SARS-CoV-2 Variants are Selecting for Spike Protein Mutations that Increase Protein Stability

David Shorthouse[1], Benjamin A. Hall[1*]

1 Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom, WC1E 6BT

* To whom correspondence should be addressed, b.hall@ucl.ac.uk

ABSTRACT: The emergence of variants of SARS-CoV-2 with mutations in their spike protein are a major cause for concern for the efficacy of vaccines, and control of the pandemic. We show that mutations in the spike protein of SARS-CoV-2 are selecting for amino acid changes that result in a more thermodynamically stable protein than expected from background. We suggest that computationally efficient analysis of mutational stability may aid in early screening of variants for potential danger

Since the emergence of SARS-CoV-2 in late 2019, over 2 million people have died as a result of infection[1]. As the global pandemic continues, the emergence of viral variants with RNA mutations is an expected phenomena, caused by random errors in RNA copying, and selected for by evolutionary pressure[2]. Variants contain mutations in the spike protein that confer an advantage to the virus, such as increased ACE2 receptor binding[3], glycosylation/cleavage site alterations[4], and immune evasion[5], as well as protein stability[6]. Understanding these properties helps infer how a variant may differ from another mutational profile, and provides insights into the mechanisms by which variants differ, such as increased infectivity or vaccine resistance[7]. The WHO classifies variants in SARS-CoV-2 into major categories, the two most important: "Variants of Concern" and "Variants of Interest" are assigned to emerging variants likely to have a different phenotype and mutational profile to the original SARS-CoV-2[8].

Changes in Gibbs Free Energy (called ΔΔG) is a measure of the thermodynamic energy change between two states, and we apply this concept to the comparison of energy change between wild type (WT) and mutant proteins. Prediction of the changes in ΔΔG are routinely used in protein engineering for optimization of enzymes or stabilisation of protein complexes[9], and we have recently shown that they can be predictive of mutations that destabilise or damage a protein in a cancer context[10–12]. Whilst stability of mutations has been assessed in the SARS-CoV-2 spike protein[13,14], variant analysis has not yet been performed. Mutations that stabilise the SARS-CoV-2 spike protein are likely to lead to a greater lifespan of a protein before thermal unfolding. The requirement for calculation of predicted ΔΔG values is protein structural information, which was recently published for the SARS-CoV-2 spike protein[15].

We calculate the ΔΔG of mutation for every possible missense mutation in the SARS-CoV-2 spike protein. With this "background" mutation rate we show that mutations to the spike protein observed in emerging SARS-CoV-2 variants have a lower ΔΔG, and a higher proportion of stabilising mutations than expected. This suggests an important role for protein stability when considering the evolution of SARS-CoV-2.

The SARS-CoV-2 spike protein is composed of a trimer of 3 identical subunits (Figure 1a) that sits in the membrane of the virion and interacts with the human ACE2 receptor to facilitate infection of a host cell. The structure of the spike protein was recently elucidated, enabling the calculation of predictive ΔΔG values for mutations. The Alpha variant (WHO designation), first identified in December 2020 in the United Kingdom[16] has been found to be more transmissible than the original virus, with an increased affinity for binding the human ACE2 receptor[17], and by April 2020 had become the most dominant variant in the UK. The Alpha variant carries 23 common mutations across its genome, 7 of which are amino acid substitution mutations in the spike protein, 6 of which are at locations for which crystallographic data is available (Figure 1b).

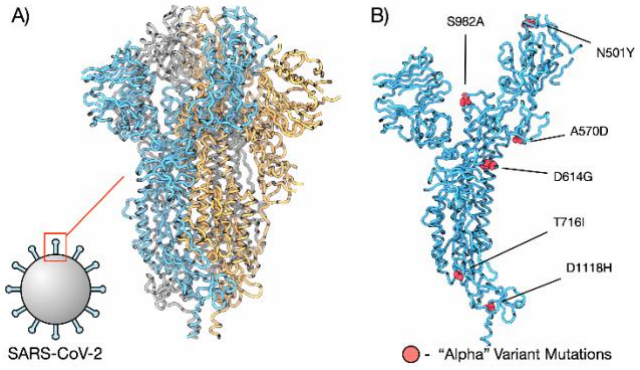

Figure 1. Structure of SARS-CoV-2 spike protein. A) Full structure of COVID spike protein structure (PDBID 6VXX), subunits are coloured blue, orange, and grey. B) Single subunit of spike protein with residues from the "Alpha" variant coloured red and labelled.

