# Electronic Supplementary Information

## Immune Evasion of SARS-CoV-2 Variants of Concern is Driven by Low Affinity to Neutralizing Antibodies

*Matheus V. F. Ferraz*[a, b, ¶], *Emerson G. Moreira*[a,b,¶], *Danilo F. Coêlho*[a,b, ¶], *Gabriel Luz Wallau*[a,&] *and Roberto D. Lins*[a,&,*] *on behalf of Fiocruz COVID-19 Genomic Surveillance Network*

[a]*Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, PE, Brazil*
[b]*Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE, Brazil*
¶ These authors contributed equally to this work
& These authors contributed equally to this work

*Corresponding author: roberto.lins@cpqam.fiocruz.br

## 1. Experimental values for $\Delta\Delta G_{Bind}$ between RBD and hACE2 upon single mutations

Starr et al.[1] obtained deep mutational scan data for all sites of interaction between SARS-CoV-2 RBD and hACE2, transformed to values of $\Delta\log(K_D)$ according to the following relation:

$$\Delta\log\left(K_D\right) = \log\left(K_{D,WT}\right) - \log\left(K_{D,MUT}\right) \quad \text{eq. 1}$$

where $K_D$ is the dissociation constant and WT are values for wild-type and MUT for mutated variant. To transform the values into differences in binding free-energy, the following mathematical relationship was used:

$$\Delta\Delta G_{Bind} = \Delta G_{MUT} - \Delta G_{WT} = RT\left[ln(K_{D,MUT}) - ln(K_{D,WT})\right] = RTln\left(\frac{K_{D,MUT}}{K_{D,WT}}\right) = -RTln\left(\frac{K_{D,WT}}{K_{D,MUT}}\right) \quad \text{eq. 2}$$

From eq. 1 we have:

$$\Delta log(K_D) = log\left(\frac{K_{D,WT}}{K_{D,MUT}}\right) \therefore \frac{K_{D,WT}}{K_{D,MUT}} = 10^{\Delta log(K_D)} \quad \text{eq. 3}$$

Substituting eq.3 in eq. 2, we then have:

$$\Delta\Delta G_{Bind} = -RT\ln\left(10^{\Delta\log\left(K_D\right)}\right) = -\Delta\log\left(K_D\right)\cdot RT\ln\left(10\right) \quad \text{eq. 4}$$

Using T = 298 K (25 ºC) and R = 1.987 kcal·K$^{-1}$·mol$^{-1}$ we transformed all $\Delta log(K_D)$ experimental values into $\Delta\Delta G_{Bind}$ using eq. 4. Figure S1 shows the heat map plot for the values.

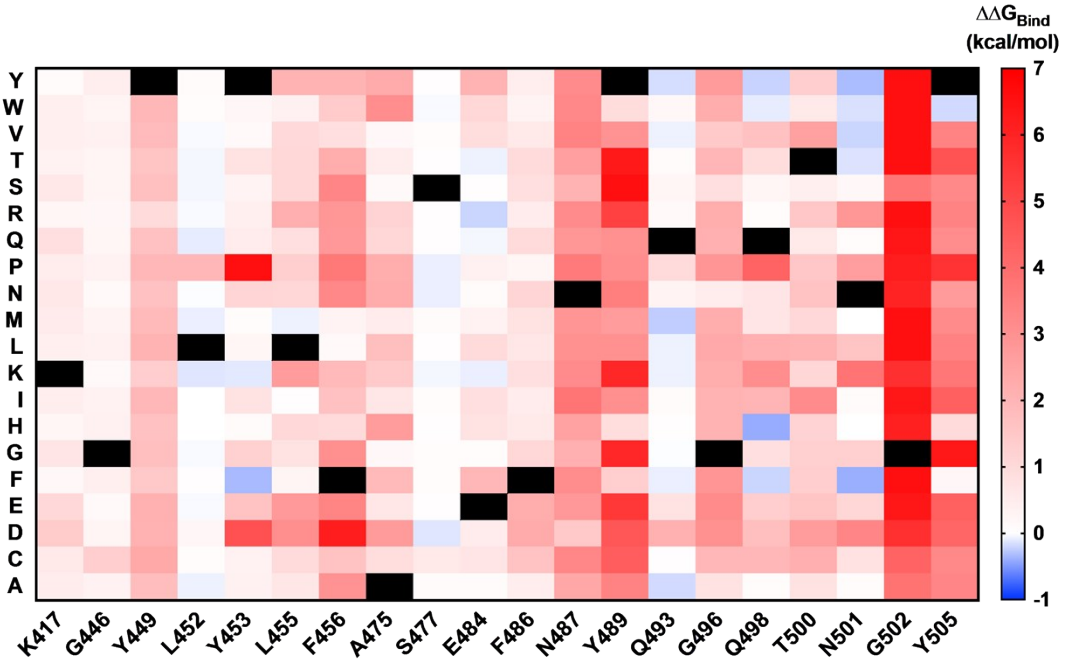



**Figure S.1 -** Heat map plot of Deep Mutational Scan ΔΔ**GBind** values converted from experimental data retrieved from Starr et al.[1] Horizontal-axis shows amino acid position in RBD and correspondent native amino acid. The vertical axis shows the identity of mutation amino acid. Values range from highly unfavorable changes (red squares) to marginally favorable change (blue), whereas black squares denote WT variant where $\Delta\Delta G_{bind}$ is not applicable.

