

# Critical Mutations of the SARS-CoV-2 Virus

L. RIDGWAY SCOTT

University of Chicago, Chicago, Illinois 60637, USA

ARIEL FERNANDEZ

Daruma Institute for Applied Intelligence, AF Innovation GmbH, 1005 Oakhurst Ave, High Point, NC 27262  
CONICET–Argentine National Research Council,  
Buenos Aires, ARGENTINA

**Abstract:** SARS-CoV-2 presents an opportunity to understand better the role of viral mutations. The Alpha and Delta variants of SARS-CoV-2 provide particular insight. We argue that looking at the mutations through a physical chemistry lens provides a deeper understanding of viral evolutionary trends. We advocate here the use of quantitative (mathematical) methods, based on physical chemistry foundations, to analyze viruses. The behavior of viral proteins depends both on structural properties (how the protein sidechains are configured in three dimensional space) and epistuctural properties (how the protein interacts with the enveloping solvent, e.g., water). In both cases, physical chemistry (and ultimately quantum mechanics) plays a dominant role. There are many barriers to entry for quantitative scientists (e.g., mathematicians) to study viruses. At the simplest level, there are multiple ways to describe a virus, by its genomic sequence (RNA or DNA) or its protein sequence. Many papers assume that the context is clear when the word ‘sequence’ is used, but the novice would be forgiven for confusion. But there are much more complicated issues of terminology and interpretation that can make it very hard to understand what is going on. The book [1] was written in part to clarify this. We use the SARS-CoV-2 virus mutations here as the basis for a primer on the tools in [1] and to exhibit the kinds of observations they can yield. We attempt here both to lower the barrier of entry to the subject and to raise the level of rigor in the discussion by showing how a much more quantitative view can be beneficial. We do this by explaining concepts in simple, quantitative terms. In many cases, this involves measuring distances between atoms in PDB files. Thus we quantify what it means to be a hydrogen bond, a salt bridge, to be underwrapped, all of which have rigorous definitions [1]. Our goal here is to look at virus mutation from a mathematical perspective, with a particular focus on the SARS-CoV-2 virus [2]. This is for two reasons. First of all, it has become one of the greatest threats to humanity of all time. But the second is more positive: SARS-CoV-2 is very widely studied, and this allows new opportunities for understanding viruses in general. Any advances could have a very wide impact. We can compare and contrast two closely related viruses by considering mutations of a single virus. This allows us to focus on particular features and their impact on disease. Certain mutations of the SARS-CoV-2 virus have been the focus of attention, especially those in the spike protein [3]. We will limit our attention here to this protein for simplicity. While pure genomic sequence analysis [4] is extremely valuable, and certainly mathematically rigorous, we will stress here a different approach. Instead we focus on the amino acid sequence with a physical chemistry perspective. This brings in new mathematical tools that have yet to be fully utilized [1]. Typical sequence analysis sees all sidechains as the same, whereas physical chemistry allows us to differentiate them, to study their interactions, and to quantify epistuctural behavior as well. Using standard sequence analysis helps us understanding what mutations are important, but adding a physical chemistry perspective informs us why they are important. Thus we can see that all mutations are not created equal, but some can be viewed as a “smoking gun” in certain contexts. We examine three mutations in detail, explaining why they play a significant role in enhancing the effectiveness of SARS-CoV-2 mutants. Each one has a different physical chemistry signature, and taken together they provide a blueprint for analysing viruses, and proteins, in general.

**KeyWords:** Viral behavior, viral behavior, Amino acid mutations, Wrapping analysis

Received: March 29, 2021. Revised: November 18, 2021. Accepted: December 15, 2021. Published: January 27, 2022.

## 1 Infectivity and Transmissibility

Infectivity and transmissibility are two key factors that determine the impact of a virus. Infectivity means the ability of a virus to penetrate the human cell and hijack cellular machinery to replicate itself. Transmissibility means the ability of a virus to move from one individual to another. There are many views on the infectivity and transmissibility of SARS-CoV-2. The infectivity mechanism of SARS-CoV-2 is generally accepted to involve the S protein binding to the ACE2 receptor. One papers says “like SARS-CoV, SARSCoV- 2 enters host

cells through the interaction of spike glycoprotein (S protein) and host angiotensin-converting enzyme 2 (ACE2) receptor” [5]. Similarly, the furin cleavage site is thought to be “responsible for its high infectivity and transmissibility” [6]. Thus there is not just one critical area of the protein responsible for its virulence. This is further exemplified by the mutation D614G [7] which is in neither the protein binding cite of S for the ACE2 receptor nor the furin cleavage cite. Indeed, there may be many mutations of interest not yet fully explicated. And there may yet be more to come in future variants yet to evolve. Thus all we can do is to comprehend better the ones that have been identified.

## 2 Amino acid mutations

Mutations in proteins do not occur directly. Rather, they are the result of mutations in RNA or DNA that then get translated into protein. Thus not all protein-level mutations are equally likely. Single-point mutations at the genomic level in viruses are common. But you have to read the genetic code [1, Figure 8.6] to know what are the resulting residues in the protein.

