH dependence have been developed based on continuum electrostatics approximations for protein and solvent (Warwicker and Watson, 1982). One such implementation (Warwicker, 2004), used in the  $pK_a$  application of the protein-sol tool (Hebditch and Warwicker, 2020), incorporates a scheme for partitioning electrostatic interactions (Bashford and Karplus, 1990) and derives  $pK_a$ s with Monte Carlo sampling of ionization states (Beroza et al., 1991). There are many other implementations of a wide variety of continuum electrostatics (implicit solvent) methods, with reviews available, for example, general (Alexov et al., 2011; Gunner and Baker, 2016) and focused on the generalized Born method (Onufriev and Case, 2019). A major advantage of continuum methods is that their speed allows application in high-throughput studies, aligned to the modern biology of omics data collection. They are complementary to the growing use of lower throughput constant-pH molecular dynamics methods (Chen et al., 2014), which explicitly couple pH titration to conformational change, the central aim of understanding and predicting pH sensing. There is room for both low- and high-throughput methods, not only because of the facilitation of omics scale scans with implicit solvent techniques but also due to the scale of conformational change that may be linked to pH dependence, which is

challenging for simulation, but where the experimental structure databases (and perhaps AlphaFold (2) (Jumper et al., 2021) or other modeling schemes) can play a role. In particular (for larger systems), cryo-electron microscopy (cryo-EM) structures are appearing at pace. Work with SARS-CoV-2 spike protein, in a subsequent section, refers to the great variety of information available when structures are solved under multiple conditions, including pH, ligand binding, and mutation.

Beyond the clear relevance of histidine side chains, several classes of ionizable groups need to be considered in models. Certain systems more or less eschew histidine altogether, for example, the thioredoxin and DsbA family of disulfide oxidoreductases, where cysteine  $pK_a$  couples to a scale of redox potential and range of biological function that varies substantially through the family (Gane et al., 1995; Collet and Bardwell, 2002). This example highlights the importance of studying models in the context of families and homologs. A common chemistry, disulfide bond making and breaking, and partial stabilization of a cysteine thiolate at the amino terminus of an  $\alpha$ -helix, is supplemented with sequence and small structural changes within a common fold to modulate  $pK_a$  and redox potential. A comparable example is the variation in heme group redox potentials, successfully captured with electrostatics calculations (Zheng and Gunner, 2008). As a side note, cysteine  $pK_a$  and reactivity (and that of other groups) appear to be important for the growing field of non-covalent inhibitor development in therapeutics (Fowler et al., 2017).

Aspartic and glutamic acid side chains are obvious candidates for pH sensing, well known from studies of acid–base catalysis in enzyme active sites, often contributing to pH bell curves of activity, perhaps the prototype example being E35 ( $pK_a$  6.3) of lysozyme (Bashford and Karplus, 1990; Bartik et al., 1994). Experimental determination of  $pK_a$ s in proteins, focusing on NMR, has been reviewed (Hass and Mulder, 2015). Other than catalytic systems, multiple instances of functional ionization for carboxylate side chains have been noted. Amino acids E69 and E74 potentially mediate pH dependence (in the pH range of 4.5–7.1) of coiled-coil formation in a variant of influenza hemagglutinin (Higgins et al., 2014). Association of some antigenic peptides and MHC complex class II proteins is pH-dependent, with buried interfacial Asp/Glu side chain  $pK_a$ s > 7 thought to be responsible (Belmares et al., 2000). In photosystem I, interfacial residues D612 and E613 of PsaB are proposed to modulate acidic pH-dependence of electron transfer in complexes with plastocyanin or cytochrome  $c_6$  (Kuhlgert et al., 2012). Two Asp side chains in the periplasmic chaperone HdeA of *Escherichia coli* mediate dimer dissociation from neutral to lower pH (Foit et al., 2013). Aspartic acid ionization is coupled to membrane protein function, for example, in G-protein-coupled signal transduction (Dror et al., 2011), and peptide insertion into a membrane (Teixeira et al., 2016) (Vila-Vicosa et al., 2018). Furthermore, protein–protein interactions can be influenced by the Asp charge state (Lin et al., 2017). It is apparent that carboxylate side chains should be considered alongside histidine side chains when considering pH dependence in biology, with the caveat that a degree of burial from solvent will normally be required to elevate a carboxylate  $pK_a$  toward

neutral pH. In support of this, an engineered, buried, V66E mutation ( $pK_a$  8.8) in staphylococcal nuclease introduces an acidic pH dependence of stability (Dwyer et al., 2000).