## 2. Computational Procedure

### 2.1. Systems Preparation

Atomic coordinates for the three-dimensional structure of the complex hACE2-SARS-CoV-2 RBD were retrieved from the Protein Data Bank (PDB) under accession code 6M0J.[2] 27 crystal structures of nAbs bound to SARS-Cov-2 RBD were downloaded from the PDB. The structures were cleaned by removing crystallographic water molecules, ligands, and ions. Only structures without missing residues in the interface between RBD and the respective antibody were considered, resulting in a total of 22 structures (accession code in Table S2). The missing residues lying far from the binding interface were modeled using Swiss-Model.[3, 4] Then, 4 steps of the FastRelax protocol, as implemented in the Rosetta package v. 3.12,[5] were used to pack and minimize the side-chains and backbone's conformations. Recognized VOCs (B.1.1.7, B.1.351 and P.1) and VOIs (P.2, B.1.429, B.1.526) with known epidemiological importance and bearing a number of key mutations in the RBD of the S protein (https://cov-lineages.org/ and https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html) (Table S1)[6] that were previously evaluated in experimental studies were recovered from the GISAID database (https://www.gisaid.org/) and used to estimate binding affinities with hACE2 and nAbs. Moreover, we also included a recent emerged variant[7] (P.3 - https://cov-lineages.org/lineages/lineage_P.3.html) bearing additional Spike mutations to investigate its binding affinities to the hACE2 and nAbs as well.

**Table S.1 - Mutations in RBD of each VOC/VOI evaluated**

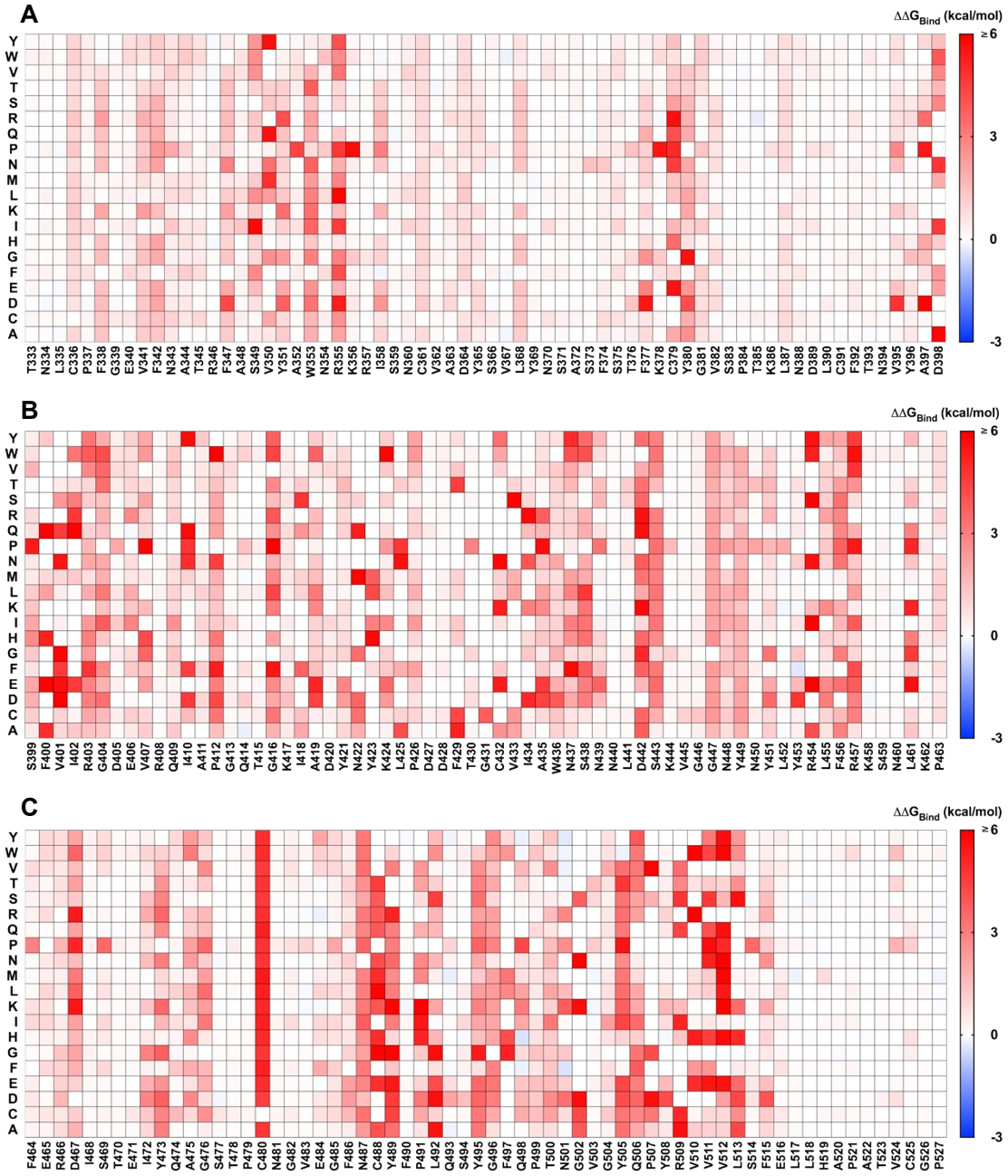
| VOC/VOI | RBD mutation | GISAID ID |
|---|---|---|
| B.1.351 | K417N E484K N501Y | EPI_ISL_968081 |
| B.1.1.7 | N501Y | EPI_ISL_601443 |
| P.1 | K417T E484K N501Y | EPI_ISL_833137 |
| B.1.429 | L452R | EPI_ISL_1402625 |
| B.1.526 | S477N E484K | EPI_ISL_1081256 |
| P.2 | E484K | EPI_ISL_792560 |
| P.3 | E484K N501Y | EPI_ISL_1122426 |