We first calculated the ΔΔG for each of the possible 19440 missense mutations in the 6VXX cryo-em structure using FoldX[18] (Table S1). As is consistent with a previous study[10], we find mutations that stabilise the protein are rare (Figure 2a). We define mutations with an induced ΔΔG of < -1 kcal/mol as strongly stabilising, and mutations with a ΔΔG > 2.5 kcal/mol as strongly destabilising, with those between zero and each threshold described as mildy stabilising and destabilising respectively. Only 767 (3.9%) of possible mutations are predicted to strongly stabilise the protein, and only 3699 (19%) have a ΔΔG < 0. With this "background" distribution, we compared to mutations found only in WHO "Variants of Concern" and "Variants of Interest" as of June 2021 (Figure 2b). Mutations found in both categories have a significantly lower ΔΔG (t-test pvalue < 0.05) than bulk population, indicating that variants may be evolutionarily selecting for stabilising or non-destabilising mutations. Considering individual mutations found in "Variants of Concern" (Figure 2c), none of the mutations observed induce a ΔΔG > 2.5 kcal/mol (defined as strongly destabilising), significantly different to the expected 34% (chi squared pvalue <0.05). Additionally 4 of 17 mutations have ΔΔG of <= ~-1 kcal/mol (N501Y, H655Y, T716I, and T1027I), showing a significant enrichment for stabilising mutations in the "Variants of Concern" (chi squared pvalue <0.05). Grouping mutants by spatial location, we find mutations in the regions involved in interaction with host proteins (n-terminus domain – NTD, and receptor binding domain- RBD) are generally destabilising (Figure S1). This is in line with these mutations generally altering binding affinity of the spike protein to the human ACE2 receptor, thus being more accommodating to changes that decrease spike protein stability[19]. There is a statistical enrichment for mutations that stabilise the spike protein compared to the mutational background in SARS-CoV-2 variants.