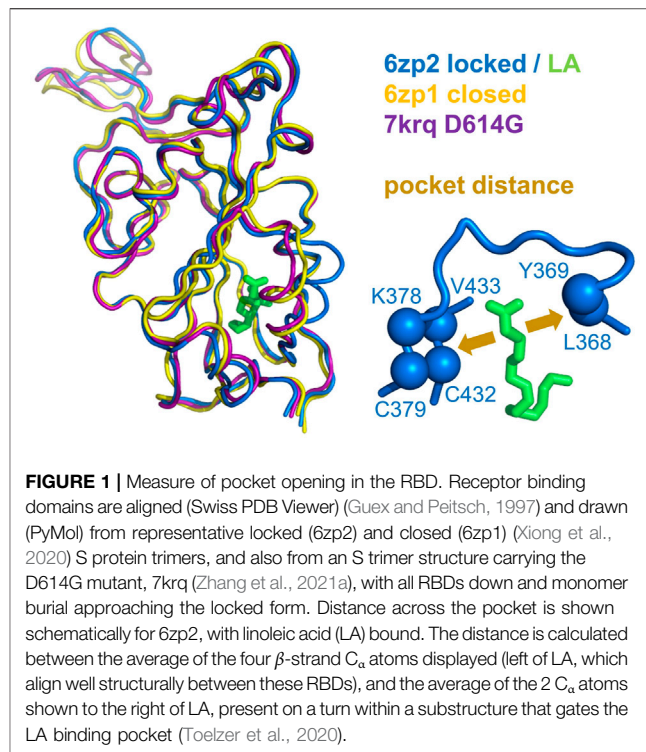
Coupling between pH sensors and their environment is a theme that was established early on, with a landmark study of the mutations arising when influenza viruses were selected for their ability to grow in cells with elevated endosomal pH, from treatment with amantadine hydrochloride (Daniels et al., 1985). A large number of the resulting mutations were present at the interface between hemagglutinin subunits and were not themselves ionizable groups. This result makes sense in that the energetics of a process that is susceptible to pH switching (conformational change, complex formation) is coupled to the pH range over which switching occurs. However, it rather complicates the issue of screening specifically for pH-sensing residues through analysis of experimental selection based on altered pH or through large-scale mutagenesis aimed at identifying pH sensors. A similar situation is apparent in acid-sensing ion channels (ASICs), where the results of moderate-scale single-site mutagenesis have been interpreted as reporting on both pH-sensing amino acids (those directly involved as part of a pH sensor) and pH-coupled amino acids (those present in regions that change upon pH switching) (Sazanavets and Warwicker, 2015).

Thus, predictions of pH dependence are more complicated than simply identifying ionizable groups with a normal  $pK_a$  value close to physiological pH (typically histidine) and that are buried from solvent. These complicating factors include sufficiently precise calculation of the balance of solvating and desolvating (dehydrating) interactions (be it with implicit or explicit solvent methods), incorporation of enough detail of changes in conformation/complexation (whether using molecular simulation/modeling methods or multiple experimental structures), consideration of the energetic coupling between structural changes and pH sensing (where experimental screens for altered pH dependence will not necessarily distinguish between them), and making sure that models do not exclude classes of groups that may be relevant (for example, carboxylate groups have well-established roles in pH dependence beyond catalytic sites). These areas are also opportunities for further development of prediction methods. A multiplicity of implicit and explicit solvent methods will facilitate implementation of combined high- and low-throughput hierarchical analyses of pH sensing, especially in regard to studies aimed at characterization of pH dependence across proteomes, as more structures and effective models become available. Included in the increasing structural database are ever more snapshots of systems under different conditions, revealing the conformation/complexation changes that pH switches act upon. Experimental screens for altered pH dependence in biological systems, making use of genomics era technologies, although not delivering uniquely pH-sensor groups, should highlight regions that couple to pH-sensing, itself a valuable insight. Supporting all areas is the common theme of bioinformatics, homology, whereby variation within a common framework and evolutionary adaptation, simplifies computation through an analysis of results for closely related systems.

## pH Dependence of SARS-CoV-2 Spike Protein Conformation

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

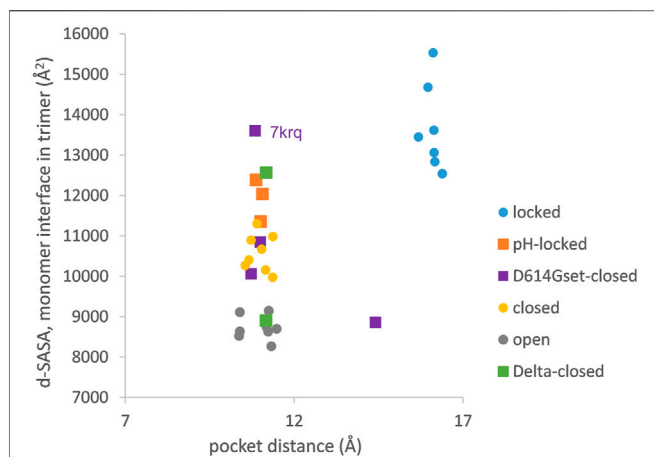
Based on the response of the S protein trimer structure to acidic pH (Zhou et al., 2020; Qu et al., 2021), a scheme has been suggested in which the trimers of a virus within the acidic pH of the secretory pathway would be protected in a more tightly packed (locked) conformation (Qu et al., 2021). This suggestion is consistent with the discovery of a locked and more tightly packed form of the SARS-CoV-2 S protein trimer at the normal pH (7.5–8) of structure solution, due to binding of a linoleic acid pocket factor in the receptor binding domain (RBD) (Toelzer et al., 2020). These observations prompt questions about which S protein amino acids are sensing pH, and how these might be coupled to pocket factor binding and packing/interface changes. Such questions highlight areas discussed in this article: regional differences in pH (inside and outside of cells), ligand binding, and interface variation identified with multiple structures and, as will become apparent, the potential influence of mutations (in the S protein). A study of S protein trimers grouped according to RBD down (with locked and closed forms treated