### 2.2. Binding free-energy calculations with Rosetta

To compute the change in binding free energy (ΔΔG) of hACE2-RBD and nAbs-RBD upon residues mutation, the Rosetta v.3.12[5] package was used. The Rosetta package is a prime-tool in protein modeling and design,[8, 9] and it has an empirical energy function[10, 11] that predicts mutational $^{\Delta\Delta G}$ with a Pearson's correlation coefficient of 0.72 with mean absolute error (MAE) of 1.7 kcal.mol$^{-1}$, as tested on a dataset of 1240 mutations.[12] The calculations were conducted using the scripting language interface RosettaScripts[13, 14] (code and parsed command-lines described below). The employed XML was adapted from a previously published script elsewhere.[12] The employed energy function was the Talaris2014.[11] Initially, the wild-type structure was repacked using harmonic restraints on pairwise bond lengths. Then, the structure was minimized maintaining the bond lengths restraints using a maximum 5.000 iterations and tolerance of 0.001 REU by the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm complying with the Armijo-Goldstein condition, as implemented in the Rosetta package. For these steps a constraint term of weight of 1.0 was added to the score function. The constraints were released, and the wild-type binding free energy ($\Delta G_{WT}$) was calculated by averaging over 10 repeats of the Ddg filter. Then, mutations were performed with the PackRotamersMover, and residues within a sphere of 8 Å of the mutation(s) site(s) were repacked. The same harmonic restraints were used to the mutated structure and the same level of convergence was used to minimize the geometry of the mutant. Then, upon constraints release the mutant binding free energy ($\Delta G_{MUT}$) was computed using the Ddg filter. The ΔΔG was calculated as $\Delta\Delta G = \Delta G_{MUT} - \Delta G_{WT}$. Prior to assessing the impact of the VOCs' mutations in the RBD to the $^{\Delta\Delta G}$ towards the binding to hACE2, we have tested the performance of the Rosetta energy function to predict the $^{\Delta\Delta G}$for mutations contained in the dataset from the deep mutational scanning, comprising ca. 4,000 mutations. Thus, the $^{\Delta\Delta G}$ for every single mutation was computed and compared to the experimental value as from the deep mutational scanning data. The performance of the Rosetta-predicted values for the $^{\Delta\Delta G}$ was measured using the MAE, calculated according to eq. 5, as follow:

$$MAE = \frac{\sum_{i=1}^{n} y_i - \bar{y}_i \vee}{n} \quad \text{eq. 5}$$

Where n consists of the total sample size, $y_i$ is the predicted $\Delta\Delta G$ from rosetta calculations, and $\bar{y}_i$ is the experimentally observed value for $\Delta\Delta G$ from the deep mutational scanning data.