For example, the mutation D (Asp) to G (Gly) requires only a single-point mutation. Similarly, P (Pro) to H (His) requires only a single-point mutation. However, the mutation P (Pro) to R (Arg) requires an intermediary genomic mutation. For example, H (His) to R (Arg) requires only a single-point mutation. This may explain why the Delta variant of SARS-CoV-2 was slow to emerge. The mutation P681R occurs in the Delta variant [8] of SARS-CoV-2, also known as lineage B.1.617.2.

In Table 1, we present data on mutations of SARS-CoV-2 leading to the five current Variants of Concern (VOC), as defined by the World Health Organization. This overview shows that most mutations only occur in a single variant. One mutation occurs in all current VOCs, as well as all current and former variants of interest (VOI), as defined by the World Health Organization. In addition, one mutation occurs in four VOCs, three mutations occur in three VOCs, and three mutations occur in two VOCs. We explain the impact of most of the mutations that occur in multiple variants.

See Table 2 regarding the naming of the variants.

## 3 D to G

The mutation D614G on the spike (S) protein has been the focus of many studies [11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. It is widely recognized that the mutation D614G on the spike (S) protein strengthens “viral survivability” [5]. It appears in all four VOCs listed in Table 1.

In Table 1, the column  $n$  indicates the number of VOCs that have a mutation at that residue. Horizontal lines delineate different regions of the protein. The lowest-numbered residues are in the N-terminal domain, followed by the RBD, residues 333 to 526. The subsequent mutations are in S1 but not in the RBD. The mutation at D614 is in a region where contacts are made with another copy of the S protein in a trimeric structure. The furin cleavage site has only two mutations in it. The remaining mutations are in S2. There is an insertion (replacement of) R214 by three residues EPE in the Omicron variant.

The paper [12] provides an example of the type of analysis we are advocating. It uses the concept of wrapping, which has been extensively explained [1], to explore epistructural effects. It also examines the role of salt bridges and hydrogen bonds, and thus it provides a starting point that introduces several quantitative measures of protein interactions.

### 3.1 Wrapping analysis

Briefly, the change from D to G reduces the number of wrappers (hydrophobic carbonaceous groups) from the vicinity of a particular mainchain-mainchain, intramolecular, hydrogen bond, the principle type of bond that supports protein structure. But in addition, it removes a salt bridge between D and a nearby R that frees the arginine to rotate, potentially further reducing wrapping. This double whammy makes the implicated hydrogen bond vulnerable to water attack.

Underwrapped hydrogen bonds may be viewed as magnets for agents (nonpolar groups) that can protect them from water. Thus they become binding sites. Why this is important for SARS-CoV-2 is explained in [12]. Thus we do not repeat the arguments here. Instead, we focus on the quantitative details of wrapping in this case, as this has not been presented before for the D614G mutation.

### 3.2 Wrapping defined

The concept of wrapping of a hydrogen bond is quite simple. We count the number of nonpolar carbonaceous groups within a cutoff distance of the donor and acceptor. More precisely, we measure the distance to the C-alpha carbon (CA) for each donor and acceptor sidechain. By considering large groups of hydrogen bonds, it has been established that having less than 19 wrappers for a cutoff distance of 6.5 Ångstroms is problematic.

In Table 3 we list all of the atoms proximal to the donor and acceptor atoms for the underwrapped hydrogen bond featured in [12]. There are three copies of the S protein in the PDB file 6vxx, reflecting the fact that the protein forms a trimeric structure. We are focusing on the hydrogen bond in copy A, but it has 3 wrappers from copy B, indicating the close proximity of the two copies.

We see that one atom (number 3960) appears in both spheres of influence. Thus there are only 18 wrappers for this hydrogen bond. Therefore this is already a marginal hydrogen bond in the native virus, and the mutation takes it beyond the pale.

### 3.3 No hydrogen bond

It is also observed in [12] that D614G does not form a sidechain hydrogen bond with the hydroxyl group in T859, contrary to a previous suggestion referenced in [12]. Again, this can be understood by simply computing distances between atoms in the 6vxx PDB structure. The two oxygens in ASP A 614 are greater than 6 Ångstroms from the oxygen OG1 in THR B 859. This distance is not nearly small enough to be considered an O-H-O hydrogen bond [1, Chapter 6], as the O-H distance would be at best 5 Ångstroms. Moreover, sidechain-sidechain hydrogen bonds, even when the distance is favorable, must be viewed cautiously since the sidechains are free to move, potentially disrupting the hydrogen bond.

### 3.4 Salt-bridge network

As noted previously, the D614G mutation eliminates a putative salt bridge. This could potentially lead to a substantial charge imbalance. But salt bridges often occur as a network, with charge matching occurring with multiple sidechains. We list in Table 4 the distances from the CZ atom in ARG A 646, a reasonable candidate for the center of positive charge, to negatively charged side chains within a distance of 15 Ångstroms. We see that the mutation D614G eliminates only one of the charge partners for ARG A 646 within the trimer. Thus the effect of the mutation D614G on charge imbalance is less critical than might occur in other cases.

