bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Potential Neutralizing Antibodies Discovered for Novel Corona Virus Using Machine Learning

Rishikesh Magar [¥], *Prakarsh Yadav* [‡] *and Amir Barati Farimani* [1¥§‡]

[¥] Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

[‡] Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

[§]Machine Learning Department, School of Computer Science, Carnegie Mellon University Pittsburgh, PA 15213

Abstract

The fast and untraceable virus mutations take lives of thousands of people before the immune system can produce the inhibitory antibody. Recent outbreak of novel coronavirus infected and killed thousands of people in the world. Rapid methods in finding peptides or antibody sequences that can inhibit the viral epitopes of COVID-19 will save the life of thousands. In this paper, we devised a machine learning (ML) model to predict the possible inhibitory synthetic antibodies for Corona virus. We collected 1933 virus-antibody sequences and their clinical patient neutralization response and trained an ML model to predict the antibody response. Using graph featurization with variety of ML methods, we screened thousands of hypothetical antibody sequences and found 8 stable antibodies that potentially inhibit COVID-19. We combined bioinformatics, structural biology, and Molecular Dynamics (MD) simulations to verify the stability of the candidate antibodies that can inhibit the Corona virus.

[1] Corresponding Author, e-mail: barati@cmu.edu, Website: www.baratilab.com

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Keywords: Coronavirus, COVID-19, Machine Learning, Antibody Engineering, Bio-informatics, Structural Biology

Introduction

The biomolecular process for recognition and neutralization of viral particles is through the process of viral antigen presentation and recruitment of appropriate B cells to synthesize the neutralizing antibodies.[1] Theoretically, this process allows the immune system to stop any viral invasion, but this response is slow and often requires days, even weeks before adequate immune response can be achieved.[2,3] This poses a challenging question: Can the process of antibody discovery be accelerated to counter highly infective viral diseases?

With the rapid expansion of available biological data, such as DNA/protein sequences and structures[4], it is now possible to model and predict the complex biological phenomena through machine learning (ML) approaches. Given sufficient training data, ML can be used to learn a mapping between the viral epitope and effectiveness of its complementary antibody.[5] Once such mapping is learnt, it can be used to predict potentially neutralizing antibody for a given viral sequence[6].

ML can essentially learn the complex antigen-antibody interactions faster than human immune system, leading to the generation of synthetic inhibitory antibodies acting as a bridge, which can overcome the latency between viral infection and human immune system response. This bridge can potentially save the life of many especially during an outbreak and pandemic. One such instance is the spread of coronavirus disease (COVID-19)[7].

With incredibly high infectivity and mortality rate, COVID-19 has become a global scare.[8,9] To compound the problem, there are no proven therapeutics to aid the suffering patients[2,8,10–14]. Only viable treatment at the moment is symptomatic and there is a desperate need for developing therapeutics to counter COVID-19. Recently, the proteomics sequences of 'WH-Human 1' coronavirus became available through Metagenomic RNA sequencing of a patient in Wuhan.[4,15] WH-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Human 1 is 89.1% similar to a group of SARS-like coronaviruses.[4] With this sequence available, it is possible to find potential inhibitory antibodies by scanning thousands of antibody sequences and discovering the neutralizing ones[16–18]. However, this requires very expensive and time-consuming experimentation to discover the inhibitory responses to Corona virus in a timely manner. In addition, computational and physics-based models require the bound crystal structure of antibody-antigen complex, however; only a few of these structures have become available.[19,20,21,22] In the case of COVID-19, the bound antigen-antibody crystal structure is not available to-date[23,24]. Given this challenge and the fact that ML models require a large amount of data, the ML approach should rely on the sequences of the antibody-antigen rather than the crystal structures[25].

In this paper, we have collected a dataset comprised of antibody-antigen sequences of variety of viruses including HIV, Influenza, Dengue, SARS, Ebola, Hepatitis, etc. with their patient clinical/biochemical IC_{50} data. Using this dataset (we call it VirusNet), we trained and benchmarked different shallow and deep ML models and selected the best performing model. Based on SARS 2006 neutralizing antibody scaffold[26], we created thousands of potential antibody candidates by mutation and screened them with our best performing ML model. Finally, molecular dynamics (MD) simulations were performed on the neutralizing candidates to check their structural stability. We predict 8 structures that were stable over the course of simulation and are potential neutralizing antibodies for COVID-19.

In addition, we interpreted the ML method to understand what alterations in the sequence of binding region of the antibody would most effectively counter the viral mutation(s) and restore the ability of the antibody to bind to the virus[27]. This information is critical in terms of antibody design and engineering and reducing the dimension of combinatoric mutations needed to find a neutralizing antibody.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Methods

1. Dataset

The majority of the data in the training set is composed of HIV antibody-antigen complex (1887 samples). Most of the samples for the HIV training set were obtained from the Compile, Analyze and Tally NAb panels (CATNAP) database from the Los Alamos National Laboratory (LANL) [28,29]. From CATNAP, data was collected for monoclonal antibodies, 2F5, 4E10 and 10E8, which bind with GP41[30–32]. Using CATNAP's functionality for identifying epitope alignment, we selected FASTA sequence of the antigen corresponding to the site of alignment, in the antibody. We generated a dataset with 1831 training examples comprising of antibodies – antigen sequences and their corresponding IC_{50} values. The CATNAP output is comprised of site of antigen sequence alignment for each of the antibodies with respect to 2F5, 4E10 and 10E8. Using the co-crystalized structure of (2F5-ELDKWAS) in (PDB:1TJG)[30] as template, the antibody fragment that comes in contact with the antigen was found by considering amino acids within 7Å of the antigen in the co-crystallized structure.