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

The picture of the SARS-CoV-2 spike protein encountered in the first year of the COVID-19 pandemic is therefore of a trimer that can exist in a more tightly packed form (termed locked) either with pocket factor binding at neutral pH or at acidic pH in the absence of pocket factor. However, the spike protein is changing, leading to the question of what the consequences may be for the relative stability of the locked, closed, and open forms, which is the subject of this section. Predictions for potential acid sensor residues are already available (Lobo and Warwicker, 2021). Here, the focus is on how these are situated in the context of interface and RBD pocket changes. Data from previous work for the change in solvent accessible surface area (d-SASA) upon spike monomer incorporation into trimers (in





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

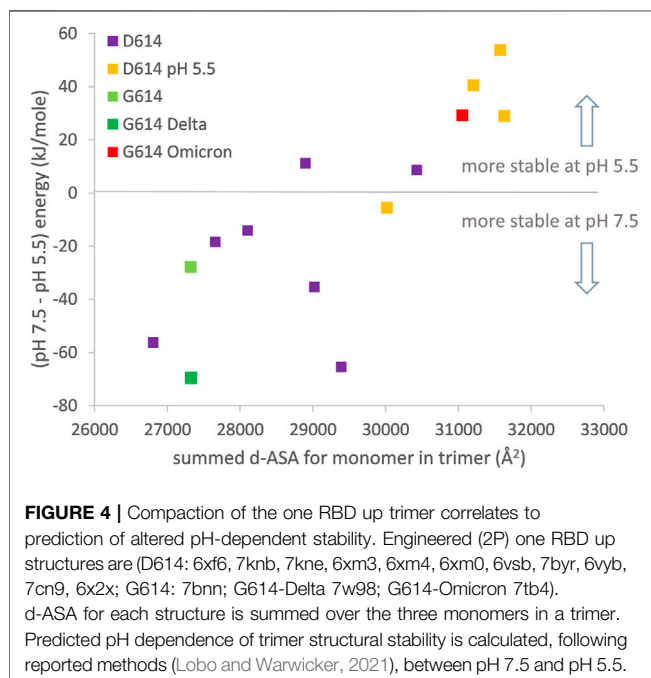
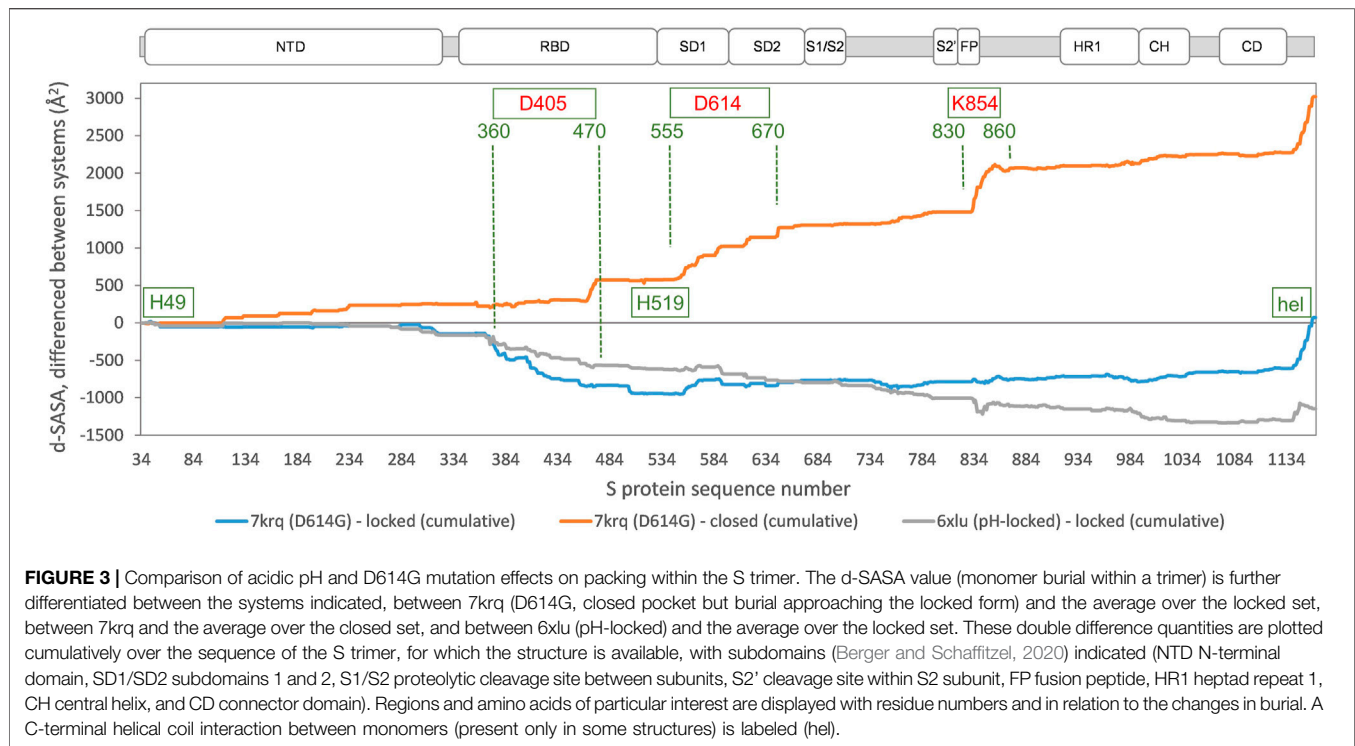
various trimer forms) are supplemented with equivalent calculations for two recent structures of the Delta variant (RCSB/PDB (Berman et al., 2007) identifiers 7v7n (10.2210/pdb7V7N/pdb) and 7sbk (Zhang et al., 2021b)). A new measure is constructed of the distance across the RBD pocket in which linoleic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