### 3.5 All variants

As indicated in Table 1, the mutation D614G occurs in all of the mutants listed there. But in addition it is found as well in the current and former variants of interest Epsilon, Eta, Iota, Kappa, Lambda, Mu, Theta, and Zeta.

### 3.6 Conclusions about D614G

We have seen that a quantitative analysis of the mutation D614G allows clarification of its physical chemistry role. We were able to rule out a putative hydrogen bond that had been asserted in the literature, explain in more detail the dehydron effect, and make an assessment of the charge imbalance caused by the mutation. Analysis at this level of detail is not found in typical papers

on SARS-CoV-2, and this can lead to unnecessary confusion.

## 4 Furin cleavage site

The furin cleavage site has a signature protein sequence RRAR in positions 682–685. These arginines are in a part of the S protein which has the amino acid sequence PRRA at positions 681–684. This sequence is present in SARS-CoV-2 [21] but is not present in the purported [22] precursor bat virus CoV RaTG13 [23], or other known viruses (see e.g. [24, Table 1] and [25, Figure 3B]), in the region of focus regarding possible gain of function [6].

Furin cleavage occurs after the RRAR signature sequence, and before the subsequent S (serine). Thus this part of SARS-CoV-2 is often written

....NSPRRAR|S....,

where the vertical bar indicates the cleavage point.

The role of the cleavage is well understood. The spike protein has two subunits: S1 and S2. S1 contains the receptor binding domain discussed subsequently in section 5. S2 contains the peptide responsible for fusing the membranes of virion and human cells. The spike protein has to be cleaved by furin at the S1/S2 junction to function properly [26].

### 4.1 The role of charge

The role of the arginines in the furin cleavage site is well recognized. Arginine is a proton acceptor, and this makes it basic. The term ‘polybasic’ is used to indicate the multiple R’s [27] in the furin cleavage site of SARS-CoV-2. The Delta variant of SARS-CoV-2 adds one additional R near the cleavage point, making it even more basic.

The Alpha variant (B.1.1.7) [28] has the mutation P681H [29]. Like R, histidine (H) can be positively charged, indicating that the mutated site is more basic than in the native virus, although less basic than in Delta (see [1, Table 4.2] and the discussion in [1, Section 4.5]).

In the context of the furin cleavage site, the concepts of basic and acidic can be understood simply in terms of charge. The definition of the basic-acidic pH scale is in terms of proton affinity. Charge in sidechains is a result of this affinity and thus can be used as a quantitative marker. Proline (P) is not charged and Arginine (R) is typically positively charged at normal pH. Histidine is not always positively charged at normal pH, but it can be positive under suitable conditions.

There is a quantitative measure of the charge of a sidechain called  $pK_a$ . This is the pH at which half of the residues would be in each of the two protonation states, charged or uncharged. The  $pK_a$  of His is 6.5, so at normal pH 7, or anywhere near there, one would have

a substantial probability of H being positive charged. We can thus think of H having a partial positive charge at normal pH. So it is certainly more basic than proline, although not as basic as arginine. As observed in [30], there is a natural positivity (basicity) scale for the residues

$$P \rightarrow H \rightarrow R$$

that correlates with the infectivity scale

$$\text{SARS-CoV-2} \rightarrow \text{Alpha} \rightarrow \text{Delta}.$$

## 4.2 The end of the road?

Several other mutations of the furin cleavage site have been extensively studied [27]. It is said that the furin cleavage site for SARS-CoV-2 is suboptimal, since there are other such sites, e.g., in “a highly pathogenic H5N1 avian influenza haemagglutinin containing seven basic amino acids” [27]. So it is conceivable that yet another mutation near the furin cleavage site could lead to greater basicity and greater harm. Indeed, this obtains in the Omicron variant, which has the mutations P681H and N679K. But the mutation P681R is at the end of the road: there is not another more basic sidechain for R to mutate to.

## 5 RBD

The receptor binding domain (RBD) is on the S1 subunit of the S protein, residues 333 to 526. It is postulated that “infectivity is proportional to the binding affinity between SARS-CoV-2 spike glycoprotein (S protein) and host ACE2 receptor” [5]. The binding free energy is estimated computationally in [5, 34, 35, 36, 37] for various mutations in the RBD. Experimental studies on many mutations in the RBD have been done in [11]. A systematic study of all possible mutations has been done experimentally in [33]; both expression and binding affinities were measured.

### 5.1 Charge imbalance

Not all mutations are subtle to analyze. The mutation E484K in the RBD appears [9] simply to make the charge interaction with the ACE2 receptor more favorable. Going from negatively charged to positively charged is a substantial change, and it likely makes a substantial improvement. Experimental confirmation of the change in binding energy is found in [33, Figure 3B].