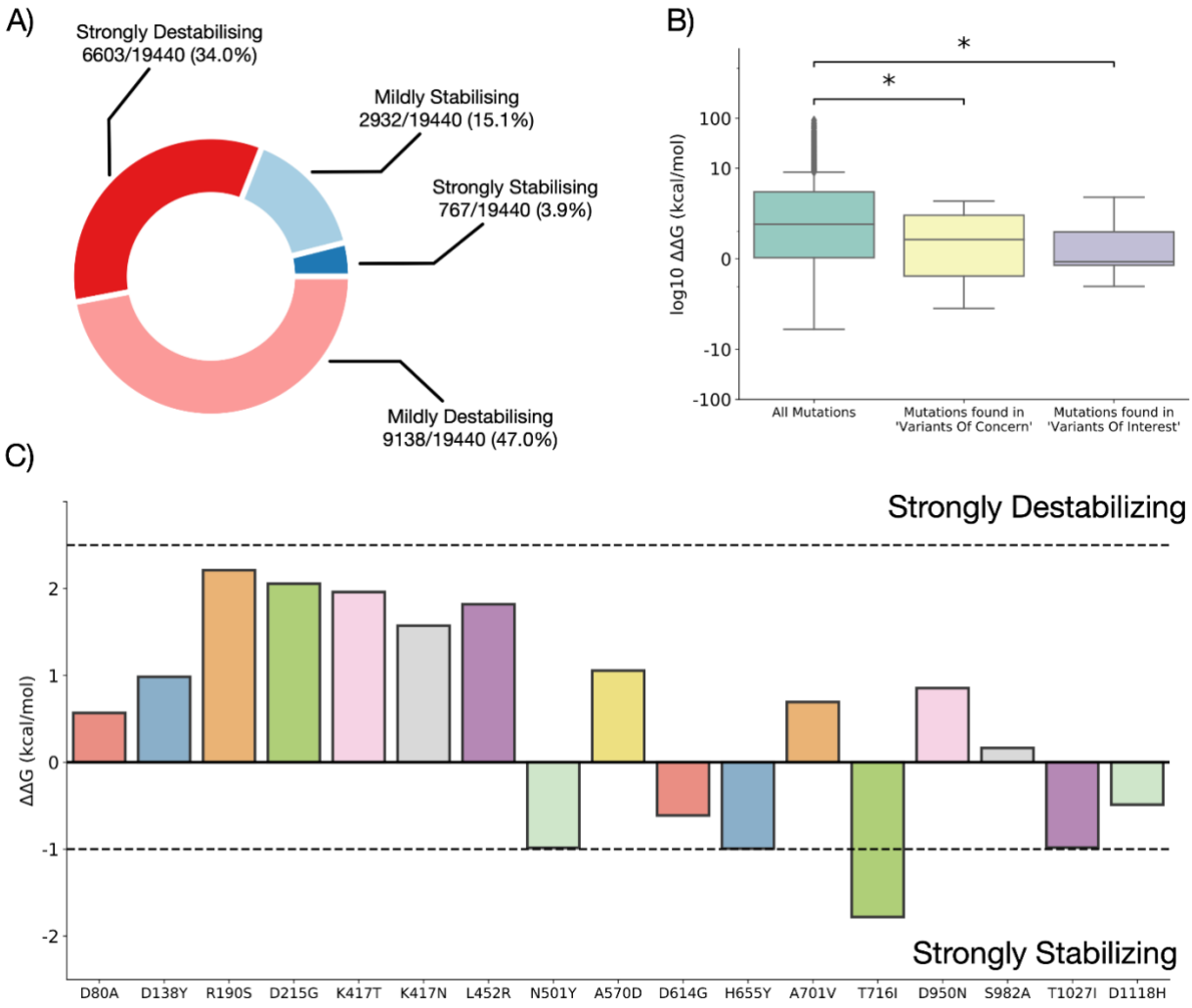

Figure 2. Saturation screen of SARS-CoV-2 spike protein. A) Proportion of mutations that are strongly stabilising (ΔΔG < -1 kcal/mol), strongly destabilising (ΔΔG > 2.5 kcal/mol), mildly stabilising (0 > ΔΔG < -1 kcal/mol, or mildly destabilising (0 < ΔΔG < 2.5 kcal/mol). B) ΔΔG for all 19440 mutations compared to those in WHO "Variants of Concern" and "Variants of Interest". * represents ttest

pvalue <= 0.05. C) ΔΔG (kcal/mol) for all mutations observed in WHO "Variants of Concern".

We next calculated the ΔΔG distribution for mutations in each variant (Figure 3, variants and mutations included in Table S2). Of the 10 variants studied, 7 have a statistically significantly lower ΔΔG than the bulk mutational background (ttest pvalue <=0.05), and 5 variants (Alpha, Gamma, Eta, Theta, and Iota) have a mean ΔΔG less than 0, indicating that the protein will be stabilised with respect to the original variant. We next calculated the expected ΔΔG for each variant given the number of mutations occurring in it (Figure S2), and find that all variants aside from Beta and Eta have a lower ΔΔG than expected.