In order to establish the location of interface differences (for monomer burial in trimer) and how they change in a D614G structure with similar burial to locked forms, differences between

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

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

Finally, the S protein of the Omicron variant (<https://www.ecdc.europa.eu/en/covid-19/variants-concern>) carries S371L, S373P, and S375F mutations in the sequence next to the RBD pocket gate. There are many further mutations in the Omicron S protein, which are being discussed in the context of various features, including the extent to which they mediate escape of an immune response primed by vaccine or prior infection (Greaney et al., 2021; Zahradník et al., 2021), and how immune escape mutations may be compensated to maintain ACE2 binding (Mannar et al., 2021) (Han et al., 2022). Interesting mutations in the Omicron spike protein, with respect to packing, are D614G and the S371L, S373P, and S375F cluster. The position equivalent to 373 in the RBD of coronavirus HKU9 is also a proline, and since this amino acid lies in a segment connecting two sides of the LA-binding pocket (in the SARS-CoV-2 spike), a proline substitution could alter main chain conformation and pocket gating. Indeed, the equivalent pocket region in HKU9 (5gyq) (Huang et al., 2016) looks to be closed, with a similar outcome anticipated for the RBD of the SARS-CoV-2 Omicron variant (Zahradník et al., 2021). If this pocket is closed in the Omicron variant spike, then linoleic acid would not bind and thus would not be a route to locking (compacting) the spike protein structure. It is intriguing that early reports of the Omicron spike structure indicate a capacity to form relatively compact structures in both closed (all RBD down)



and open forms (Hong et al., 2022). Reports suggest that for the one RBD up form, this may be due, in part, to enhanced interactions between monomers (S375F–F486) (Zhou et al., 2021).

The major thrust of pH dependence predictions for non-Omicron variants is that increased compaction of the spike

trimer leads to partial dehydration of some Asp/Glu side chains and a disfavoring of a more compact form at neutral pH, unless balanced by linoleic acid binding to the RBD pocket (Lobo and Warwick, 2021). In order to include the Omicron variant spike protein (which may not bind linoleic acid) in this picture, it is necessary to study the one RBD up form since this is the only currently available structure (January 2022) that has a complete set of three RBDs resolved (7tb4) (Zhou et al., 2021). Furthermore, to maintain equivalence between the engineered status of the spike protein ectodomain, analysis is restricted to the 2P (K986P, V987P) stabilized form. In keeping with the previous work (Lobo and Warwick, 2021), increased trimer compaction correlates with a relative lowering of stability at pH 7.5 compared with pH 5.5 (Figure 4). The Omicron spike protein is predicted to be as compact as one RBD up trimers that have been stabilized by lowering the pH of structure solution to 5.5, although the Omicron spike was not solved at acidic pH. Presumably this increased compaction arises from interactions (including S375F–F486) that are absent in other SARS-CoV-2 variant spike proteins and that balance the predicted electrostatic destabilization at neutral pH. Since relatively compact trimers have been reported for both all RBD down and RBD up Omicron spike protein structures at neutral pH (Hong et al., 2022), likely in the absence of RBD pocket factor binding, it is reasonable to suggest that they may be better protected against S1 shedding than spike trimer of other SARS-CoV-2 variants. Furthermore, although a predicted pH dependence of stability remains for the Omicron variant spike, with an overall shift in the balance of trimer stabilizing interactions, it is unclear how this will translate to the prevalence of different spike conformational forms in different pH environments. One possibility is that with

relatively compact closed and open (at least one RBD up) forms, with similar electrostatic destabilization (due to partial Asp/Glu burial), transition between the forms would be facilitated.

## CONCLUDING REMARKS

The roles of factors that contribute to establishing pH sensing in biology have been discussed, including pH variation within and outside of cells, separating groups that are genuinely pH sensing (i.e., with pH titration that can be modulated) and those that couple to pH sensing (often through interactions at an interface), the importance of focusing on all groups that could titrate in a physiological pH range, and the roles that homology and omics data can play in large-scale screening for pH sensors.

In the example chosen, SARS-CoV-2 spike protein, reference is made to predictions of pH dependence with a large set of S protein trimer structures, classified into different conformational forms (Lobo and Warwicker, 2021). Although this area is currently busy with data collection, many mechanisms remain unknown, including the pH dependence of S trimer conformation and how that might relate to function. Several themes are apparent in the S protein work, including the use of a large number of input structures to capture variation, close examination of ionization site burial, and how that changes with structural classes, and proposing that Asp/Glu side chains could be responsible for observed pH-dependent effects. This

proposal is yet to be tested, indeed an alternate hypothesis is that histidines H49 and H519 could mediate pH dependence (Qu et al., 2021), but these give small predicted pH dependence (Lobo and Warwicker, 2021) and are located in regions without large interface changes between structural forms (Figure 3). One challenge facing analyses of pH dependence is how to construct experiments that continue to test models and predictions as the systems investigated become larger and direct biophysical readouts more complex to measure.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## FUNDING

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

## ACKNOWLEDGMENTS

The author thanks the staff who support the computational shared facility (CSF) at the University of Manchester.