Such major modifications to charge interactions can also lead to conformational changes which support “formation of new hydrogen bonds due to E484K mutations [and] provides additional stability to the RBD-hACE2 binding” [9]. We leave as exercises to

examine the putative new hydrogen bonds using the approach explained in section 3.3. The E484K mutation occurs in both the B.1.351 (Beta) and P.1 (Gamma) variants of SARS-CoV-2, but not in the more virulent strains Alpha and Delta. Perhaps further analysis is needed to explain this.

### 5.2 Charge matching

Similarly, improved intermolecular salt bridges and a sidechain hydrogen bond are listed as major contributors to the binding energy as the result of the mutation N501Y [36]. Experimental confirmation of the change in binding energy is found in [33, Figure 3B], although this figure also predicts that mutation to F and V, instead of Y, would also be favorable. The sidechains F, V, and Y are among the best wrappers, so it is possible that wrapping also plays a role in this mutation. We also leave this as an exercise for further study. Intriguingly, the N501Y mutation does not appear in the Delta variant, although it is in Alpha, Beta, and Gamma. Again, maybe this omission is significant in some way.

### 5.3 Something else

Even with powerful new tools, we cannot expect to explain everything. The effect of the K417N mutation, which occurs in the Beta and Gamma variants, has been independently studied in [37, Table 1] and in [38, Figure 1]. Interestingly, this mutation has little beneficial effect on binding affinity according to the data in [33, Figure 3B].

### 5.4 Computational limitations

Computational simulations have certain limitations, as they require estimating pair potentials for atoms that in reality interact quantum mechanically. One particular difficulty is in estimating the non-bonded (van der Waals) interaction energies [39]. It would be an interesting exercise to compare the various computational simulations with the experimental results in [33].

## 6 Enter Omicron

We have already observed in section 4.2 that the Omicron variant may optimize the furin cleavage site with increased positive charge. But there are other unique mutations (among VOCs) harbored by Omicron that have been previously been identified as being of significant interest. We consider two of them in the RBD.

The Omicron variant exhibits a mutation N440K which “produced ten times higher infectious viral titers than a prevalent A2a strain, and over 1000 folds higher

titers than a much less prevalent A3i strain prototype in Caco2 cells” [31]. It is known that this mutation increases binding to ACE2 [33, Figure 3].

Omicron is the first variant of concern to harbor the S477N mutation. This occurs in a loop region [32]. It is known that this mutation increases binding to ACE2 [33, Figure 3].

It is also known that the mutation G339D increases binding to ACE2 [33, Figure 3]. On the other hand, the mutations Q493R, G496S, Q498R, and Y505H are not more beneficial to binding, suggesting another role for the changes of charge.

## 7 Additional topics of interest

There are many additional topics that could be covered. Here we mention briefly a few of them.

### 7.1 Mink mutations

Additional spike protein mutations have been found in other species. For example, Y453F, F486L, and N501T have been found in mink. These are of concern because the “recurrent emergence of these mutations in phylogenetically distant lineages, and their relative scarcity in SARS-CoV-2 samples circulating in humans, further supports these as strong candidates for ongoing host-adaptation of the virus to transmission in minks. Indeed, many of the cases identified in humans seem putatively linked to mink farm outbreaks in the Netherlands and Denmark” [40]. It is known that the mutation Y453F increases binding to ACE2 [33, Figure 3].

### 7.2 Evolutionary confusion

The use of the evolution model is sometimes a cause for misunderstanding. The fact that it is a model and not gospel is often lost in educational debates. But it is a very useful model provided we use it properly.

In one paper, it is said that “SARS-CoV-2 virus is expected to continue evolving over time in human populations” [11]. This might be interpreted to mean that the mutations could go on forever, wandering aimlessly in sequence space. But we can think of the virus as trying to solve an optimization problem, seeking the best interaction with the host to insure the viability of the virus. Once it gets to a local minimum, it may stay resident for a very long time. And if it gets to a global minimum, it will not stray very much. Mutations will continue, but since they will not be beneficial, they will not survive. So the dynamic picture is more complicated.

## 7.3 Antibody evasion

The optimization problem that a virus faces as it evolves is multifaceted. We have discussed optimizing for binding affinity and infectivity. But another concern is the attempt by a virus to evade antibody detection and immunosurveillance in general, as observed in the Delta Plus variant [8]. The effect of mutations on antibody interactions with SARS-Cov-2 are studied in [38].

A far more subtle strategy for immune evasion is glycosylation, a post-translational modification that greatly modulates interaction with the natural and vaccine-induced immune repertoire [41]. Glycosylation is also exploited by the influenza and other viruses to evade immune surveillance [42]. This is precisely the reason why vaccines against flu virus or SARS-CoV-2 are far less efficient than those against viruses that do not adopt this evasion strategy [43]. The camouflaging patterns for this “sugar coating” are not predictable with current physico-chemical approaches, and their effect on protein folding are very complicated and cannot currently be assessed.

## 7.4 Therapeutics

Improved understanding of virus structure and function can be used to design new antiviral drugs and vaccines [19, 41, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55]. The type of analysis described here can improve the precision of the process of therapeutic agent development. Most of the current epitopes currently adopted are not efficient because of their susceptibility to get glycosylated [41]. Alternative stages in the maturation of the virion need to be considered to identify epitopes that are not subject to glycosylation [54].