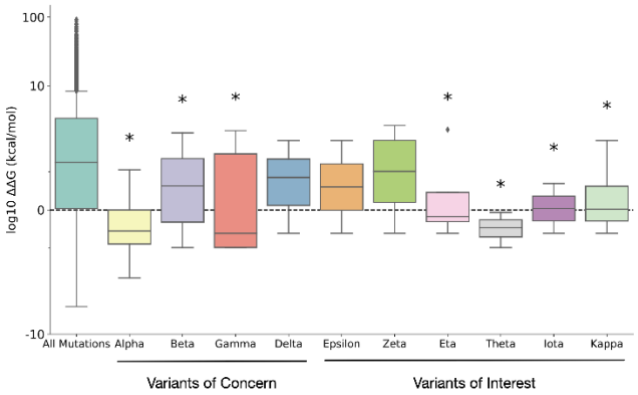

Figure 3. ΔΔG (kcal/mol) for all mutations observed in WHO "Variants of Concern" and "Variants of Interest" compared to all 19440 possible mutations in the SARS-CoV-2 spike protein. * represents ttest pvalue <=0.05.

Finally, to study the potential evolutionary order and gain insights into mechanisms of mutational selection, we calculated the ΔΔG for every combination of mutations in each variant, (Figures S3-S12). For the Alpha variant there is a consistent trend towards stabilisation, with all combinations of 5 or more mutations resulting in a predicted stabilisation of the protein with respect to the original. We observe combinations that result in a positive ΔΔG, which are likely to be evolutionarily less favourable (when considering stability alone), and expect that these combinations would be less likely to occur in the evolutionary history than stabilising combinations. Furthermore, some variants, such as the Beta variant first identified in South Africa in May 2020, contain combinations of mutations with a ΔΔG expected to be highly destabilising, and whilst the final ΔΔG of all mutations is still predicted to be strongly destabilising, it is lower than the most extreme combinations. A potential driver of selection of some mutants may be that they stabilise the protein complex enough for it to function, whilst retaining the advantageous properties unrelated to stability, such as ACE2 receptor binding, that are otherwise destabilising for the protein structure.

This work highlights that mutations with a stabilising effect on the SARS-CoV-2 spike protein are one of the key drivers of evolution of the virus. That variants are more stable than expected by chance shows that evolution favours mutations with a stabilising effect, and it may be that mutations that destabilise a protein but have other influences, such as K417N[17], which alters ACE2 binding affinity, are offset or preceded by mutations that stabilise the structure. We note however, that not all mutations in all variants can be considered, due to missing regions of the cryo-em structure, and as such this study does not necessarily represent the true ΔΔG for each variant.

We further note that a large number of mutations are found within flexible unresolved loops within the protein, and what these mutations may be impacting is not necessarily known. Furthermore, we study only the structure in its "closed" conformation as we feel this is the most physiologically relevant of the existing structures, but further work will need to address the impact of the dynamics of the structure on mutational stability. A final potential confounder to this study is that we do not know the role of glycosylation site mutation on protein stability. Despite these confounders, we highlight that stability of the SARS-CoV-2 spike protein is an important consideration for future study of variants, and is likely one of a number of driving forces in the evolution of the virus. Finally, we suggest that efficient folding calculations of newly sequenced variants such as those within this study (taking only ~150 CPU hours to screen every possible mutation in the spike protein), offer a computationally inexpensive method to highlight advantageous mutations to surveil for.

ASSOCIATED CONTENT

Supporting Information.
The following files are available free of charge.
Supplementary Information: Methods for ΔΔG calculation and supplementary figures (PDF)
Table S1: Predicted ΔΔG for every possible mutations in SARS-CoV-2 structure PDBID 6VXX. (XLSX)
Table S2: SARS-CoV-2 Variants of Concern (Alpha, Beta, Gamma, and Delta), and Variants of Interest (Epsilon, Zeta, Eta, Theta, Iota, and Kappa) as of June 2021. (XLSX)

AUTHOR INFORMATION

Corresponding Author – Benjamin A Hall, b.hall@ucl.ac.uk

Present Addresses – Department of Medical Physics and Biomedical Engineering, UCL, London, United Kingdom, WC1E 6BT

Author Contributions

DS and BAH conceived the study and wrote the manuscript. DS generated all data and performed all analysis.

Funding Sources

This work was supported by the Medical Research Council (grant no. MR/S000216/1). B.A.H. acknowledges support

from the Royal Society (grant no. UF130039). The authors declare no competing financial interest.