To make the dataset more diverse and train a more robust ML model, we included more available antibody-antigen sequences and their neutralization potential. To do this, we compiled the sequences of Influenza, Dengue, Ebola, SARS, Hepatitis, etc.[26,33–86] by searching the keywords of "virus, antibody" on RCSB server[87] and selected the neutralizing complex by reading their corresponding publications. Furthermore, for each neutralizing complex, the contact residues at the interface of antibody and antigen were selected. To select the antigen contact sequences, all amino acids within 5Å of corresponding antibody were chosen. (Supporting Information) To select the antibody contact sequences, all amino acids within 5Å of the antigen were chosen. In total, 102 sequences of antibody-antigen complexes were mined and added to the 1831 samples, resulting in total number of 1933 training samples.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

2. Graph Featurization and Machine Learning

For effective representation of molecular structure of amino acids, the individual atoms of amino acids of antibody and antigen were treated as undirected graph, where the atoms are nodes and bonds are edges[88]. It has been shown that graph representation is better in transferring the chemistry and topology of molecular structure compared to Extended Connectivity Fingerprints (ECFP)[88,89]. We construct these molecular graphs using RDkit[90]. Embeddings are generated to encode relevant features about the molecular graph[91,92]. These embeddings encode information like the type of atom, valency of an atom, hybridization state, aromaticity etc. First, each antibody and antigen were encoded into separate embeddings and then concatenated into a single embedding for the entire antibody-antigen complex. We then apply mean pooling over the features for this concatenated embedding to ensure dimensional consistency across the training data. The pooled information is then passed to classifier algorithms like XGBoost[93], Random Forest[94], Multilayer perceptron, Support Vector Machine (SVM)[95] and Logistic Regression which then predict whether the antibody is capable of neutralizing the virus.

3. Hypothetical Antibody Library Generation

In order to find potential antibody candidates for COVID-19, 2589 different mutant strains of antibody sequences were generated based on the sequence of SARS neutralizing antibodies. The reason we selected these antibodies as initial scaffolds is that the genome of COVID-19 [4] is 79.8% identical to "Tor2" isolate of SARS (Accession number: AY274119)[96]. Using 4 different antibody variants of SARS (PDB: 2GHW, 3BGF, 6NB6, 2DD8)[26,76,81,86], point mutations were applied to all

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

the amino acids in the binding region of antibody. (see Supporting Information for COVID-19 antigen and antibody interactions) To find out the binding region of these antibodies for sequence generation, all amino acids within 5Å of their respective antigen were chosen. To assess the biological feasibility of these mutant sequences, we scored each mutation by using the BLOSUM62 matrix[97].

4. Molecular Dynamics Simulations

To assess the stability of proposed antibody structures, we performed molecular dynamics (MD) simulations of each of antibody structure in a solvated environment[98]. The simulation of solvated antibody was carried out using GROMACS-5.1.4[99–101], and topologies for each antibody were generated according the GROMOS 54a7[102] forcefield. The protein was centered in a box, extending 1 nanometer from surface of the protein. This box was the solvated by the SPC216 model water atoms, pre-equilibrated at 300K. The antibody system in general carried a net positive charge and it was neutralized by the counter ions. Energy minimization was carried out using steepest descent algorithm, while restraining the peptide backbone to remove the steric clashes in atoms and subsequently optimize solvent molecule geometry. The cut-off distance criteria for this minimization were forces less than 100.0 kJ/mol/nm or number of steps exceeding 50,000. This minimized structure was the sent to two rounds of equilibration at 300K. First, an NVT ensemble for 50 picoseconds and a 2-femtosecond time step. Leapfrog dynamics integrator was used with Verlet scheme, neighbor-list was updated every 10 steps. All the ensembles were under Periodic Boundary Conditions and harmonic constraints were applied by the LINCS algorithm[103]; under this scheme the long-range electrostatic interactions were computed by Particle Mesh Ewald (PME) algorithm[104]. Berendsen thermostat was used for temperature coupling and pressure coupling was done using the Parrinello-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Rahman barostat[105,106]. The last round of NPT simulation ensures that the simulated system is at physiological temperature and pressure. The system volume was free to change in the NPT ensemble but in fact did not change significantly during the course of the simulation. Following the rounds of equilibration, production run for the system was carried out in NPT and no constraints for a total of 15 nanoseconds, under identical simulation parameters.

Results and Discussions

The flowchart of COVID-19 antibody discovery using ML has four major steps (Figure 1): 1. Collecting data and curating the dataset for training set. 2. Featurization, embedding and benchmarking ML models and selecting the best performing one. 3. Hypothetical antibody library generation and ML screening for neutralization and 4. Checking the stability of proposed antibodies. This workflow enables the rapid screening of large space of potential antibodies to neutralize COVID-19. In general, this workflow can be used for high throughput screening of antibodies for any type of virus by only knowing the sequences of antigen epitopes.

To better understand the diversity and similarity of the sequences that were used in the training set, t-Distributed Stochastic Neighbor Embedding (t-SNE) of all different viruses were computed (Figure 2a). t-SNE axes shows the directions of the maximum variance in the dataset, therefore, the dimensionality of the data can be reduced to lower dimensions (here two). HIV antigen shows the most variations on t-SNE components where viruses such as Influenza, Dengue and H1N1 are very close to each other. The neutralizing antibodies were also projected using t-SNE to show the variations in the available neutralizing sequences. (Figure 2b) Unlike antigen variations, antibody sequences are much closer to the center of t-SNE with a few scattered ones. The comparison of Figure 2a and Figure 2b shows that the neutralizing antibodies are not sequence-diverse compared to virus antigens. This difference demonstrates that a large space of potential antibodies can be screened and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

used for finding novel antibodies. The labels in the dataset are comprised of the neutralization panel data, IC_{50} values for monoclonal antibodies and pseudo-typed viruses. (Figure 2c) The IC_{50} labels were collected from 49 published neutralization studies and were collected from Los Alamos HIV Database (for 1831 samples in our training set). For some cases in CATNAP, personal communication with the authors were made to resolve sequence name ambiguities between different laboratories. For 102 samples of various viruses collected from RCSB server, all of them neutralize their antigen based on biochemical assays. These samples were labeled by setting their IC_{50} to zero. Since classification is performed on the training dataset, $IC_{50} \leq 10$ are set to neutralizing and $IC_{50} > 10$ to non-neutralizing (Figure 2c). To visualize the diversity of the virus types used in the dataset other than HIV, the distribution of 13 more viruses were presented in Figure 2d. Influenza, Dengue, SARS, Ebola and then Hepatitis have relatively larger samples in the dataset.