## REFERENCES

- Afrose, F., Martfeld, A. N., Greathouse, D. V., Koeppel, R. E., and 2nd (2021). Examination of pH Dependency and Orientation Differences of Membrane Spanning Alpha Helices Carrying a Single or Pair of Buried Histidine Residues. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1863 (1), 183501. doi:10.1016/j.bbmem.2020.183501
- Alexov, E., Mehler, E. L., Baker, N., M. Baptista, A., Huang, Y., Milletti, F., et al. (2011). Progress in the Prediction of pK<sub>a</sub> Values in Proteins. *Proteins* 79 (12), 3260–3275. doi:10.1002/prot.23189
- Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994). Prediction of pH-Dependent Properties of Proteins. *J. Mol. Biol.* 238 (3), 415–436. doi:10.1006/jmbi.1994.1301
- Bartik, K., Redfield, C., and Dobson, C. M. (1994). Measurement of the Individual pK<sub>a</sub> Values of Acidic Residues of Hen and turkey Lysozymes by Two-Dimensional 1H NMR. *Biophysical J.* 66 (4), 1180–1184. doi:10.1016/S0006-3495(94)80900-2
- Bashford, D., and Karplus, M. (1990). pK<sub>a</sub>'s of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry* 29 (44), 10219–10225. doi:10.1021/bi00496a010
- Belmares, M. P., Rabinowitz, J. D., Liu, W., Mellins, E. D., and McConnell, H. M. (2000). pH Stability of HLA-DR4 Complexes with Antigenic Peptides. *Biochemistry* 39 (47), 14558–14566. doi:10.1021/bi001544g
- Benčina, M. (2013). Illumination of the Spatial Order of Intracellular pH by Genetically Encoded pH-Sensitive Sensors. *Sensors* 13 (12), 16736–16758. doi:10.3390/s131216736
- Berger, I., and Schaffitzel, C. (2020). The SARS-CoV-2 Spike Protein: Balancing Stability and Infectivity. *Cell Res* 30 (12), 1059–1060. doi:10.1038/s41422-020-00430-4
- Berman, H., Henrick, K., Nakamura, H., and Markley, J. L. (2007). The Worldwide Protein Data Bank (wwPDB): Ensuring a Single, Uniform Archive of PDB Data. *Nucleic Acids Res.* 35, D301–D303. doi:10.1093/nar/gkl971

- Beroza, P., Fredkin, D. R., Okamura, M. Y., and Feher, G. (1991). Protonation of Interacting Residues in a Protein by a Monte Carlo Method: Application to Lysozyme and the Photosynthetic Reaction center of Rhodospirillum rubrum. *Proc. Natl. Acad. Sci.* 88 (13), 5804–5808. doi:10.1073/pnas.88.13.5804
- Boyken, S. E., Benhaim, M. A., Busch, F., Jia, M., Bick, M. J., Choi, H., et al. (2019). De Novo design of Tunable, pH-Driven Conformational Changes. *Science* 364 (6441), 658–664. doi:10.1126/science.aav7897
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Jr., et al. (2020). Distinct Conformational States of SARS-CoV-2 Spike Protein. *Science* 369 (6511), 1586–1592. doi:10.1126/science.abd4251
- Chan, P., and Warwicker, J. (2009). Evidence for the Adaptation of Protein pH-Dependence to Subcellular pH. *BMC Biol.* 7, 69. doi:10.1186/1741-7007-7-69
- Chen, W., Morrow, B. H., Shi, C., and Shen, J. K. (2014). Recent Development and Application of Constant pH Molecular Dynamics. *Mol. Simulation* 40 (10–11), 830–838. doi:10.1080/08927022.2014.907492
- Chiche, J., Brahimi-Horn, M. C., and Pouységur, J. (2010). Tumour Hypoxia Induces a Metabolic Shift Causing Acidosis: a Common Feature in Cancer. *J. Cel Mol Med* 14 (4), 771–794. doi:10.1111/j.1582-4934.2009.00994.x
- Collet, J.-F., and Bardwell, J. C. A. (2002). Oxidative Protein Folding in Bacteria. *Mol. Microbiol.* 44 (1), 1–8. doi:10.1046/j.1365-2958.2002.02851.x
- Daniels, R., Downie, J. C., Hay, A. J., Knossow, M., Skehel, J. J., Wang, M. L., et al. (1985). Fusion Mutants of the Influenza Virus Hemagglutinin Glycoprotein. *Cell* 40 (2), 431–439. doi:10.1016/0092-8674(85)90157-6
- Demaurex, N. (2002). pH Homeostasis of Cellular Organelles. *Physiology* 17, 1–5. doi:10.1152/physiologyonline.2002.17.1.1
- Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., et al. (2011). Activation Mechanism of the  $\alpha$ -adrenergic Receptor. *Proc. Natl. Acad. Sci.* 108 (46), 18684–18689. doi:10.1073/pnas.1110499108
- Dwyer, J. J., Gittis, A. G., Karp, D. A., Lattman, E. E., Spencer, D. S., Stites, W. E., et al. (2000). High Apparent Dielectric Constants in the interior of a Protein Reflect Water Penetration. *Biophysical J.* 79 (3), 1610–1620. doi:10.1016/S0006-3495(00)76411-3