DATA AND SOFTWARE AVAILABILITY

All methodological details required to reproduce data in this manuscript, and all data generated in this study are available in the supporting information. Scripts and example input/output to reproduce all calculations in this work are available at https://github.com/shorthouse-mrc/COVID_structure.

REFERENCES

(1) Peacock, T. P.; Penrice-Randal, R.; Hiscox, J. A.; Barclay, W. S. SARS-CoV-2 One Year on: Evidence for Ongoing Viral Adaptation. *J. Gen. Virol.* **2021**, *102*. https://doi.org/10.1099/jgv.0.001584.

(2) Plante, J. A.; Mitchell, B. M.; Plante, K. S.; Debbink, K.; Weaver, S. C.; Menachery, V. D. The Variant Gambit: COVID-19's next Move. *Cell Host Microbe* **2021**, *29*, 508–515. https://doi.org/10.1016/j.chom.2021.02.020.

(3) Ozono, S.; Zhang, Y.; Ode, H.; Sano, K.; Tan, T. S.; Imai, K.; Miyoshi, K.; Kishigami, S.; Ueno, T.; Iwatani, Y.; Suzuki, T.; Tokunaga, K. SARS-CoV-2 D614G Spike Mutation Increases Entry Efficiency with Enhanced ACE2-Binding Affinity. *Nat. Commun.* **2021**, *12*, 848. https://doi.org/10.1038/s41467-021-21118-2.

(4) Tortorici, M. A.; Veesler, D. Structural Insights into Coronavirus Entry. In *Adv. Virus Res.* **2019**; Vol. 105, pp 93–116. https://doi.org/10.1016/bs.aivir.2019.08.002.

(5) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S. J.; Robertson, D. L. SARS-CoV-2 Variants, Spike Mutations and Immune Escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. https://doi.org/10.1038/s41579-021-00573-0.

(6) Berger, I.; Schaffitzel, C. The SARS-CoV-2 Spike Protein: Balancing Stability and Infectivity. *Cell Res.* **2020**, *30*, 1059–1060. https://doi.org/10.1038/s41422-020-00430-4.

(7) Planas, D.; Bruel, T.; Grzelak, L.; Guivel-Benhassine, F.; Staropoli, I.; Porrot, F.; Planchais, C.; Buchrieser, J.; Rajah, M. M.; Bishop, E.; Albert, M.; Donati, F.; Prot, M.; Behillil, S.; Enouf, V.; Maquart, M.; Smati-Lafarge, M.; Varon, E.; Schortgen, F.; Yahyaoui, L.; Gonzalez, M.; de Sèze, J.; Péré, H.; Veyer, D.; Sève, A.; Simon-Lorière, E.; Fafi-Kremer, S.; Stefic, K.; Mouquet, H.; Hocqueloux, L.; van der Werf, S.; Prazuck, T.; Schwartz, O. Sensitivity of Infectious SARS-CoV-2 B.1.1.7 and B.1.351 Variants to Neutralizing Antibodies. *Nat. Med.* **2021**, *27*, 917–924. https://doi.org/10.1038/s41591-021-01318-5

(8) SARS-CoV-2 Variant Classifications and Definitions https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html (accessed 2021 -06 -14).

(9) Eijsink, V. G. H.; Bjørk, A.; Gåseidnes, S.; Sirevåg, R.; Synstad, B.; Burg, B. van den; Vriend, G. Rational Engineering of Enzyme Stability. In *J. Biotechnol.* **2004**; Vol. 113, 105–120. https://doi.org/10.1016/j.jbiotec.2004.03.026.

(10) Shorthouse, D.; Hall, M. W. J.; Hall, B. A. Computational Saturation Screen Reveals the Landscape of Mutations in Human Fumarate Hydratase. *J. Chem. Inf Model.* **2021**, *61*, 1970–1980. https://doi.org/10.1021/acs.jcim.1c00063.