To benchmark the performance of different ML models on the VirusNet dataset and select the best performing one, XGBoost, Random Forest (RF), Multilayer perceptron (MLP), Support Vector Machines (SVM), and Logistic Regression (LR) were used (Figure 3a). The five-fold cross validation on 80%-20% split, train, and test resulted in best accuracy for XG-Boost model. The performance and ranking of models follow the order of XGBoost (90.57%)> RF (89.18%)>LR (81.17%)> MLP (78.23)> SVM (75.49%). Since the training data is sparse in the case of VirusNet (See Figure 2a and Figure 2b), XGBoost selects the sparse features input by pruning and learning the underlying sparsity patterns. To test the robustness of the XGBoost on completely unseen virus types, for each left-out virus type, the model was trained on all the sequences in the VirusNet except for the left-out. For example, for Influenza, all the sequences and labels of Influenza were removed from the training set and the trained model on the remaining dataset were tested on all the Influenzas's sequences and consequently the classification accuracy were reported (Figure 3b). The accuracies for the out of class

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

test is as follows: Influenza (84.61%), Dengue (100%), Ebola (75%), Hepatitis (75%), SARS (100%). The out of class results demonstrate that our model is capable of generalizing the prediction to a completely novel virus epitope. Since COVID-19 is completely a new virus, we can conclude that our model prediction performance will be accurate. The fact that our model prediction is 100% for SARS out of class test demonstrate its capability of effectively predicting the antibodies for COVID-19 which is from SARS family.

Next, using the best performing model (XGBoost), all the hypothetical candidates in the library were evaluated for neutralization. Out of all the candidates, some of them are invalid mutations screened using BLOSUM62 matrix[97] (Figure 3c). 18 final candidates are both valid mutations and can neutralize COVID-19 with high confidence probability of 0.9895 (shown with green color in Figure 3c).

Interpretability of the ML models is very important in both explaining the underlying biological and chemical understanding of neutralization and providing design guidelines for antibody engineering. One of the significant advantages of ensemble methods such as XGBoost is their interpretability. By taking advantage of this property, the important features that are giving rise to neutralization were ranked and scored (Figure 3d). The input features to the model contains atomic level attributes such as atom type, valency, hybridization, etc. To collectively translate the important atomic features into important amino acid features, the scores of amino acids with unique atomic features were summed up and ranked (Figure 3d). Some of the atomic features were common among all the amino acids (e.g. Carbon, implicit valency, Oxygen, etc.) therefore; we ignored them. However, some of the other features like aromaticity or having Sulfur are unique and we considered those in amino acid features. Based on the unique features appeared in ranking, Methionine (M) is the most important one. (Figure 3d)

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

The predicted sequences from the ML model were then used to model the novel structure of potentially neutralizing antibodies. The predicted sequences were projected onto their progenitor antibody and the changes in amino acid sequence were modelled as follows: Simple point mutations were introduced by modifying the target amino acid using Coot [107,108](Crystallographic Object-Oriented Toolkit). Coot environment allowed us to predict the stereochemical effect of each point mutation and appropriately compensate for it. Using such an approach, we were able to accurately model the putative structures of the antibodies. The modelled structures were then passed to MD simulations for stability check.

To check the stability of predicted structures energetically, 20 MD simulations (18 point mutations+2 wildtype (WT)) in total were performed (Figure 4a). Structures with low Root Mean Square Deviation (RMSD) and low contact distance are in a stable conformation, whereas structures with high RMSD and high contact distances are in an unstable conformation. RMSD and contact distance for WT structures have lower values, demonstrating stability, therefore; the contact distances versus RMSD is a good indicator of the stability of a protein over the course of a simulation (Figure 4b).

Once mutation introduced in the crystallographic structure, it will cause it to deviate from WT structure's RMSD and contact distance. We performed simulations for all the 18 point-mutant structures and their mean contact distance versus RMSD[109,110] were computed for their respective trajectories (Figure 4b) (see Supporting Information). Based on the two WT structures mean RMSD and contact distances, we selected the mutations which have mean contact distance and RMSD values less than 0.488 nm and 0.35 nm, respectively. (the shaded triangle region in Figure 4b). Candidates with higher values of mean RMSD and contact distances are unstable and will potentially fail to neutralize the COVID-19.

In order to be more comprehensive, we created co-mutations out of 5 stable point mutations (C3, C7, C14, C17, C18, see Table S1 in Supporting Information for the list of all 18 candidates). This resulted

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

in 5 new structures (Co1, Co2, Co3, Co4, Co5 in Table S2) that were screened using XGBoost for neutralization. Among all 5 co-mutations, Co5 did not neutralize. To check the stability of these 4 neutralizing co-mutations, MD simulations were performed and Co1, Co2 and Co4 were found to be stable (Figure 4b). The list of the final 8 stable mutations and co-mutations are tabulated in Table 1 and the PDB structures are available as Supporting Information.