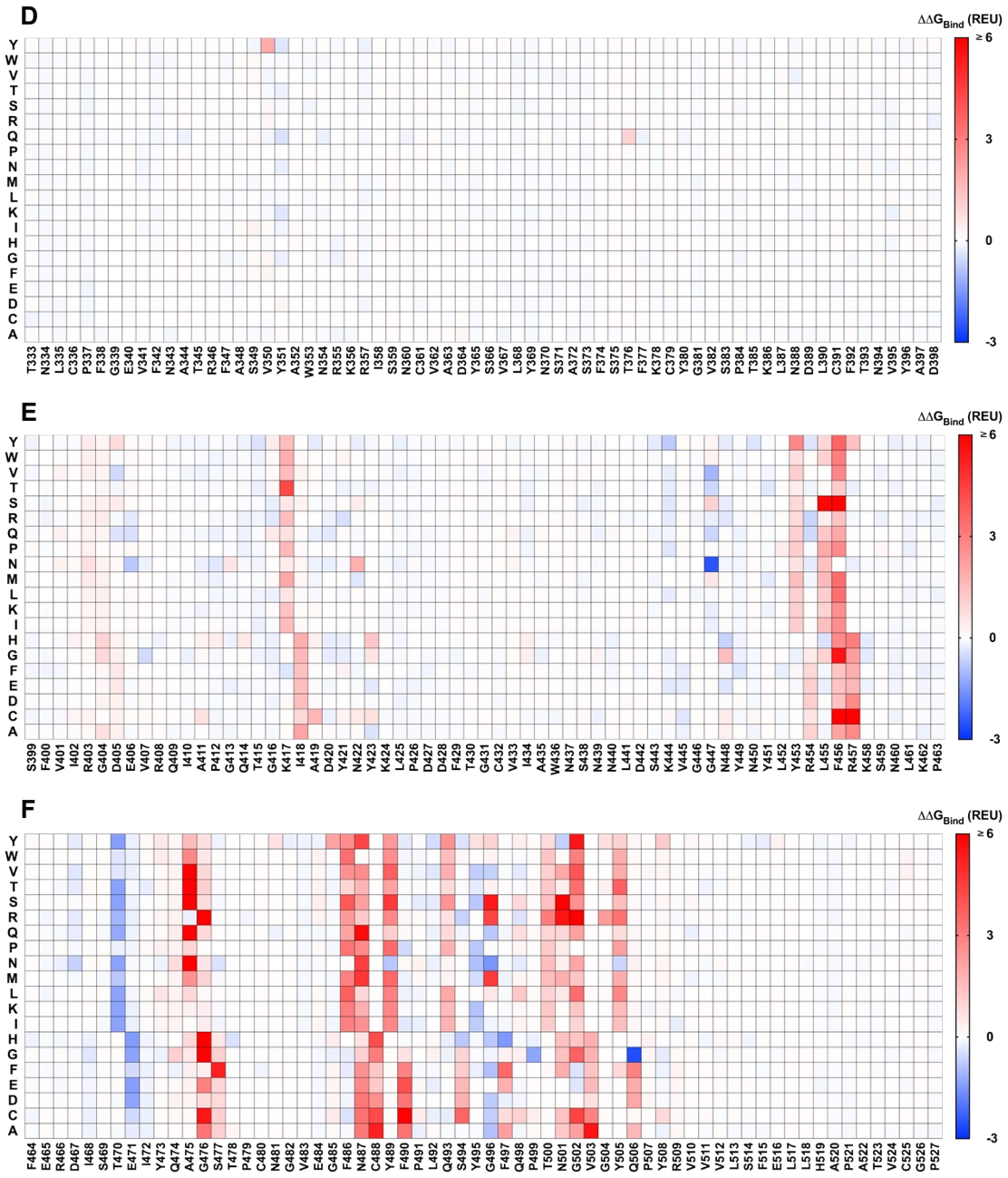



**Figure S.2** Comparison of heatmap plots for the ΔΔGBind value upon residue mutation (A-C) The shown values are derived from converted from the deep mutational scannign data from Starr et al.[1] (D-F) Predicted values calculated using the Talaris2014 energy function of the Rosetta suite of software. The x-axis indicates the code for the native amino acid in the RBD, whereas the y-axis displays the identiy of which the native amino acid was mutated. The values range from blue (moderately favorable) to red (highly disfavorable).

### 2.3. Command-lines and XML scripts for the calculations with the Rosetta package

A recent study used the Rosetta package to compute the relative ΔGBind between SARS-CoV-1 and -2 RBD targeting hACE2.[15] For these calculations, the Rosetta energy function 2015 (REF2015) was used, while the Talaris14 energy function was employed in this study. Despite parametrical improvements in the REF2015 energy function over Talaris14, we opted to use the latter since it has been extensively benchmarked for the estimation of ΔΔGBind in protein – protein binding upon mutation,[12] the task we have carried out in this work. However, it is worth noting that while out of the scope of this work, we have calculated the ΔGBind of SARS-CoV-2 RBD and hACE2 using REF2015, and obtained exactly the same result (ΔΔGBind value) as Chowdhury *et al*,[15] even though using a different combination of movers.

#### 2.3.1. Rosetta Command-lines

The following command-lines were parsed:

*Relax the PDB*: $rosetta/main/source/bin/rosetta_scripts.macosgccrelease -database rosetta/main/database/ -s *.pdb -parser:protocol relax.xml -ex1 -ex2 -use_input_sc

*Binding free-energy evaluation*:
$rosetta/main/source/bin/rosetta_scripts.macosgccrelease -database rosetta/main/database/ -s *.pdb -parser:protocol ddg.xml -parser:script_vars file=mutation*.resfile -ex1 -ex2 -use_input_sc -restore_talaris_behavior

#### 2.3.2. RosettaScripts

*Relax* (**relax.xml**)

```
<ROSETTASCRIPTS>
<MOVERS>
<FastRelax name="fstrlx" repeats="4"/>
</MOVERS>
<FILTERS>
<Geometry name="omega" omega="150" cart_bonded="100" confidence="0"/>
<Rmsd name="rmsd" confidence="0" superimpose="1"/>
</FILTERS>
<PROTOCOLS>
<Add filter_name="omega"/>
<Add mover_name="fstrlx"/>
<Add filter_name="rmsd"/>
</PROTOCOLS>
</ROSETTASCRIPTS>
```

Binding free-energy evaluation (**ddg.xml**)

```
<ROSETTASCRIPTS>
 <SCOREFXNS>
  <ScoreFunction name="fa_talaris2014" weights="talaris2014"/>
  <ScoreFunction name="fa_talaris2014_cst" weights="talaris2014">
   <Reweight scoretype="atom_pair_constraint" weight="1.0"/>
   <Set fa_max_dis="9.0"/>
  </ScoreFunction>
 </SCOREFXNS>
     <TASKOPERATIONS>
          <InitializeFromCommandline name="init"/>
          <ReadResfile name="resfile" filename="%%file%%" />
     </TASKOPERATIONS>
     <RESIDUE_SELECTORS>
  <Task name="resselector" fixed="0" packable="0" designable="1"
task_operations="resfile"/>
  <Neighborhood name="bubble" selector="resselector" distance="8.0"/>
  <PrimarySequenceNeighborhood name="bubble_adjacent" selector="bubble"
lower="1" upper="1"/>
  <StoredResidueSubset name="restore_neighbor_shell"
subset_name="neighbor_shell"/>
  <Not name="everythingelse" selector="restore_neighbor_shell"/>
     </RESIDUE_SELECTORS>
 <TASKOPERATIONS>
  <OperateOnResidueSubset name="repackonly" selector="restore_neighbor_shell">
   <RestrictToRepackingRLT/>
  </OperateOnResidueSubset>
  <OperateOnResidueSubset name="norepack" selector="everythingelse">
   <PreventRepackingRLT/>
  </OperateOnResidueSubset>
  <UseMultiCoolAnnealer name="multicool" states="6"/>
  <ExtraChiCutoff name="extrachizero" extrachi_cutoff="0"/>
  <InitializeFromCommandline name="commandline_init"/>
  <RestrictToRepacking name="restrict_to_repacking"/>
 </TASKOPERATIONS>
     <FILTERS>
          <Ddg name="ddg_filter" threshold="1000" repeats="3" jump="1"/>
     </FILTERS>
     <MOVERS>
          <MinMover name="minimize" scorefxn="fa_talaris2014_cst" chi="1"
bb="1" type="lbfgs_armijo_nonmonotone" tolerance="0.00001" max_iter="500"/>
          <StoreResidueSubset name="neighbor_shell_storer"
subset_name="neighbor_shell" residue_selector="bubble_adjacent" />
          <PackRotamersMover name="repack" scorefxn="fa_talaris2014"
task_operations="commandline_init,repackonly,norepack,multicool"/>
          <PackRotamersMover name="mut_and_pack"
task_operations="resfile,multicool,norepack"/>
          <FilterReportAsPoseExtraScoresMover name="dg_wt"
report_as="ddg_wt"  filter_name="ddg_filter" />
```

```
        <FilterReportAsPoseExtraScoresMover name="dg_mut"
report_as="ddg_mut" filter_name="ddg_filter" />
        <ScoreMover name="apply_score" scorefxn="fa_talaris2014_cst"
verbose="0"/>
        <AddConstraintsToCurrentConformationMover name="addcst"
use_distance_cst="1" coord_dev="0.5" min_seq_sep="0" max_distance="9"
CA_only="1" bound_width="0.0" cst_weight="0.0"/>
         <ClearConstraintsMover name="clearcst"/>
    </MOVERS>
    <PROTOCOLS>
        <Add mover_name="addcst"/>
        <Add mover_name="apply_score"/>
        <Add mover_name="neighbor_shell_storer"/>
        <Add mover_name="repack"/>
        <Add mover_name="addcst"/>
        <Add mover_name="minimize"/>
        <Add mover_name="clearcst"/>
        <Add mover_name="dg_wt"/>
        <Add mover_name="mut_and_pack"/>
        <Add mover_name="addcst"/>
        <Add mover_name="minimize"/>
        <Add mover_name="clearcst"/>
        <Add mover_name="dg_mut"/>
    </PROTOCOLS>
</ROSETTASCRIPTS>
```