(11) Lee, M.; Shorthouse, D.; Mahen, R.; Hall, B. A.; Venkitaraman, A. R. Cancer-Causing *BRCA2* Missense Mutations Disrupt an Intracellular Protein Assembly Mechanism to Disable Genome Maintenance. *Nucleic Acids Res.* **2021**, *49,* 5588–5604. https://doi.org/10.1093/nar/gkab308.

(12) Fowler, J. C.; King, C.; Bryant, C.; Hall, M. W. J.; Sood, R.; Ong, S. H.; Earp, E.; Fernandez-Antoran, D.; Koeppel, J.; Dentro, S. C.; Shorthouse, D.; Durrani, A.; Fife, K.; Rytina, E.; Milne, D.; Roshan, A.; Mahububani, K.; Saeb-Parsy, K.; Hall, B. A.; Gerstung, M.; Jones, P. H. Selection of Oncogenic Mutant Clones in Normal Human Skin Varies with Body Site. *Cancer Discov.* **2021**, *11*, 340–361. https://doi.org/10.1158/2159-8290.CD-20-1092.

(13) Teng, S.; Sobitan, A.; Rhoades, R.; Liu, D.; Tang, Q. Systemic Effects of Missense Mutations on SARS-CoV-2 Spike Glycoprotein Stability and Receptor-Binding Affinity. *Brief. Bioinform.* **2021**, *22*, 1239–1253. https://doi.org/10.1093/bib/bbaa233.

(14) Smaoui, M. R.; Yahyaoui, H. Unraveling the Stability Landscape of Mutations in the SARS-CoV-2 Receptor-Binding Domain. *Sci. Rep.* **2021**, *11*, 9166. https://doi.org/10.1038/s41598-021-88696-5.

(15) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(16) Public Health England. Investigation of Novel SARS-COV-2 Variant Variant of Concern 202012/01. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf (accessed 2021 -06 -14).

(17) Ramanathan, M.; Ferguson, I. D.; Miao, W.; Khavari, P. A. SARS-CoV-2 B.1.1.7 and B.1.351 Spike Variants Bind Human ACE2 with Increased Affinity. *Lancet Infect. Dis.* **2021**, *21*, 1070. https://doi.org/10.1016/S1473-3099(21)00262-0.

(18) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382–W388. https://doi.org/10.1093/nar/gki387.

(19) Thomson, E. C.; Rosen, L. E.; Shepherd, J. G.; Spreafico, R.; da Silva Filipe, A.; Wojcechowskyj, J. A.; Davis, C.; Piccoli, L.; Pascall, D. J.; Dillen, J.; Lytras, S.; Czudnochowski, N.; Shah, R.; Meury, M.; Jesudason, N.; de Marco, A.; Li, K.; Bassi, J.; O'Toole, A.; Pinto, D.; Colquhoun, R. M.; Culap, K.; Jackson, B.; Zatta, F.; Rambaut, A.; Jaconi, S.; Sreenu, V. B.; Nix, J.; Zhang, I.; Jarrett, R. F.; Glass, W. G.; Beltramello, M.; Nomikou, K.; Pizzuto, M.; Tong, L.; Cameroni, E.; Croll, T. I.; Johnson, N.; di Iulio, J.; Wickenhagen, A.; Ceschi, A.; Harbison, A. M.; Mair, D.; Ferrari, P.; Smollett, K.; Sallusto, F.; Carmichael, S.; Garzoni, C.; Nichols, J.; Galli, M.; Hughes, J.; Riva, A.; Ho, A.; Schiuma, M.; Semple, M. G.; Openshaw, P. J. M.; Fadda, E.; Baillie, J. K.; Chodera, J. D.; Rihn, S. J.; Lycett, S. J.; Virgin, H. W.; Telenti, A.; Corti, D.; Robertson, D. L.; Snell, G. Circulating SARS-CoV-2 Spike N439K Variants Maintain Fitness While Evading Antibody-Mediated Immunity. *Cell* **2021**, *184*. https://doi.org/10.1016/j.cell.2021.01.037

For Table of Contents Only