Conclusion

We have developed a machine learning model for high throughput screening of synthetic antibodies to discover antibodies that potentially inhibit the COVID-19. Our approach can be widely applied to other viruses where only the sequences of viral coat protein-antibody pairs can be obtained. The ML models were trained on 14 different virus types and achieved over 90% fivefold test accuracy. The out of class prediction is 100% for SARS and 84.61% for Influenza, demonstrating the power of our model for neutralization prediction of antibodies for novel viruses like COVID-19. Using this model, the neutralization of thousands of hypothetical antibodies was predicted, and 18 antibodies were found to be highly efficient in neutralizing COVID-19. Using MD simulations, the stability of predicted antibodies were checked and 8 stable antibodies were found that can neutralize COVID-19. In addition, the interpretation of ML model revealed that mutating to Methionine and Tyrosine is highly efficient in enhancing the affinity of antibodies to COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Acknowledgements

The authors gratefully acknowledge the use of the supercomputing resource Arjuna provided by the Pittsburgh Supercomputing Center (PSC). This work is supported by Center for Machine Learning in Health (CMLH) at Carnegie Mellon University (https://www.cs.cmu.edu/cmlh-cfp) and start-up fund from Mechanical Engineering Department at CMU. The authors would like to thank Prof. Reeja Jayan for her support and Junhan Li for his help in collecting the data.

Supporting Information Available

The RMSD and contact distance plots for all the trajectories versus time, the structure of virus antibody complex and the residues at the contact region, native contacts in antigen-antibody complex, the interaction of COVID-19 epitope with 2GHW antibody, tables for all neutralizing point mutations, co-mutations and their neutralization potentials, the structures of stable antibodies in PDB format, and IC_{50} data interpretation are available online. The VirusNet dataset will be available upon request.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

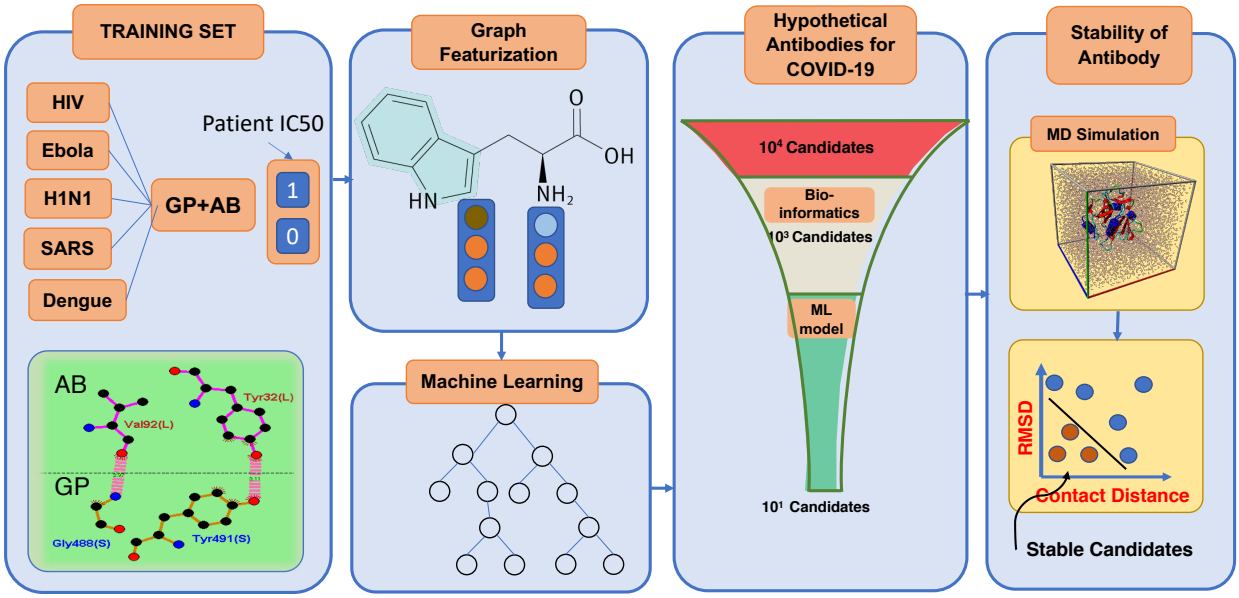


Figure 1. Designing antibodies or peptide sequences that can inhibit the COVID-19 virus requires high throughput experimentation of vastly mutated sequences of potential inhibitors. The screening of thousands of available strains of antibodies are prohibitively expensive, and not feasible due to lack of available structures. However, machine learning models can enable the rapid and inexpensive exploration of vast sequence space on the computer in a fraction of seconds. We collected 1933 virus-antibody sequences with clinical patient IC_{50} data. Graph featurization of antibody-antigen sequences creates a unique molecular representation. Using graph representation, we benchmarked and used a variety of shallow and deep learning models and selected XGBoost because of its superior performance and interpretability. We trained our model using a dataset including 1,933 diverse virus epitope and the antibodies. To generate the hypothetical antibody library, we mutated the SARS scaffold antibody of 2006 (PDB:2GHW) and generated thousands of possible candidates. Using the ML model, we classified these sequences and selected the top 18 sequences that will neutralize COVID-19 with high confidence. We used MD simulations to check the stability of the 18 sequences and rank them based on their stability.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