### 2.4. Binding free-energy calculation with FoldX

For the calculations using FoldX, the relaxed input PDBs were geometry-optimized using the FoldX force field and the RepairPDB command. The binding energy of the wild-type complex ($\Delta G_{WT}$) was calculated using the AnalyseComplex command and was assumed as being corresponding to the interaction energy term. Then, mutations in the RBD for the different variants were introduced using a single run of the BuildModel. The binding energy of the mutated complex ($\Delta G_{MUT}$) was calculated with the AnalyseComplex, and the $\Delta\Delta G_{Bind}$ was taken as $\Delta\Delta G_{Bind} = (\Delta G_{MUT} - \Delta G_{WT})$. For all FoldX calculations, the pH was set as 7.0 and a solution ionic strength of 0.05M. The full command-lines were as following:

*Repair PDB*: $foldx --command=RepairPDB --pdb=*.pdb --ionStrength=0.05 --pH=7 --vdwDesign=2 --pdbHydrogens=false –water=predict

*Mutation:* $foldx --command=BuildModel --pdb=*Repaired.pdb --mutant-file=mutations_list.txt --ionStrength=0.05 --pH=7 --water=predict --vdwDesign=2 --pdbHydrogens=false

*Binding free-energy evaluation:* $foldx --command=AnalyseComplex --analyseComplexChains=A,E --pdb=*.pdb

Figure S.3 below shows the $\Delta\Delta G_{Bind}$ for the complexes hACE2-RBD and RBD-nAbs upon mutations of each lineage using FoldX force-field:

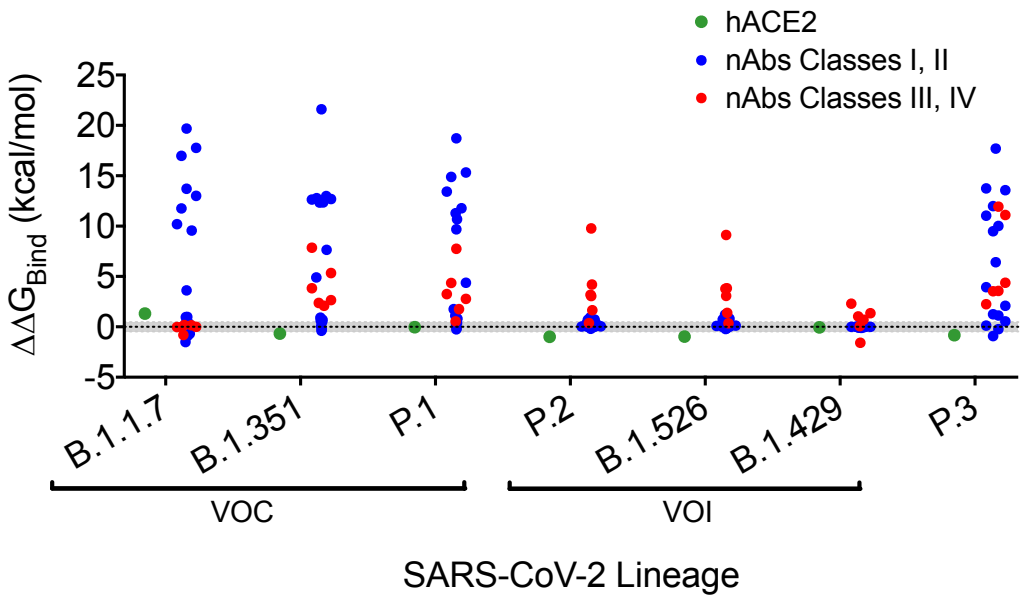



**Figure S.3 -** Variation in the binding free-energy upon mutation of each lineage for the hACE2-RBD and nAbs-RBD complexes computed using FoldX. The hatched area indicates the MAE of the FoldX energy function centered at 0.