## 8 Conclusions

Using quantitative tools based on physical chemistry can yield more precise understanding of viral behavior. It facilitates comparing and contrasting viral mutations. Using such tools earlier in the pipeline in the future could make predictions of mutation impact more reliable.

## References

- [1] L. Ridgway Scott and Ariel Fernández. *A Mathematical Approach to Protein Biophysics*. Springer-Verlag, 2017.
- [2] Cody B Jackson, Michael Farzan, Bing Chen, and Hyeryun Choe. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, pages 1–18, 2021.
- [3] William T Harvey, Alessandro M Carabelli, Ben Jackson, Ravindra K Gupta, Emma C Thomson,

- Ewan M Harrison, Catherine Ludden, Richard Reeve, Andrew Rambaut, Sharon J Peacock, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, 19(7):409–424, 2021.
- [4] Lucy van Dorp, Mislav Acman, Damien Richard, Liam P Shaw, Charlotte E Ford, Louise Ormond, Christopher J Owen, Juanita Pang, Cedric CS Tan, Florencia AT Boshier, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, 83:104351, 2020.
- [5] Jiahui Chen, Rui Wang, Menglun Wang, and Guo-Wei Wei. Mutations strengthened SARS-CoV-2 infectivity. *Journal of Molecular Biology*, 432(19):5212–5226, 2020.
- [6] Shuai Xia, Qiaoshuai Lan, Shan Su, Xinling Wang, Wei Xu, Zezhong Liu, Yun Zhu, Qian Wang, Lu Lu, and Shibo Jiang. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal transduction and targeted therapy*, 5(1):1–3, 2020.
- [7] Cody B Jackson, Lizhou Zhang, Michael Farzan, and Hyeryun Choe. Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. *Biochemical and Biophysical Research Communications*, 538:108–115, 2021.
- [8] Saathvik R Kannan, Austin N Spratt, Alisha R Cohen, S Hasan Naqvi, Hitendra S Chand, Thomas P Quinn, Christian L Lorson, Siddappa N Byraredy, and Kamal Singh. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *Journal of Autoimmunity*, 124:102715, 2021.
- [9] Nilesh Joshi, Adish Tyagi, and Sandeep Nigam. Molecular level dissection of critical spike mutations in SARS-CoV-2 variants of concern (VOCs): A simplified review. *ChemistrySelect*, 6(31):7981, 2021.
- [10] Lucy G Thorne, Mehdi Bouhaddou, Ann-Kathrin Reuschl, Lorena Zuliani-Alvarez, Ben Polacco, Adrian Pelin, Jyoti Batra, Matthew VX Whelan, Manisha Ummadi, Ajda Rojc, et al. Evolution of enhanced innate immune evasion by the sars-cov-2 b. 1.1. 7 uk variant. *bioRxiv*, 2021.
- [11] Qianqian Li, Jiajing Wu, Jianhui Nie, Li Zhang, Huan Hao, Shuo Liu, Chenyan Zhao, Qi Zhang, Huan Liu, Lingling Nie, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*, 182(5):1284–1294, 2020.
- [12] Ariel Fernández. Structural impact of mutation D614G in SARS-CoV-2 spike protein: enhanced infectivity and therapeutic opportunity. *ACS medicinal chemistry letters*, 11(9):1667–1670, 2020.
- [13] Yujiro Toyoshima, Kensaku Nemoto, Saki Matsumoto, Yusuke Nakamura, and Kazuma Kiyotani. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of human genetics*, 65(12):1075–1082, 2020.
- [14] Jessica A Plante, Yang Liu, Jianying Liu, Hongjie Xia, Bryan A Johnson, Kumari G Lokugamage, Xianwen Zhang, Antonio E Muruato, Jing Zou, Camila R Fontes-Garfias, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592(7852):116–121, 2021.
- [15] Adam S Lauring and Emma B Hodcroft. Genetic variants of SARS-CoV-2—what do they mean? *Jama*, 325(6):529–531, 2021.
- [16] Hsu-Yu Chen, Chun Huang, Lu Tian, Xiaoli Huang, Chennan Zhang, George N Llewellyn, Geoffrey L Rogers, Kevin Andresen, Maurice RG O’Gorman, Ya-Wen Chen, et al. Cytoplasmic tail truncation of SARS-CoV-2 spike protein enhances titer of pseudotyped vectors but masks the effect of the D614G mutation. *Journal of Virology*, pages JVI–00966, 2021.
- [17] Sirwan MA Al-Jaf, Sherko Subhan Niranj, and Zana Hameed Mahmood. Rapid, inexpensive methods for exploring SARS CoV-2 D614G mutation. *medRxiv*, 2021.
- [18] Utsav Pandey, Rebecca Yee, Lishuang Shen, Alexander R Judkins, Moiz Bootwalla, Alex Ryutov, Dennis T Maglinte, Dejerianne Ostrow, Mimi Precit, Jaclyn A Biegel, et al. High prevalence of SARS-CoV-2 genetic variation and D614G mutation in pediatric patients with COVID-19. *Open Forum Infectious Diseases*, 8(6), 11 2021. ofaa551.
- [19] Kakon Nag, Juwel Chandra Baray, Maksudur Rahman Khan, Asif Mahmud, Jikrul Islam, Sanat Myti, Rostum Ali, Enamul Haq Sarker, Samir Kumar, Mobarak Hossain Chowdhury, et al. An mRNA-based vaccine candidate against SARS-CoV-2 elicits stable immuno-response with single dose. *Vaccine*, 39(28):3745–3755, 2021.
- [20] Neha Periwal, Shravan B Rathod, Ranjan Pal, Priya Sharma, Lata Nebhnani, Ravi P Barnwal, Pooja Arora, Kinshuk Raj Srivastava, and Vikas Sood. In silico characterization of mutations circulating

