Supporting Information

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

David Shorthouse, Benjamin A. Hall*

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

*Corresponding author Email: B.Hall@ucl.ac.uk
Experimental procedures:

ΔΔG Calculation:

To study the mutational landscape of the SARS-CoV-2 spike protein from PDBID 6VXX, the structure was initially relaxed and repaired using the RepairPDB command in Foldx as follows:

```
$foldx --command=RepairPDB --pdb=6vxx.pdb --ionStrength=0.05 --pH=7 --vdwDesign=2
```

RepairPDB was repeated on the structure six times to minimize its energy. The relaxed structure was then used to calculate the ΔΔG. PositionScan was run on each residue in the protein structure sequentially using the following command:

```
$foldx --command=PositionScan --pdb=6vxx_repaired.pdb --ionStrength=0.05 --pH=7 --vdwDesign=2 --pdbHydrogens=false --positions=100
```

To run PositionScan on the 100th residue. PositionScan mutates a target residue sequentially from wildtype (WT) to each amino acid possibility, calculating the ΔΔG relative to wildtype each time. The protein backbone is unchanged, but the energy cost or gain from inducing a different side chain is measured. Histidine protonation state is calculated in each case from the input pH (7) and the surrounding side chains.

Mutations:

Mutations in SARS-CoV-2 variants were obtained from CoVariants3 (https://covariants.org/).

Expected mutational ΔΔG:

To calculate the expected mutational ΔΔG for a variant (Figure S1), 1,000,000 samples of the same number of mutations in the variant were taken from the structure. For each sample the ΔΔG was calculated and the median of the distribution taken as the expected value. The value observed for the variant was removed from the expected to generate the ΔΔG difference.

Mutational ΔΔG combinations:

To calculate the ΔΔG for combinations of mutations in each variant, every possible combination of mutations in each variant was calculated. Each combination was then generated 15 times and average ΔΔG calculated using the Foldx BuildModel command:

```
$foldx --command=BuildModel --pdb=6vxx_repaired.pdb --mutant-file=mutantfile.txt --numberOfRuns=15 --pH=7 --vdwDesign=2 --ionStrength=0.05
```

Where mutant-file.txt is a file containing the mutational combination to be modelled separated by a comma. For example, to model mutations L452R, D614G, and D950N in the Delta variant the file would contain:

```
LA452R,DA614G,DA950N;
```

Supplementary Figures:
Figure S1: Mutational $\Delta \Delta G$ for mutations coloured by location in the spike protein. (NTD – N-terminal domain, RBD – Receptor Binding Domain, HR1 – Heptapeptide repeat 1).

Figure S2: Difference between median expected $\Delta \Delta G$ for each variant and observed $\Delta \Delta G$ (Kcal/mol)
Figure S3: Upset plot for mutation combinations in SARS-CoV-2 Alpha variant.

Figure S4: Upset plot for mutation combinations in SARS-CoV-2 Beta variant
Figure S5: Upset plot for mutation combinations in SARS-CoV-2 Gamma variant.

Figure S6: Upset plot for mutation combinations in SARS-CoV-2 Delta variant.
Figure S7: Upset plot for mutation combinations in SARS-CoV-2 Epsilon variant.

Figure S8: Upset plot for mutation combinations in SARS-CoV-2 Zeta variant.
Figure S9: Upset plot for mutation combinations in SARS-CoV-2 Eta variant.

Figure S10: Upset plot for mutation combinations in SARS-CoV-2 Theta variant.
**Figure S11:** Upset plot for mutation combinations in SARS-CoV-2 Iota variant.

**Figure S12:** Upset plot for mutation combinations in SARS-CoV-2 Kappa variant.
Table S1: Table containing predicted ΔΔG for every possible mutations in SARS-CoV-2 structure PDBID 6VXX (available as XLSX)
| WHO Label | PANGO Lineage | Location Identified | Mutations Present |
|-----------|--------------|---------------------|------------------|
| **Alpha** | B.1.1.7       | United Kingdom      | N501Y, A570D, D614G, P681R* |
| **Beta**  | B.1.526       | United States       | D80A, H655Y, D614G, A717V |
| **Gamma** | P.1.2         | Brazil              | E484K*, L452R, D614G, E1092K, H1101Y, V1176F* |
| **Delta** | B.1.617.2     | India               | S13I*, W152C*, L452R, D614G, D64G, H655Y, T106A, V1176F* |
| **Epsilon** | B.1.427      | United States       | Q52R, A67V, E484K*, D614G, G677R* |
| **Zeta**  | P.3.2         | Brazil              | E484K*, F565L, D614G, V1176F* |
| **Eta**   | B.1.525       | Multiple Countries  | L18F*, T20N*, P681R* |
| **Theta** | B.1.17.1      | United Kingdom      | E484K*, N501Y, D614G, P681R* |
| **Iota**  | B.1.526       | United States       | L5F*, T95I*, D418H* |
| **Kappa** | B.1.617.1     | India               | G142D, E154K*, L452R, E484Q*, D614G, P681R* |

* indicates a mutated residue is not included in the 6VXX structure.
**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.

**References:**

(1) 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 (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

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

(3) Emma B. Hodcroft. CoVariants: SARS-CoV-2 Mutations and Variants of Interest https://covariants.org/ (accessed 2021-06-15).