### 2.5. Protein-protein interaction analysis

The Protein-protein interactions (PPI) analysis in the binding interface was carried out to map the interactions profile between the WT and variant RBDs with the antibodies considered in this study. The g_mindist program of the Gromacs package v. 4.6.7 was employed to identify the total number of antibodies interacting with each residue of the RBDs. To this end, a residue was considered to make any contact with the antibodies within a distance cut-off value of 4 Å. In addition, The Protein Interactions Calculator (http://pic.mbu.iisc.ernet.in/) was used to compute hydrophobic interactions, disulphide bridges, hydrogen bonds, ionic aromatic and cation-Pi interactions, by keeping all the interaction cut-off value as the default. The results for all the systems were compiled and disclosed on a .zip file.

### 2.6. Electrostatic potential calculations

Surface electrostatic potentials were plotted for the RBDs (wild-type and variants) and rendered utilizing the PyMol software (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). PDB2PQR[16] was used to generate pqr files by assigning partial atomic charges for the protein from the AMBER ff99 all-atom force field[17] at pH=7 with the solvent described by a dielectric constant of 78 and saline (NaCl) concentration of 0.15 M with ions of radius 1.5 Å. The low-dielectric cavity (εr) was assumed as 4, using Bondi atomic radii. The linearized Poisson-Boltzmann equation (LPBE) was solved using the APBS code.[18]

## 3. Antibodies dataset used for $\Delta\Delta G_{Bind}$ computational calculations

All PDB files were downloaded from RSC Protein Data Bank (https://www.rcsb.org/) and only structures determined by means of X-Ray crystallography and RBD with 100% identity to Wuhan lineage were used. Table S2 shows the structures dataset.

## 4. Data Accessibility

The rbds_nab.zip file contains i. All the PDB files of the systems utilized in this study (WT and variant RBDs in the free form; the WT and variant RBDs bound to the antibodies); ii. All provided tables from the PIC webserver depicting the computed interactions.

**Table S2. Antibodies dataset used for computational calculations**

| | | | $\Delta\Delta G_{Bind}$* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **PDB ID** | **nAb Name** | **nAb Class†** | **B.1.351** | **B.1.1.7** | **P.1** | **B.1.429** | **B.1.526** | **P.2** | **P.3** |
| 6XC4 | CC12.3 | I, II | 0.25 (0.92) | -0.59 (0.35) | -0.51 (1.11) | -0.55 (0.01) | -2.9 (0.09) | -1.6 (0.05) | -0.419 (1.12) |
| 6XKQ | CV07-250 | I, II | 0.6 (-0.24) | -0.95 (0.99) | 0.06 (-0.25) | -0.83 (0.02) | 0.39 (0.13) | -0.96 (0.39) | -0.49 (2.1) |
| 7B3O | STE90-C11 | I, II | 16.85 (12.33) | 15.8 (16.98) | 15.57 (9.68) | -0.45 (0.02) | -0.52 (0.01) | -0.76 (0.03) | 15.65 (9.49) |
| 7BEI | COVOX-150 | I, II | 15.99 (12.63) | 15.26 (11.76) | 16.5 (10.69) | -0.23 (-0.02) | 0.45 (0.81) | 0.46 (0.71) | 15.84 (11.98) |
| 7BEJ | COVOX-158& | I, II | 6.79 (7.64) | 6.36 (10.19) | 6.71 (11.29) | -0.25 (0.01) | -0.23 (0.15) | -0.12 (-0.03) | 6.26 (6.41) |
| 7BEK | COVOX-158& | I, II | 14.16 (12.77) | 14.96 (13.01) | 13.65 (14.89) | 0.08 (0.01) | -0.01 (0.03) | -0.01 (0.03) | 15.17 (13.75) |
| 7BEM | COVOX-269 | I, II | 14.53 (12.69) | 12 (13.71) | 14.74 (13.42) | -1.1 (-0.10) | -0.63 (-0.01) | -1.26 (0.02) | 11.92 (11.03) |
| 7C01 | CB6 | I, II | 1.54 (0.71) | -1.8 (-0.89) | 1.07 (1.75) | -1.02 (0.08) | -0.25 (-0.21) | -0.98 (-0.2) | -1.72 (-0.91) |
| 7CH4 | BD-604 | I, II | 12.16 (12.99) | 9.11 (19.69) | 10.84 (15.32) | -0.35 (-0.02) | 0.11 (-0.05) | -0.45 (-0.03) | 8.93 (13.56) |
| 7CH5 | BD-629 | I, II | 1.41 (0.06) | 1.05 (0.97) | 1.29 (-0.13) | -0.35 (0) | -0.94 (0.15) | -0.07 (0.35) | 1.08 (1.24) |
| 7CHB | BD-236 | I, II | 21.79 (12.35) | 17.39 (9.56) | 20.68 (11.77) | -0.96 (-0.09) | -0.56 (-0.03) | -0.47 (-0.01) | 17.4 (10.01) |
| 7K8M | C102 | I, II | 6.86 (21.6) | 15.25 (17.76) | 14.43 (18.71) | -0.03 (0.02) | 0.67 (-0.16) | -0.66 (-0.06) | 15.31 (17.7) |
| 7KLG | 15033 | I, II | 1.03 (0.49) | -0.94 (-0.69) | 2.97 (0.37) | -0.64 (-0.01) | 2.742 (0.74) | 2.99 (0.72) | 3.13 (0.12) |
| 7KLH | 15033-7 | I, II | 4.5 (0.62) | 0.3 (-1.51) | 3.8 (0.8) | -0.4 (-0.03) | 2.3 (1.22) | 2.6 (0.94) | 2.68 (-0.22) |
| 7CDI | P2C-1F11 | I, II | 0.96 (4.91) | -0.23 (3.62) | 0.83 (4.37) | -0.28 (-0.04) | 0.25 (0.3) | -0.21 (0.32) | 0.04 (3.93) |
| 7BZ5 | B38 | I, II | 0.79 (-0.39) | -0.59 (0.52) | 0.85 (0.29) | -1.25 (0) | -0.94 (0.01) | -0.58 (0.02) | -0.88 (0.55) |
| 6XKP | CV07-270 | III, IV | 4.72 (3.84) | -1.32 (0) | 4.31 (3.26) | 4.2 (2.3) | 4.51 (3.84) | 4.67 (3.18) | 4.55 (3.56) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 7BEH | COVOX-316 | III, IV | 8.45 (2.38) | -0.55 (-0.8) | 8.59 (2.78) | -1.49 (0.07) | 8.46 (3.08) | 8.2 (3.05) | 8.34 (2.25) |
| 7BWJ | P2B-2F6 | III, IV | 6.29 (2.1) | -0.9 (0) | 0.09 (1.75) | -1.11 (1.04) | 0.07 (1.38) | 0.17 (1.65) | 0.02 (11.1) |
| 7L5B | 2-15 | III, IV | 6.92 (2.66) | -2.59 (0.16) | 6.37 (0.55) | -1.2 (0.7) | 7.77 (0.35) | 7.61 (0.37) | 6.08 (3.53) |
| 7KMG | LY-CoV555 | III, IV | 11.55 (7.85) | -0.54 (-0.02) | 11.54 (7.75) | 2.32 (1.35) | 11.4 (9.13) | 14.35 (9.78) | 11.54 (11.94) |
| 7CDJ | P2C-1A3 | III, IV | 9.7 (5.35) | -1.65 (0.18) | 12.54 (4.36) | -3.35 (-1.58) | 9.83 (3.78) | 9.88 (4.2) | 9.6 (4.38) |