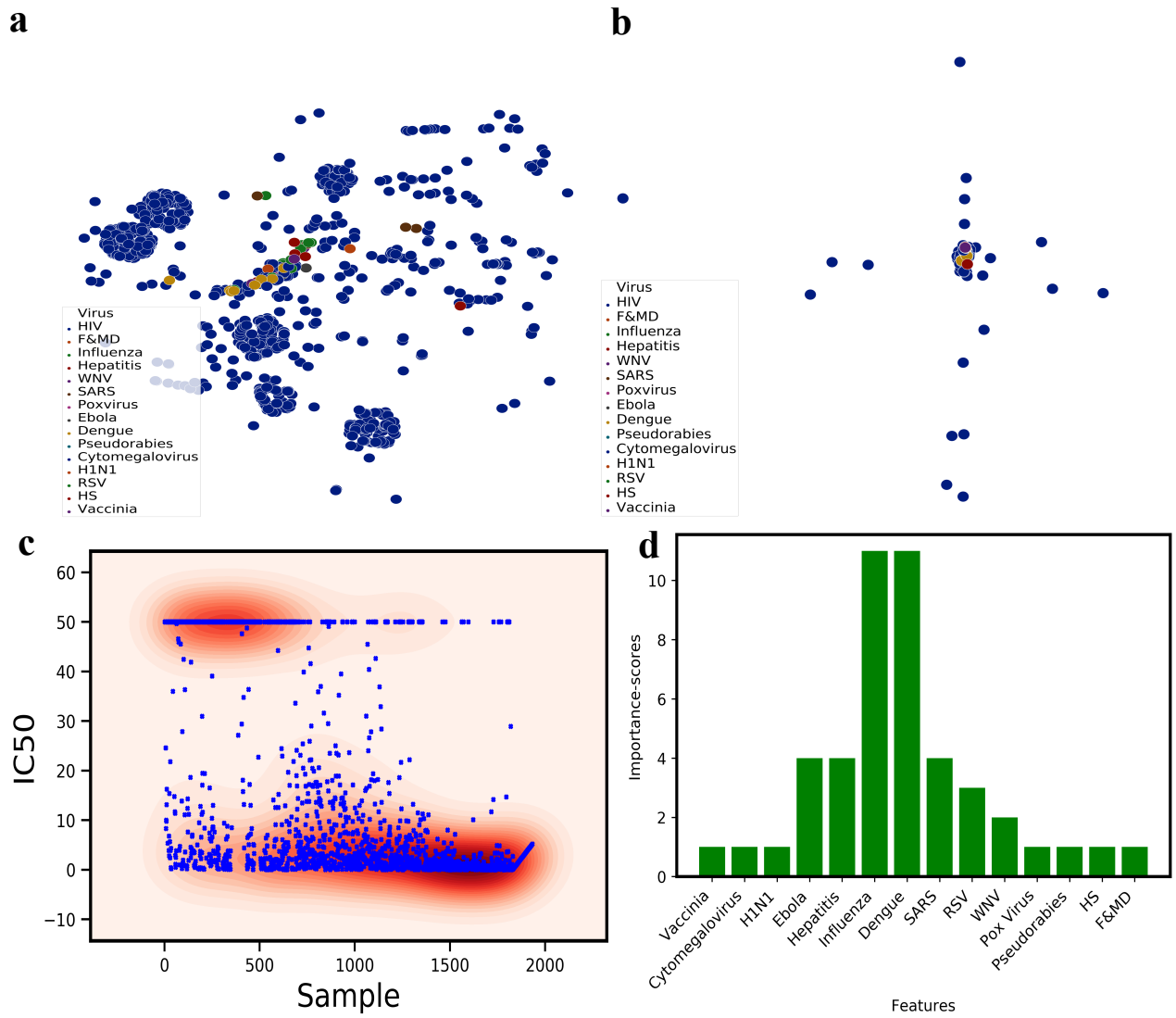


Figure 2. **a)** t-Distributed Stochastic Neighbor Embedding (t-SNE) of all the viruses epitopes used in the training dataset, revealing biological similarity and diversity of the sequences used in the dataset. **b)** t-SNE of all the therapeutics antibody sequences used in the training set for variety of different virus types. The majority of the broadly neutralizing antibodies such as 2F5 is clustered at the center of this plot. **c)** Patient clinical IC50 data obtained from various sources and the distribution of the neutralizing (IC_{50}<10) and Non- neutralizing (IC_{50}>10) samples. **d)** The number of samples for each virus class except HIV. For HIV, we collected 1883 samples. Influenza and Dengue has 10+ samples