- Fields, W., and Kielian, M. (2015). Interactions Involved in pH protection of the Alphavirus Fusion Protein. *Virology* 486, 173–179. doi:10.1016/j.virol.2015.08.028
- Foit, L., George, J. S., Zhang, B. W., Brooks, C. L., 3rd, and Bardwell, J. C. A. (2013). Chaperone Activation by Unfolding. *Proc. Natl. Acad. Sci.* 110 (14), E1254–E1262. doi:10.1073/pnas.1222458110
- Fowler, N. J., Blanford, C. F., de Visser, S. P., and Warwicker, J. (2017). Features of Reactive Cysteines Discovered through Computation: from Kinase Inhibition to Enrichment Around Protein Degrons. *Sci. Rep.* 7 (1), 16338. doi:10.1038/s41598-017-15997-z
- Fraternali, F., Amodeo, P., Musco, G., Nilges, M., and Pastore, A. (1999). Exploring Protein Interiors: the Role of a Buried Histidine in the KH Module Fold. *Proteins* 34 (4), 484–496. doi:10.1002/(sici)1097-0134(19990301)34:4<484:aid-prot8>3.0.co;2-p
- Gane, P. J., Freedman, R. B., and Warwicker, J. (1995). A Molecular Model for the Redox Potential Difference between Thioredoxin and DsbA, Based on Electrostatics Calculations. *J. Mol. Biol.* 249 (2), 376–387. doi:10.1006/jmbi.1995.0303
- Gopi, S., Devanshu, D., Krishna, P., and Naganathan, A. N. (2018). pStab: Prediction of Stable Mutants, Unfolding Curves, Stability Maps and Protein Electrostatic Frustration. *Bioinformatics* 34 (5), 875–877. doi:10.1093/bioinformatics/btx697
- Greaney, A. J., Starr, T. N., and Bloom, J. D. (2021). An Antibody-Escape Calculator for Mutations to the SARS-CoV-2 Receptor-Binding Domain. *bioRxiv* 2021, 471236. doi:10.1101/2021.12.04.471236
- Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An Environment for Comparative Protein Modeling. *Electrophoresis* 18 (15), 2714–2723. doi:10.1002/elps.1150181505
- Gunner, M. R., and Baker, N. A. (2016). Continuum Electrostatics Approaches to Calculating pKas and Ems in Proteins. *Methods Enzymol.* 578, 1–20. doi:10.1016/bs.mie.2016.05.052
- Han, P., Li, L., Liu, S., Wang, Q., Zhang, D., Xu, Z., et al. (2022). Receptor Binding and Complex Structures of Human ACE2 to Spike RBD from Omicron and Delta SARS-CoV-2. *Cell* 185, 1–11. doi:10.1016/j.cell.2022.01.001
- Hass, M. A. S., and Mulder, F. A. A. (2015). Contemporary NMR Studies of Protein Electrostatics. *Annu. Rev. Biophys.* 44, 53–75. doi:10.1146/annurev-biophys-083012-130351
- Hebditch, M., and Warwicker, J. (2020). Protein-Sol pKa: Prediction of Electrostatic Frustration, with Application to Coronaviruses. *Bioinformatics* 36, 5112–5114. doi:10.1093/bioinformatics/btaa646
- Higgins, C. D., Malashkevich, V. N., Almo, S. C., and Lai, J. R. (2014). Influence of a Heptad Repeat Stutter on the pH-dependent Conformational Behavior of the central Coiled-Coil from Influenza Hemagglutinin HA2. *Proteins* 82 (9), 2220–2228. doi:10.1002/prot.24585
- Hong, Q., Han, W., Li, J., Xu, S., Wang, Y., Li, Z., et al. (2022). Molecular Basis of SARS-CoV-2 Omicron Variant Receptor Engagement and Antibody Evasion and Neutralization. *bioRxiv* 2001, 475532. doi:10.1101/2022.01.10.475532
- Huang, C., Qi, J., Lu, G., Wang, Q., Yuan, Y., Wu, Y., et al. (2016). Putative Receptor Binding Domain of Bat-Derived Coronavirus HKU9 Spike Protein: Evolution of Betacoronavirus Receptor Binding Motifs. *Biochemistry* 55 (43), 5977–5988. doi:10.1021/acs.biochem.6b00790
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Klusch, N., Murphy, B. J., Mills, D. J., Yildiz, Ö., and Kühlbrandt, W. (2017). Structural Basis of Proton Translocation and Force Generation in Mitochondrial ATP Synthase. *Elife* 6, e33274. doi:10.7554/eLife.33274
- Kuhlgert, S., Drepper, F., Fufezan, C., Sommer, F., and Hippler, M. (2012). Residues PsaB Asp612 and PsaB Glu613 of Photosystem I Confer pH-dependent Binding of Plastocyanin and Cytochrome C6. *Biochemistry* 51 (37), 7297–7303. doi:10.1021/bi300898j
- Kurotani, A., Tokmakov, A. A., Sato, K.-I., Stefanov, V. E., Yamada, Y., and Sakurai, T. (2019). Localization-specific Distributions of Protein pI in Human Proteome Are Governed by Local pH and Membrane Charge. *BMC Mol. Cel Biol* 20 (1), 36. doi:10.1186/s12860-019-0221-4
- Kyte, J., and Doolittle, R. F. (1982). A Simple Method for Displaying the Hydrophobic Character of a Protein. *J. Mol. Biol.* 157 (1), 105–132. doi:10.1016/0022-2836(82)90515-0
