In response to COVID-19, we used our in-silico design platform to propose mutations to SARS-CoV-1 neutralizing antibodies to achieve and optimize binding to the receptor binding domain of SARS-CoV-2 spike protein. The data contained here is a standard .csv file. Each row represents a unique mutant antibody proposed by our machine learning model, derived from a publicly available starting template antibody M396. The data for each mutant antibody (i.e., each row) are along the columns and are the result of several software-based calculations as described in the main text and below. Fields (column headers), in the order they appear in the file, are:

* Antibody\_ID
  + *unique antibody identifier*

|  |
| --- |
| * Complex |

* + *an identifier for the protein structure used for software calculations also included in Supplementary Materials*
* Structure\_MD5\_Hash
  + *MD5 hash protein structure used for software calculations*
* Antibody
  + *identifier for the starting template antibody from which the mutant antibody was derived*
* Mutation
  + *the mutations performed to produce the mutant antibody from the starting template antibody*
  + *see conversion between our internal sequence ordering and original ordering)*
* Antibody\_Sequence
  + *sequence of the mutant antibody for which calculations were performed*
* FoldX\_Average\_Whole\_Model\_DDG
  + *Antibody-antigen free energy estimates using FoldX [1] software for the mutant antibody for the whole Antibody-antigen complex*
* FoldX\_Average\_Interface\_Only\_DDG
  + *Antibody-antigen free energy estimates using FoldX [1] software for the mutant antibody for the interface of the Antibody-antigen complex*
* Rosetta\_Flex\_DDG
  + *Antibody-antigen bindings estimates using Rosetta [2] software for the mutant antibody using the "Flex\_ddG" protocol*
* Rosetta\_Total\_Energy\_DDG
  + *Antibody-antigen bindings estimates using Rosetta [2] software for the mutant antibody using the “ddg\_monomer" algorithm*
* MMGBSA
  + *Antibody-antigen bindings estimates from LLNL molecular dynamics calculations (molecular mechanics/generalized Born solvent accessible surface area) for the mutant antibody*
* Statium
  + *Antibody-antigen predictions using Statium [3] software for the mutant antibody*
* Total\_CDR\_Length
  + *Total CDR Length from Therapeutic Antibody Profiler [4]*
* CDR\_Vicinity\_PSH\_Score
  + *CDR Vicinity PSH Score (Kyte & Doolittle) from Therapeutic Antibody Profiler [4]*
* CDR\_Vicinity\_PPC\_Score
  + *CDR Vicinity PPC Score from Therapeutic Antibody Profiler [4]*
* CDR\_Vicinity\_PNC\_Score
  + *CDR Vicinity PNC Score from Therapeutic Antibody Profiler [4]*
* SFvCSP\_Score
  + *SFvCSP Score from Therapeutic Antibody Profiler [4]*
* number\_of\_unconventional\_mutations
  + *A bioinformatic heuristic representing how evolutionarily unlikely the mutations in the mutant antibody are. Derived from BLOSUM62 [5]*
* Sum\_of\_Rosetta\_Flex\_single\_point\_mutations
  + *Sum of Rosetta Flex calculations for each single point mutations present in mutant*
* Sum\_of\_Rosetta\_Total\_Energy\_single\_point\_mutations
  + *Sum of Rosetta Total Energy calculations for each single point mutations present in mutant*
* Rosetta\_Flex\_calculations\_single\_point\_mutations
  + *Rosetta Flex calculations of single point mutations for each mutation present in the mutant antibody*
  + *ordering corresponds to ordering in field labeled “Residue\_locations\_allowed\_to\_mutate\_LLNL”*
  + *Note not all locations are mutated for every mutant.*
* Rosetta\_Total\_Energy\_calculations\_single\_point\_mutations
  + *Rosetta Total Energy calculations of single point mutations for each mutation present in the mutant antibody.*
  + *ordering corresponds to ordering in field labeled “Residue\_locations\_allowed\_to\_mutate\_LLNL”*
  + *Note not all locations are mutated for every mutant.*
* Residue\_locations\_allowed\_to\_mutate\_LLNL
  + *Residue locations allowed to mutate using our internal LLNL numbering of sequences. Note not all locations are mutated for every mutant*
* Residue\_locations\_allowed\_to\_mutate\_original
  + *Residue locations allowed to mutate using original numbering of sequences. Note not all locations are mutated for every mutant.*

Note that the residue numbering in our structures differs from the numbering in the original M396 structure. The conversion is as follows:

|  |  |
| --- | --- |
| Our numbering | Original M396 numbering |
| S31 | S31\_H |
| Y32 | Y32\_H |
| T33 | T33\_H |
| W47 | W47\_H |
| G50 | G50\_H |
| I51 | I51\_H |
| T52 | T52\_H |
| I54 | I53\_H |
| L55 | L54\_H |
| I57 | I56\_H |
| A58 | A57\_H |
| N59 | N58\_H |
| Y60 | Y59\_H |
| A61 | A60\_H |
| Q62 | Q61\_H |
| D99 | D95\_H |
| T100 | T96\_H |
| V101 | V97\_H |
| M102 | M98\_H |
| G103 | G99\_H |
| G104 | G100\_H |
| N271 | N27\_L |
| G273 | G29\_L |
| S274 | S30\_L |
| K275 | K31\_L |
| W335 | W91\_L |
| D336 | D92\_L |
| S337 | S93\_L |
| S338 | S94\_L |
| D340 | D95A\_L |
| Y341 | Y96\_L |

[1] Schymkowitz J, et al. The FoldX web server: an online force field. Nucleic Acids Res, 2005 Jul 1, Volume 33, Issue Web Server issue, p.W382-8

[2] Leaver-Fay A, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in Enzymology, 2011; Vol. 487, p 545-574.

[3] DeBartolo J, et al. Genome-wide prediction and validation of peptides that bind human prosurvival Bcl-2 proteins. PLoS computational biology. 2014 Jun 26;10(6):e1003693.

[4] Raybould, M. I., et al. Five computational developability guidelines for therapeutic antibody profiling. Proceedings of the National Academy of Sciences, 116(10), 4025-4030 (2019).

[5] Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences, 89(22), 10915-10919.