- in SARS-CoV-2 structural proteins. *Journal of Biomolecular Structure and Dynamics*, pages 1–16, 2021.
- [21] Bryan A Johnson, Xuping Xie, Birte Kalveram, Kumari G Lokugamage, Antonio Muruato, Jing Zou, Xianwen Zhang, Terry Juelich, Jennifer K Smith, Lihong Zhang, et al. Furin cleavage site is key to SARS-CoV-2 pathogenesis. *BioRxiv*, 2020.
  - [22] Yuri Deigin and Rossana Segreto. SARS-CoV-2's claimed natural origin is undermined by issues with genome sequences of its relative strains: Coronavirus sequences RaTG13, MP789 and RmYN02 raise multiple questions to be critically addressed by the scientific community. *BioEssays*, page 2100015, 2021.
  - [23] Ariel Fernández. Molecular biology clues portray SARS-CoV-2 as a gain-of-function laboratory manipulation of bat CoV RaTG13. *ACS Medicinal Chemistry Letters*, 12(6):941–942, 2021.
  - [24] Laura E Frazier, Bailey Lubinski, Tiffany Tang, Susan Daniel, Javier A Jaimes, and Gary R Whittaker. Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A. 23.1 identified in Uganda. *bioRxiv*, 2021. <https://doi.org/10.1101/2021.06.30.450632> doi:10.1101/2021.06.30.450632.
  - [25] Rossana Segreto, Yuri Deigin, Kevin McCairn, Alejandro Sousa, Dan Sirotkin, Karl Sirotkin, Jonathan J Couey, Adrian Jones, and Daoyu Zhang. Should we discount the laboratory origin of covid-19? *Environ Chem Lett*, 19:2743–2757, 2021.
  - [26] Yiran Wu and Suwen Zhao. Furin cleavage sites naturally occur in coronaviruses. *Stem cell research*, 50:102115, 2021.
  - [27] Thomas P Peacock, Daniel H Goldhill, Jie Zhou, Laury Baillon, Rebecca Frise, Olivia C Swann, Ruthiran Kugathasan, Rebecca Penn, Jonathan C Brown, Raul Y Sanchez-David, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nature Microbiology*, pages 1–11, 2021.
  - [28] Jingsong Zhang, Yang Zhang, Jun-Yan Kang, Shuiye Chen, Yongqun He, Benhao Han, Mo-Fang Liu, Lina Lu, Li Li, Zhigang Yi, et al. Potential transmission chains of variant B.1.1.7 and co-mutations of SARS-CoV-2. *Cell Discovery*, 7(1):1–10, 2021.
  - [29] Suman Pokhrel, Benjamin R Kraemer, Scott Burkholz, and Daria Mochly-Rosen. Natural variants in SARS-CoV-2 Spike protein pinpoint structural and functional hotspots with implications for prophylaxis and therapeutic strategies. *Scientific Reports*, 11(1):1–10, 2021.
  - [30] Ariel Fernández and L. Ridgway Scott. Origins of the SARS Corona Virus 2. Research Report UC/CS TR-2021-11, Dept. Comp. Sci., Univ. Chicago, 2021.
  - [31] Dixit Tandel, Divya Gupta, Vishal Sah, and Krishnan Harinivas Harshan. N440k variant of sars-cov-2 has higher infectious fitness. *bioRxiv*, 2021.
  - [32] Amit Singh, Georg Steinkellner, Katharina Köchl, Karl Gruber, and Christian C Gruber. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. *Scientific reports*, 11(1):1–11, 2021.
  - [33] Tyler N Starr, Allison J Greaney, Sarah K Hilton, Daniel Ellis, Katharine HD Crawford, Adam S Dingens, Mary Jane Navarro, John E Bowen, M Alejandra Tortorici, Alexandra C Walls, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, 182(5):1295–1310, 2020.
  - [34] Ahmed L Alaofi and Mudassar Shahid. Mutations of SARS-CoV-2 RBD may alter its molecular structure to improve its infection efficiency. *Biomolecules*, 11(9):1273, 2021.
  - [35] Abbas Khan, Tauqir Zia, Muhammad Suleman, Taimoor Khan, Syed Shujait Ali, Aamir Ali Abbasi, Anwar Mohammad, and Dong-Qing Wei. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *Journal of cellular physiology*, 2021.
  - [36] Fedaa Ali, Amal Kasry, and Muhamed Amin. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. *Medicine in drug discovery*, 10:100086, 2021.
  - [37] Filip Fratev. The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A free energy of perturbation study. *BioRxiv*, 2020.
  - [38] Meng Yuan, Deli Huang, Chang-Chun D Lee, Nicholas C Wu, Abigail M Jackson, Xueyong Zhu, Hejun Liu, Linghang Peng, Marit J van Gils, Rogier W Sanders, et al. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science*, 2021.