* Values outside parenthesis are for Rosetta (REU units); values inside parenthesis are for Foldx (kcal.mol$^{-1}$). † According to Barnes et al.[19] classes I and II comprise nAbs with epitopes that overlap with hACE2 binding site (class I binds Up/Down conformations, class II binds Up only); classes III and IV epitopes do not overlap with hACE2 binding site (class III binds Up conformation only, class IV binds Up/Down conformations). Since all PDBs evaluated in this work are for nAbs complexed with isolated RBD domain, the classifications were based by aligning to hACE2-RBD complex (PDB: 6M0J) and evaluating if the epitopes overlap with hACE2 binding site.
& The same nAb is represented in two slightly different forms of binding to RBD.

## Table S3. Acknowledgment table containing sequences used in this study

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

| Accession ID | Originating Laboratory | Submitting Laboratory | Authors |
|---|---|---|---|
| EPI_ISL_1081256 | Maryland Public Health Laboratory | Maryland Public Health Laboratory | Maryland Department of Health Laboratories Administration |
| EPI_ISL_1122426 | Cebu TB Reference Laboratory | Philippine Genome Center | Francis A. Tablizo, Kenneth M. Kim, Carlo M. Lapid, Marc Jerrone R. Castro, Maria Sofia L. Yangzon, Benedict A. Maralit, Marc Edsel C. Ayes, Eva Maria Cutiongco-de la Paz, Alethea R. de Guzman, Jan Michael C. Yap, Jo-Hannah S. Llames, Sheila Mae M. Araiza, Kris P. Punayan, Irish Coleen A. Asin, Candice Francheska B. Tambaoan, Asia Louisa U. Chong, Karol Sophia Agape R. Padilla, Rianna Patricia S. Cruz, El King D. Morado, Joshua Gregor A. Dizon, Razel Nikka M. Hao, Arianne A. Zamora, Devon Ray Pacial, Juan Antonio R. Magalang, Marissa Alejandria, Celia Carlos, Anna Ong-Lim, Edsel Maurice Salvaña, John Q. Wong, Jaime C. Montoya, Maria Rosario Singh-Vergeire and Cynthia P. Saloma |
| EPI_ISL_1402625 | UCSC Genomics Institute | UCSC Genomics Institute | Mark Akeson, Ikenna Anigbogu, Eric Beraut, Isabel Bjork, Molly Cassatt-Johnstone, Terren Chang, Russell Corbett-Detig, Namrita Dhillon, Maximilian Haeussler, David Haussler, Angie Hinrichs, Miten Jain, Joshua Kapp, A. Marm Kilpatrick, Jakob McBroome, Hugh Olsen, Jeremy Sanford, Beth Shapiro, Michael Stone, Bryan Thornlow, Yatish Turakhia |
| EPI_ISL_601443 | Lighthouse Lab in Milton Keynes | Wellcome Sanger Institute for the COVID-19 Genomics UK (COG-UK) consortium | The Lighthouse Lab in Milton Keynes and Alex Alderton, Roberto Amato, Sonia Goncalves, Ewan Harrison, David K. Jackson, Ian Johnston, Dominic Kwiatkowski, Cordelia Langford, John Sillitoe on behalf of the Wellcome Sanger Institute COVID-19 Surveillance Team (http://www.sanger.ac.uk/covid-team) |
| EPI_ISL_792560 | Laboratorio de Ecologia de Doencas Transmissiveis na Amazonia, Instituto Leonidas e Maria Deane - Fiocruz Amazonia | Laboratorio de Ecologia de Doencas Transmissiveis na Amazonia, Instituto Leonidas e Maria Deane - Fiocruz Amazonia | Valdinete Nascimento, Victor Souza, André Corado, Fernanda Nascimento, George Silva, Ágatha Costa, Karina Pessoa, Debora Duarte, Luciana Gonçalves, Maria Júlia Brandão, Michele Jesus, Felipe Naveca on behalf of the Fiocruz COVID-19 Genomic Surveillance Network |
| EPI_ISL_833137 | Laboratorio de Ecologia de Doencas Transmissiveis na Amazonia, Instituto Leonidas e Maria Deane - Fiocruz Amazonia | Laboratorio de Ecologia de Doencas Transmissiveis na Amazonia, Instituto Leonidas e Maria Deane - Fiocruz Amazonia | Valdinete Nascimento, Victor Souza, André Corado, Fernanda Nascimento, George Silva, Ágatha Costa, Debora Duarte, Karina Pessoa, Matilde Mejía, Luciana Gonçalves, Maria Júlia Brandão, Michele Jesus, Felipe Naveca on behalf of the Fiocruz COVID-19 Genomic Surveillance Network |
| EPI_ISL_968081 | Public Health Virology-Forensic and Scientific Services (PHV-FSS) | Public Health Virology-Forensic and Scientific Services (PHV-FSS) | Alyssa T. Pyke et al. |