- Lin, J., Pozharski, E., and Wilson, M. A. (2017). Short Carboxylic Acid-Carboxylate Hydrogen Bonds Can Have Fully Localized Protons. *Biochemistry* 56 (2), 391–402. doi:10.1021/acs.biochem.6b00906
- Lobo, V. R., and Warwicker, J. (2021). Predicted pH-dependent Stability of SARS-CoV-2 Spike Protein Trimer from Interfacial Acidic Groups. *Comput. Struct. Biotechnol. J.* 19, 5140–5148. doi:10.1016/j.csbj.2021.08.049
- Malevanets, A., Chong, P. A., Hansen, D. F., Rizk, P., Sun, Y., Lin, H., et al. (2017). Interplay of Buried Histidine Protonation and Protein Stability in Prion Misfolding. *Sci. Rep.* 7 (1), 882. doi:10.1038/s41598-017-00954-7
- Mannar, D., Saville, J. W., Zhu, X., Srivastava, S. S., Berezuk, A. M., Tuttle, K. S., et al. (2021). SARS-CoV-2 Omicron Variant: ACE2 Binding, Cryo-EM Structure of Spike Protein-ACE2 Complex and Antibody Evasion. *bioRxiv*, 473380. doi:10.1101/2021.12.19.473380
- Mellman, I. (1992). The Importance of Being Acid: the Role of Acidification in Intracellular Membrane Traffic. *J. Exp. Biol.* 172, 39–45. doi:10.1242/jeb.172.1.39
- Narayan, A., Gopi, S., Lukose, B., and Naganathan, A. N. (2020). Electrostatic Frustration Shapes Folding Mechanistic Differences in Paralogous Bacterial Stress Response Proteins. *J. Mol. Biol.* 432 (17), 4830–4839. doi:10.1016/j.jmb.2020.06.026
- Narayan, A., and Naganathan, A. N. (2018). Switching Protein Conformational Substates by Protonation and Mutation. *J. Phys. Chem. B* 122 (49), 11039–11047. doi:10.1021/acs.jpcc.8b05108
- Onufriev, A. V., and Case, D. A. (2019). Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* 48, 275–296. doi:10.1146/annurev-biophys-052118-115325
- Papa, G., Mallery, D. L., Albecka, A., Welch, L. G., Cattin-Ortolá, J., Luptak, J., et al. (2021). Furin Cleavage of SARS-CoV-2 Spike Promotes but Is Not Essential for Infection and Cell-Cell Fusion. *Plos Pathog.* 17 (1), e1009246. doi:10.1371/journal.ppat.1009246
- Prabhakara, C., Godbole, R., Sil, P., Jahnavi, S., Gulzar, S.-e. -J., van Zanten, T. S., et al. (2021). Strategies to Target SARS-CoV-2 Entry and Infection Using Dual Mechanisms of Inhibition by Acidification Inhibitors. *Plos Pathog.* 17 (7), e1009706. doi:10.1371/journal.ppat.1009706
- Pyzik, M., Sand, K. M. K., Hubbard, J. J., Andersen, J. T., Sandlie, I., and Blumberg, R. S. (2019). The Neonatal Fc Receptor (FcRn): A Misnomer? *Front. Immunol.* 10, 1540. doi:10.3389/fimmu.2019.01540
- Qu, K., Xiong, X., Ciazynska, K. A., Carter, A. P., and Briggs, J. A. G. (2021). Structures and Function of Locked Conformations of SARS-CoV-2 Spike. *bioRxiv*. doi:10.1101/2021.03.10.434733
- Rajkumar, P., and Pluznick, J. L. (2018). Acid-base Regulation in the Renal Proximal Tubules: Using Novel pH Sensors to Maintain Homeostasis. *Am. J. Physiology-Renal Physiol.* 315 (5), F1187–F1190. doi:10.1152/ajprenal.00185.2018
- Robinson, M., Schor, S., Barouch-Bentov, R., and Einav, S. (2018). Viral Journeys on the Intracellular Highways. *Cell. Mol. Life Sci.* 75 (20), 3693–3714. doi:10.1007/s00018-018-2882-0
- Sazanavets, I., and Warwicker, J. (2015). Computational Tools for Interpreting Ion Channel pH-Dependence. *PLoS One* 10 (4), e0125293. doi:10.1371/journal.pone.0125293
- Scheiblbrandner, S., Breslmayr, E., Csarman, F., Paukner, R., Führer, J., Herzog, P. L., et al. (2017). Evolving Stability and pH-dependent Activity of the High Redox Potential Botrytis Aclada Laccase for Enzymatic Fuel Cells. *Sci. Rep.* 7 (1), 13688. doi:10.1038/s41598-017-13734-0
- Strauch, E.-M., Fleishman, S. J., and Baker, D. (2014). Computational Design of a pH-Sensitive IgG Binding Protein. *Proc. Natl. Acad. Sci.* 111 (2), 675–680. doi:10.1073/pnas.1313605111
- Sulea, T., Rohani, N., Baardsnes, J., Corbeil, C. R., Deprez, C., Cepero-Donates, Y., et al. (2020). Structure-based Engineering of pH-dependent Antibody Binding for Selective Targeting of Solid-Tumor Microenvironment. *MAbs* 12 (1), 1682866. doi:10.1080/19420862.2019.1682866
- Tan, Y.-J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995). Perturbed pKa-Values in the Denatured States of Proteins. *J. Mol. Biol.* 254 (5), 980–992. doi:10.1006/jmbi.1995.0670
- Teixeira, V. H., Vila-Viçosa, D., Reis, P. B. P. S., and Machuqueiro, M. (2016). pKa Values of Titrable Amino Acids at the Water/Membrane Interface. *J. Chem. Theor. Comput.* 12 (3), 930–934. doi:10.1021/acs.jctc.5b01114