- [39] Eric Cancès and L. Ridgway Scott. van der Waals interactions between two hydrogen atoms: The Slater-Kirkwood method revisited. *SIAM Journal on Mathematical Analysis*, 50(1):381–410, 2018.
- [40] Lucy van Dorp, Cedric CS Tan, Su Datt Lam, Damien Richard, Christopher Owen, Dorothea Berchtold, Christine Orengo, and François Balloux. Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv*, 2020.
- [41] Ariel Fernández. Glycosylation of SARS-CoV-2 steers evolutionary outcomes in the postvaccination phase. *ACS Pharmacology & Translational Science*, 4(1):410–412, 2021.
- [42] Yasunori Watanabe, Thomas A Bowden, Ian A Wilson, and Max Crispin. Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1863(10):1480–1497, 2019.
- [43] Deborah Chang and Joseph Zaia. Why glycosylation matters in building a better flu vaccine. *Molecular & Cellular Proteomics*, 18(12):2348–2358, 2019.
- [44] Lindsey S Jung, Tamara M Gund, and Mahesh Narayan. Comparison of binding site of remdesivir and its metabolites with NSP12-NSP7-NSP8, and NSP3 of SARS CoV-2 virus and alternative potential drugs for COVID-19 treatment. *The Protein Journal*, 39(6):619–630, 2020.
- [45] Ariel Fernández. Therapeutically targeted destabilization of the quaternary structure of the spike protein in the dominant G614 strain of SARS-CoV-2. *ACS pharmacology & translational science*, 3(5):1027–1029, 2020.
- [46] Ariel Fernández. Achilles’ heel of SARS-CoV-2 structure. *ACS Pharmacology & Translational Science*, 3(5):1030–1031, 2020.
- [47] Ariel Fernández. Targeted disassembling of SARS-CoV-2 as it gets ready for cell penetration. *ACS Medicinal Chemistry Letters*, 11(11):2055–2057, 2020.
- [48] Ariel Fernández. Defusing SARS-CoV-2: Emergency brakes in a vaccine failure scenario. *ACS Pharmacology & Translational Science*, 3(6):1425–1426, 2020.
- [49] Ariel Fernández. COVID-19 evolution in the post-vaccination phase: Endemic or extinct? *ACS Pharmacology & Translational Science*, 4(1):403–405, 2020.
- [50] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine HD Crawford, et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host & Microbe*, 29(1):44–57, 2021.
- [51] Shaojuan Zhao, Huajun Zhang, Xinglou Yang, Haiwei Zhang, Ying Chen, Yancheng Zhan, Xiaoqing Zhang, Rendi Jiang, Meiqin Liu, Lan Liu, et al. Identification of potent human neutralizing antibodies against SARS-CoV-2 implications for development of therapeutics and prophylactics. *Nature Communications*, 12(1):1–11, 2021.
- [52] Ariel Fernández. Toward the next-generation COVID-19 vaccines that circumvent antigenic drift while defusing viral infection. *ACS Pharmacology & Translational Science*, 4(2):1018–1020, 2021.
- [53] Adam Leach, F Tudor Ilca, Zulaikha Akbar, Mathieu Ferrari, Emma M Bentley, Giada Mattiuzzo, Shimobi Onuoha, Ami Miller, Hanif Ali, and Terence H Rabbitts. A tetrameric ACE2 protein broadly neutralizes SARS-CoV-2 spike variants of concern with elevated potency. *Antiviral Research*, 194:105147, 2021.
- [54] Ariel Fernández. SARS-CoV-2 glycosylation suggests that vaccines should have adopted the S1 subunit as antigen. *ACS Pharmacology & Translational Science*, 4(2):1016–1017, 2021.
- [55] Vaishali P Waman, Neeladri Sen, Mihaly Varadi, Antoine Daina, Shoshana J Wodak, Vincent Zoete, Sameer Velankar, and Christine Orengo. The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies. *Briefings in Bioinformatics*, 22(2):742–768, 2021.

## Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)

| residue | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $o$ | $n$ |
|---------|----------|---------|----------|----------|-----|-----|
| L18     |          |         | F        |          |     | 1   |
| T19     |          |         |          | R        |     | 1   |
| T20     |          |         | N        |          |     | 1   |
| P26     |          |         | S        |          |     | 1   |
| A67     |          |         |          |          | V   | 1   |
| D80     |          | A       |          |          |     | 1   |
| T95     |          |         |          |          | I   | 1   |
| D138    |          |         | Y        |          |     | 1   |
| G142    |          |         |          | D        | D   | 2   |
| R190    |          |         | S        |          |     | 1   |
| L212    |          |         |          |          | I   | 1   |
| R214*   |          |         |          |          | E   | 1   |
| D215    |          | G       |          |          |     | 1   |
| G339    |          |         |          |          | D   | 1   |
| S371    |          |         |          |          | L   | 1   |
| S375    |          |         |          |          | F   | 1   |
| K417    |          | N       | T        |          | N   | 3   |
| N440    |          |         |          |          | K   | 1   |
| G446    |          |         |          |          | S   | 1   |
| L452    |          |         |          | R        |     | 1   |
| S477    |          |         |          |          | N   | 1   |
| T478    |          |         |          | K        | K   | 2   |
| E484    |          | K       | K        |          | A   | 3   |
| Q493    |          |         |          |          | R   | 1   |
| G496    |          |         |          |          | S   | 1   |
| Q498    |          |         |          |          | R   | 1   |
| N501    | Y        | Y       | Y        |          | Y   | 4   |
| Y505    |          |         |          |          | H   | 1   |
| T547    |          |         |          |          | K   | 1   |
| A570    | D        |         |          |          |     | 1   |
| D614    | G        | G       | G        | G        | G   | 5   |
| H655    |          |         | Y        |          | Y   | 2   |
| N679    |          |         |          |          | K   | 1   |
| P681    | H        |         |          | R        | H   | 3   |
| A701    |          | V       |          |          |     | 1   |
| T716    | I        |         |          |          |     | 1   |
| N764    |          |         |          |          | K   | 1   |
| D796    |          |         |          |          | Y   | 1   |
| N856    |          |         |          |          | K   | 1   |
| D950    |          |         |          | N        |     | 1   |
| Q954    |          |         |          |          | H   | 1   |
| N969    |          |         |          |          | K   | 1   |
| L981    |          |         |          |          | F   | 1   |
| S982    | A        |         |          |          |     | 1   |
| T1027   |          |         | I        |          |     | 1   |
| D1118   | H        |         |          |          |     | 1   |
| V1176   |          |         | F        |          |     | 1   |
| total   | 7        | 7       | 12       | 7        | 30  | 63  |

Table 1: Mutations of concern. Deletions have been omitted. Based on data on SARS-CoV-2 Variants of Concern (VOC) from [9, Table 1], [10], and Tulio de Oliveira's Tweet regarding Omicron (B.1.1.529).

| $\alpha$ | $\beta$ | $\gamma$ | $\delta$  | $o$       |
|----------|---------|----------|-----------|-----------|
| Alpha    | Beta    | Gamma    | Delta     | Omicron   |
| B.1.1.7  | B.1.351 | P.1      | B.1.617.2 | B.1.1.529 |

Table 2: Mutation nomenclature for Variants of Concern.

| distance to CA | atom # | atom type | sidechain |
|----------------|--------|-----------|-----------|
| 4.52           | 3960   | CB        | GLN A 613 |
| 5.88           | 3961   | CG        | GLN A 613 |
| 1.53           | 3969   | CB        | ASP A 614 |
| 4.75           | 3977   | CB        | VAL A 615 |
| 4.65           | 3978   | CG1       | VAL A 615 |
| 5.16           | 3979   | CG2       | VAL A 615 |
| 5.69           | 3941   | CG        | LEU A 611 |
| 5.05           | 3942   | CD1       | LEU A 611 |
| 5.40           | 3943   | CD2       | LEU A 611 |
| 6.41           | 3960   | CB        | GLN A 613 |
| 5.99           | 4058   | CG2       | THR A 645 |
| 4.72           | 4063   | CB        | ARG A 646 |
| 6.14           | 4064   | CG        | ARG A 646 |
| 1.52           | 4074   | CB        | ALA A 647 |
| 5.29           | 4214   | CG2       | ILE A 666 |
| 5.07           | 4224   | CB        | ALA A 668 |
| 5.07           | 13023  | CB        | PRO B 862 |
| 3.94           | 13024  | CG        | PRO B 862 |
| 4.56           | 13025  | CD        | PRO B 862 |

Table 3: Wrappers of the Asp A 614 to ALA A 647 backbone hydrogen bond. Top are the wrappers near the CA atom of Asp A 614. Bottom are the wrappers near the CA atom of Ala A 647. Data taken from the PDB file 6vxx.

| distance to CZ | atom # | atom type | sidechain |
|----------------|--------|-----------|-----------|
| 10.82          | 3971   | OD1       | ASP A 614 |
| 12.69          | 3972   | OD2       | ASP A 614 |
| 10.43          | 13062  | OD1       | ASP B 867 |
| 9.45           | 13063  | OD2       | ASP B 867 |
| 9.09           | 13071  | OE1       | GLU B 868 |
| 11.13          | 13072  | OE2       | GLU B 868 |

Table 4: Salt-bridge network formed with Arg A 646. Data taken from the PDB file 6vxx.