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

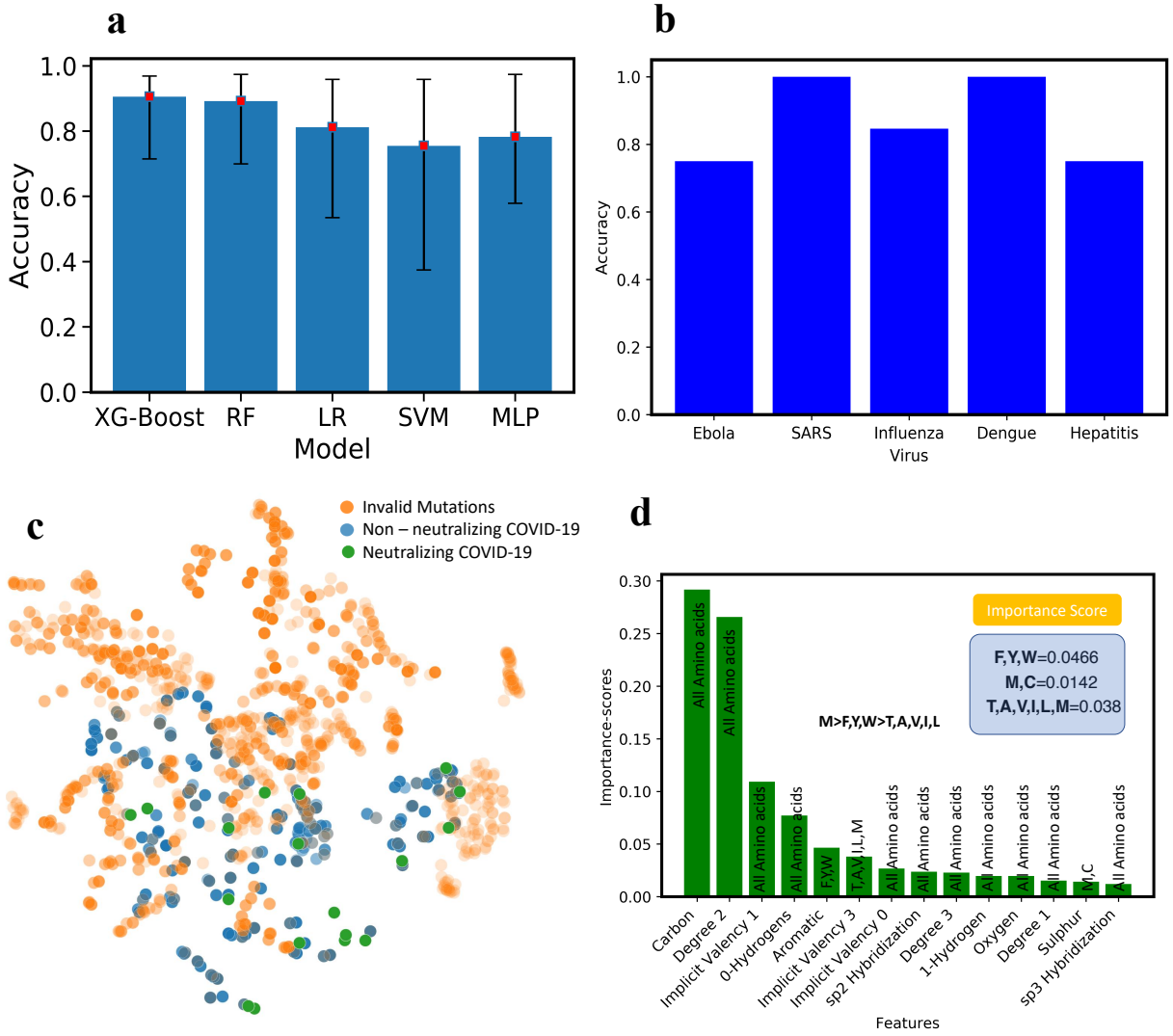


Figure 3. **a)** The test accuracy with five-fold cross validation for XG-Boost, Random Forrest (RF), Logistic Regression (LR), Support Vector Machine (SVM) and Deep Learning (Multilayer Perceptron. XGBoost has the highest performance with (90.75%). **b)** Out of training class test accuracy for influenza, Dengue, Ebola, Hepatitis, and SARS. To perform this test, for example for influenza, all the influenza virus-antibody sequences were removed from the training set and the obtained model were tested on all samples of Influenza and the accuracy is reported here. **c)** Blossum validated mutations, non-neutralizing and neutralizing antibody sequences. To achieve more confidence, we set the threshold of prediction probability to 0.9895 in XGBoost and found 18 neutralizing antibody sequences (the green points). **d)** Interpretability of ML model: to understand what mutations are playing the key roles in neutralization, XGBoost feature importance used with ranked atomic level features. Through connecting the atomic features with each of 20 amino acids, M was found to be the most important amino acids in neutralization followed by F, Y, W.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

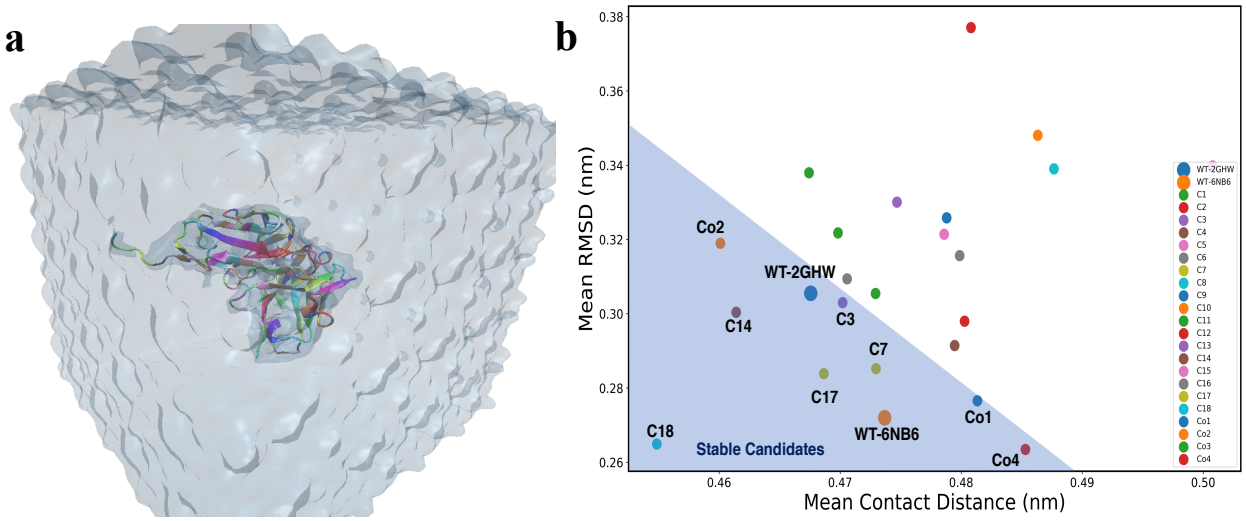


Figure 4. **a)** The snapshot of MD simulation of mutated proteins. Each protein is solvated in a box of water and simulated to collect the statistical data on the stability of mutants and co-mutants. **b)** Mean Root Mean Square Deviation (RMSD) versus Mean contact distances for each candidate averaged over the whole trajectory.

Structure	Mutation
C3	2GHW–I51M
C7	2GHW–R150H
C14	2GHW–T204N
C17	6NB6–R56H
C18	6NB6–K58S
Co1	2GHW-I51M, R150H, T204N
Co2	2GHW-I51M, R150H
Co4	6NB6-R56H, K58S

Table 1: The final neutralizing candidates obtained through screening with ML model, MD simulation for stability and Bioinformatics. The detailed list of sequences is available in the Supporting Information.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

References

(1) Dörner, T.; Radbruch, A. Antibodies and B Cell Memory in Viral Immunity. *Immunity* **2007**, *27* (3), 384–392. https://doi.org/10.1016/j.immuni.2007.09.002.