## 5. Associated References


1. T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veesler and J. D. Bloom, *Cell*, 2020, **182**, 1295-1310 e1220.
2. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang and X. Wang, *Nature*, 2020, **581**, 215-220.
3. N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714-2723.
4. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res*, 2018, **46**, W296-W303.
5. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, *Methods Enzymol*, 2011, **487**, 545-574.

6. C. M. Voloch, R. da Silva Francisco, L. G. P. de Almeida, C. C. Cardoso, O. J. Brustolini, A. L. Gerber, A. P. d. C. Guimarães, D. Mariani, R. M. da Costa, O. C. Ferreira, T. S. Frauches, C. M. B. de Mello, I. d. C. Leitão, R. M. Galliez, D. S. Faffe, T. M. P. P. Castiñeiras, A. Tanuri and A. T. R. de Vasconcelos, *Journal of Virology*, 2021, DOI: 10.1128/JVI.00119-21, JVI.00119-00121.
7. F. A. Tablizo, K. M. Kim, C. M. Lapid, M. J. R. Castro, M. S. L. Yangzon, B. A. Maralit, M. E. C. Ayes, E. M. Cutiongco-de la Paz, A. R. De Guzman, J. M. C. Yap, J.-H. S. Llames, S. M. M. Araiza, K. P. Punayan, I. C. A. Asin, C. F. B. Tambaoan, A. L. U. Chong, K. S. A. R. Padilla, R. P. S. Cruz, E. K. D. Morado, J. G. A. Dizon, R. N. M. Hao, A. A. Zamora, D. R. Pacial, J. A. R. Magalang, M. Alejandria, C. Carlos, A. Ong-Lim, E. M. Salvaña, J. Q. Wong, J. C. Montoya, M. R. Singh-Vergeire and C. P. Saloma, 2021, medRxiv:2021.03.03.21252812.
8. K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan and J. Meiler, *Biochemistry*, 2010, **49**, 2987-2998.
9. J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P. S. Huang, T. Huber, T. M. Jacobs, J. R. Jeliazkov, D. K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K. R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I. C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J. W. Labonte, J. K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J. H. Lubin, S. Lyskov, J. Maguire, L. Malmstrom, E. Marcos, O. Marcu, N. A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. O'Conchuir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R. E. Pavlovicz, M. Pethe, B. G. Pierce, K. B. Pilla, B. Raveh, P. D. Renfrew, S. S. R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A. M. Sevy, N. G. Sgourakis, L. Shi, J. B. Siegel, D. A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F. D. Teets, S. B. Thyme, R. Y. Wang, A. Watkins, L. Zimmerman and R. Bonneau, *Nat Methods*, 2020, **17**, 665-680.
10. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme and J. J. Gray, *J Chem Theory Comput*, 2017, **13**, 3031-3048.
11. A. Bazzoli, S. P. Kelow and J. Karanicolas, *PLoS One*, 2015, **10**, e0140359.
12. K. A. Barlow, O. C. S, S. Thompson, P. Suresh, J. E. Lucas, M. Heinonen and T. Kortemme, *J Phys Chem B*, 2018, **122**, 5389-5399.
13. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E. M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler and D. Baker, *PLoS One*, 2011, **6**, e20161.
14. B. J. Bender, A. Cisneros, 3rd, A. M. Duran, J. A. Finn, D. Fu, A. D. Lokits, B. K. Mueller, A. K. Sangha, M. F. Sauer, A. M. Sevy, G. Sliwoski, J. H. Sheehan, F. DiMaio, J. Meiler and R. Moretti, *Biochemistry*, 2016, **55**, 4748-4763.

15. R. Chowdhury, V. S. Boorla and C. D. Maranas, *Comput Struct Biotechnol J*, 2020, **18**, 2573-2582.
16. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res*, 2004, **32**, W665-667.
17. A. Spasic, J. Serafini and D. H. Mathews, *J Chem Theory Comput*, 2012, **8**, 2497-2505.
18. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc Natl Acad Sci U S A*, 2001, **98**, 10037-10041.
19. C. O. Barnes, C. A. Jette, M. E. Abernathy, K. A. Dam, S. R. Esswein, H. B. Gristick, A. G. Malyutin, N. G. Sharaf, K. E. Huey-Tubman, Y. E. Lee, D. F. Robbiani, M. C. Nussenzweig, A. P. West, Jr. and P. J. Bjorkman, *Nature*, 2020, **588**, 682-687.