- Toelzer, C., Gupta, K., Yadav, S. K. N., Borucu, U., Davidson, A. D., Kavanagh Williamson, M., et al. (2020). Free Fatty Acid Binding Pocket in the Locked Structure of SARS-CoV-2 Spike Protein. *Science* 370, 725–730. doi:10.1126/science.abd3255
- Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why Are  $\alpha$ -Natively Unfolded? Proteins Unstructured under Physiologic Conditions? *Proteins* 41 (3), 415–427. doi:10.1002/1097-0134(20001115)41:3<415:aid-prot130>3.0.co;2-7
- Vercoulen, Y., Kondo, Y., Iwig, J. S., Janssen, A. B., White, K. A., Amini, M., et al. (2017). A Histidine pH Sensor Regulates Activation of the Ras-specific Guanine Nucleotide Exchange Factor RasGRP1. *Elife* 6, e29002. doi:10.7554/eLife.29002
- Vila-Viçosa, D., Silva, T. F. D., Slaybaugh, G., Reshetnyak, Y. K., Andreev, O. A., and Machuqueiro, M. (2018). Membrane-Induced pKa Shifts in Wt-pHLIP and its L16H Variant. *J. Chem. Theor. Comput.* 14 (6), 3289–3297. doi:10.1021/acs.jctc.8b00102
- Warwicker, J. (2004). Improved pKa Calculations through Flexibility Based Sampling of a Water-Dominated Interaction Scheme. *Protein Sci.* 13 (10), 2793–2805. doi:10.1110/ps.04785604
- Warwicker, J. (2021). A Model for pH Coupling of the SARS-CoV-2 Spike Protein Open/closed Equilibrium. *Brief Bioinform* 22 (2), 1499–1507. doi:10.1093/bib/bbab056
- Warwicker, J., and Watson, H. C. (1982). Calculation of the Electric Potential in the Active Site Cleft Due to  $\alpha$ -helix Dipoles. *J. Mol. Biol.* 157 (4), 671–679. doi:10.1016/0022-2836(82)90505-8
- White, K. A., Grillo-Hill, B. K., Esquivel, M., Peralta, J., Bui, V. N., Chire, I., et al. (2018).  $\beta$ -Catenin Is a pH Sensor with Decreased Stability at Higher Intracellular pH. *J. Cell Biol* 217 (11), 3965–3976. doi:10.1083/jcb.201712041
- White, K. A., Ruiz, D. G., Szpiech, Z. A., Strauli, N. B., Hernandez, R. D., Jacobson, M. P., et al. (2017). Cancer-associated Arginine-To-Histidine Mutations Confer a Gain in pH Sensing to Mutant Proteins. *Sci. Signal.* 10 (495), eaam9931. doi:10.1126/scisignal.aam9931
- Whittaker, G. R., Daniel, S., and Millet, J. K. (2021). Coronavirus Entry: How We Arrived at SARS-CoV-2. *Curr. Opin. Virol.* 47, 113–120. doi:10.1016/j.coviro.2021.02.006
- Xiong, X., Qu, K., Ciazynska, K. A., Hosmillo, M., Carter, A. P., Ebrahimi, S., et al. (2020). A Thermostable, Closed SARS-CoV-2 Spike Protein Trimer. *Nat. Struct. Mol. Biol.* 27 (10), 934–941. doi:10.1038/s41594-020-0478-5
- Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., et al. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 183, 739–751. doi:10.1016/j.cell.2020.09.032
- Zahradnik, J., Tuekprakhon, A., Ginn, H. M., Duyvesteyn, H. M. E., Bahar, M., Khan, S., et al. (2021). Receptor Binding and Escape from Beta Antibody Responses Drive Omicron-B.1.1.529 Evolution. *bioRxiv*, 471045. doi:10.1101/2021.12.03.471045
- Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S. M., et al. (2021a). Structural Impact on SARS-CoV-2 Spike Protein by D614G Substitution. *Science* 372 (6541), 525–530. doi:10.1126/science.abf2303
- Zhang, J., Xiao, T., Cai, Y., Lavine, C. L., Peng, H., Zhu, H., et al. (2021b). Membrane Fusion and Immune Evasion by the Spike Protein of SARS-CoV-2 Delta Variant. *Science* 374, 1353. doi:10.1126/science.abl9463
- Zheng, Z., and Gunner, M. R. (2008). Analysis of the Electrochemistry of Hemes with E(m)s Spanning 800 mV. *Proteins* 75 (3), 719–734. doi:10.1002/prot.22282
- Zhou, T., Tsybovsky, Y., Gorman, J., Rapp, M., Cerutti, G., Chuang, G.-Y., et al. (2020). Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host & Microbe* 28 (6), 867–879. doi:10.1016/j.chom.2020.11.004
- Zhou, T., Wang, L., Misasi, J., Pegu, A., Zhang, Y., Harris, D. R., et al. (2021). Structural Basis for Potent Antibody Neutralization of SARS-CoV-2 Variants Including B.1.1.529. *bioRxiv* 2027, 474307. doi:10.1101/2021.12.27.474307

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Warwicker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.