(2) Li, Z.; Yi, Y.; Luo, X.; Xiong, N.; Liu, Y.; Li, S.; Sun, R.; Wang, Y.; Hu, B.; Chen, W.; Zhang, Y.; Wang, J.; Huang, B.; Lin, Y.; Yang, J.; Cai, W.; Wang, X.; Cheng, J.; Chen, Z.; Sun, K.; Pan, W.; Zhan, Z.; Chen, L.; Ye, F. Development and Clinical Application of A Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. *J. Med. Virol.* **2020**, *n/a* (n/a). https://doi.org/10.1002/jmv.25727.

(3) Hewitt, E. W. The MHC Class I Antigen Presentation Pathway: Strategies for Viral Immune Evasion. *Immunology* **2003**, *110* (2), 163–169. https://doi.org/10.1046/j.1365-2567.2003.01738.x.

(4) Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, 1–5. https://doi.org/10.1038/s41586-020-2008-3.

(5) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackerman, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180* (4), 688-702.e13. https://doi.org/10.1016/j.cell.2020.01.021.

(6) Continuous cultures of fused cells secreting antibody of predefined specificity | Nature https://www.nature.com/articles/256495a0 (accessed Mar 9, 2020).

(7) The Species Severe Acute Respiratory Syndrome-Related Coronavirus : Classifying 2019-NCoV and Naming It SARS-CoV-2. *Nat. Microbiol.* **2020**, 1–9. https://doi.org/10.1038/s41564-020-0695-z.

(8) Fu, Y.; Cheng, Y.; Wu, Y. Understanding SARS-CoV-2-Mediated Inflammatory Responses: From Mechanisms to Potential Therapeutic Tools. *Virol. Sin.* **2020**. https://doi.org/10.1007/s12250-020-00207-4.

(9) Secondary attack rate and superspreading events for SARS-CoV-2 - The Lancet https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30462-1/fulltext (accessed Mar 9, 2020).

(10) Cao, Y.; Li, L.; Feng, Z.; Wan, S.; Huang, P.; Sun, X.; Wen, F.; Huang, X.; Ning, G.; Wang, W. Comparative Genetic Analysis of the Novel Coronavirus (2019-NCoV/SARS-CoV-2) Receptor ACE2 in Different Populations. *Cell Discov.* **2020**, *6* (1), 1–4. https://doi.org/10.1038/s41421-020-0147-1.

(11) Chang, Y.-C.; Tung, Y.-A.; Lee, K.-H.; Chen, T.-F.; Hsiao, Y.-C.; Chang, H.-C.; Hsieh, T.-T.; Su, C.-H.; Wang, S.-S.; Yu, J.-Y.; Shih, S.; Lin, Y.-H.; Lin, Y.-H.; Tu, Y.-C. E.; Tung, C.-W.; Chen, C.-Y. Potential Therapeutic Agents for COVID-19 Based on the Analysis of Protease and RNA Polymerase Docking. **2020**. https://doi.org/10.20944/preprints202002.0242.v1.

(12) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *0* (0). https://doi.org/10.1016/j.cell.2020.02.052.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(13) Nguyen, T. M.; Zhang, Y.; Pandolfi, P. P. Virus against Virus: A Potential Treatment for 2019-NCov (SARS-CoV-2) and Other RNA Viruses. *Cell Res.* **2020**, *30* (3), 189–190. https://doi.org/10.1038/s41422-020-0290-0.

(14) Tian, X.; Li, C.; Huang, A.; Xia, S.; Lu, S.; Shi, Z.; Lu, L.; Jiang, S.; Yang, Z.; Wu, Y.; Ying, T. Potent Binding of 2019 Novel Coronavirus Spike Protein by a SARS Coronavirus-Specific Human Monoclonal Antibody. *Emerg. Microbes Infect.* **2020**, *9* (1), 382–385. https://doi.org/10.1080/22221751.2020.1729069.

(15) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet* **2020**, *395* (10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.

(16) Karthick, V.; Nagasundaram, N.; Doss, C. G. P.; Chakraborty, C.; Siva, R.; Lu, A.; Zhang, G.; Zhu, H. Virtual Screening of the Inhibitors Targeting at the Viral Protein 40 of Ebola Virus. *Infect. Dis. Poverty* **2016**, *5*. https://doi.org/10.1186/s40249-016-0105-1.

(17) Miller, C. R.; Johnson, E. L.; Burke, A. Z.; Martin, K. P.; Miura, T. A.; Wichman, H. A.; Brown, C. J.; Ytreberg, F. M. Initiating a Watch List for Ebola Virus Antibody Escape Mutations. *PeerJ* **2016**, *4*, e1674. https://doi.org/10.7717/peerj.1674.

(18) Murin, C. D.; Fusco, M. L.; Bornholdt, Z. A.; Qiu, X.; Olinger, G. G.; Zeitlin, L.; Kobinger, G. P.; Ward, A. B.; Saphire, E. O. Structures of Protective Antibodies Reveal Sites of Vulnerability on Ebola Virus. *Proc. Natl. Acad. Sci.* **2014**, *111* (48), 17182–17187. https://doi.org/10.1073/pnas.1414164111.

(19) Tiller, K. E.; Tessier, P. M. Advances in Antibody Design. *Annu. Rev. Biomed. Eng.* **2015**, *17* (1), 191–216. https://doi.org/10.1146/annurev-bioeng-071114-040733.

(20) Computational predictions of protein structures associated with COVID-19 /research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19 (accessed Mar 9, 2020).

(21) Jakobsson, E.; Farimani, A. B.; Tajkhorshid, E.; Aluru, N. Combining Physics-Based and Evolution-Based Methods to Design Antibodies Against an Evolving Virus. *Biophys. J.* **2020**, *118* (3), 482a. https://doi.org/10.1016/j.bpj.2019.11.2669.

(22) Barati Farimani, A.; Aluru, N. R.; Tajkhorshid, E.; Jakobsson, E. Computational Approach to Designing Antibody for Ebola Virus. *Biophys. J.* **2016**, *110* (3), 537a. https://doi.org/10.1016/j.bpj.2015.11.2877.

(23) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural Basis for the Recognition of the SARS-CoV-2 by Full-Length Human ACE2. *Science* **2020**. https://doi.org/10.1126/science.abb2762.

(24) Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV | Cellular & Molecular Immunology https://www.nature.com/articles/s41423-020-0385-z (accessed Mar 9, 2020).

(25) JCI Insight - Predicting the broadly neutralizing antibody susceptibility of the HIV reservoir https://insight.jci.org/articles/view/130153 (accessed Mar 13, 2020).

(26) Hwang, W. C.; Lin, Y.; Santelli, E.; Sui, J.; Jaroszewski, L.; Stec, B.; Farzan, M.; Marasco, W. A.; Liddington, R. C. Structural Basis of Neutralization by a Human Anti-Severe Acute Respiratory Syndrome Spike Protein Antibody, 80R. *J. Biol. Chem.* **2006**, *281* (45), 34610–34616. https://doi.org/10.1074/jbc.M603275200.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(27) Zhang, C.; Zheng, W.; Huang, X.; Bell, E. W.; Zhou, X.; Zhang, Y. Protein Structure and Sequence Re-Analysis of 2019-NCoV Genome Does Not Indicate Snakes as Its Intermediate Host or the Unique Similarity between Its Spike Protein Insertions and HIV-. 11.

(28) Yoon, H.; Macke, J.; West, A. P.; Foley, B.; Bjorkman, P. J.; Korber, B.; Yusim, K. CATNAP: A Tool to Compile, Analyze and Tally Neutralizing Antibody Panels. *Nucleic Acids Res.* **2015**, *43* (W1), W213–W219. https://doi.org/10.1093/nar/gkv404.

(29) CATNAP Tools https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/ (accessed Mar 10, 2020).

(30) Ofek, G.; Tang, M.; Sambor, A.; Katinger, H.; Mascola, J. R.; Wyatt, R.; Kwong, P. D. Structure and Mechanistic Analysis of the Anti-Human Immunodeficiency Virus Type 1 Antibody 2F5 in Complex with Its Gp41 Epitope. *J. Virol.* **2004**, *78* (19), 10724–10737. https://doi.org/10.1128/JVI.78.19.10724-10737.2004.

(31) Kwon, Y. D.; Georgiev, I. S.; Ofek, G.; Zhang, B.; Asokan, M.; Bailer, R. T.; Bao, A.; Caruso, W.; Chen, X.; Choe, M.; Druz, A.; Ko, S.-Y.; Louder, M. K.; McKee, K.; O'Dell, S.; Pegu, A.; Rudicell, R. S.; Shi, W.; Wang, K.; Yang, Y.; Alger, M.; Bender, M. F.; Carlton, K.; Cooper, J. W.; Blinn, J.; Eudailey, J.; Lloyd, K.; Parks, R.; Alam, S. M.; Haynes, B. F.; Padte, N. N.; Yu, J.; Ho, D. D.; Huang, J.; Connors, M.; Schwartz, R. M.; Mascola, J. R.; Kwong, P. D. Optimization of the Solubility of HIV-1-Neutralizing Antibody 10E8 through Somatic Variation and Structure-Based Design. *J. Virol.* **2016**, *90* (13), 5899–5914. https://doi.org/10.1128/JVI.03246-15.

(32) Irimia, A.; Sarkar, A.; Stanfield, R. L.; Wilson, I. A. Crystallographic Identification of Lipid as an Integral Component of the Epitope of HIV Broadly Neutralizing Antibody 4E10. *Immunity* **2016**, *44* (1), 21–31. https://doi.org/10.1016/j.immuni.2015.12.001.

(33) Fleury, D.; Barrère, B.; Bizebard, T.; Daniels, R. S.; Skehel, J. J.; Knossow, M. A Complex of Influenza Hemagglutinin with a Neutralizing Antibody That Binds Outside the Virus Receptor Binding Site. *Nat. Struct. Biol.* **1999**, *6* (6), 530–534. https://doi.org/10.1038/9299.

(34) Pejchal, R.; Gach, J. S.; Brunel, F. M.; Cardoso, R. M.; Stanfield, R. L.; Dawson, P. E.; Burton, D. R.; Zwick, M. B.; Wilson, I. A. A Conformational Switch in Human Immunodeficiency Virus Gp41 Revealed by the Structures of Overlapping Epitopes Recognized by Neutralizing Antibodies. *J. Virol.* **2009**, *83* (17), 8451–8462. https://doi.org/10.1128/JVI.00685-09.

(35) Ekiert, D. C.; Friesen, R. H. E.; Bhabha, G.; Kwaks, T.; Jongeneelen, M.; Yu, W.; Ophorst, C.; Cox, F.; Korse, H. J. W. M.; Brandenburg, B.; Vogels, R.; Brakenhoff, J. P. J.; Kompier, R.; Koldijk, M. H.; Cornelissen, L. A. H. M.; Poon, L. L. M.; Peiris, M.; Koudstaal, W.; Wilson, I. A.; Goudsmit, J. A Highly Conserved Neutralizing Epitope on Group 2 Influenza A Viruses. *Science* **2011**, *333* (6044), 843–850. https://doi.org/10.1126/science.1204839.

(36) Ochoa, W. F.; Kalko, S. G.; Mateu, M. G.; Gomes, P.; Andreu, D.; Domingo, E.; Fita, I.; Verdaguer, N. A Multiply Substituted G–H Loop from Foot-and-Mouth Disease Virus in Complex with a Neutralizing Antibody: A Role for Water Molecules. *J. Gen. Virol.* **2000**, *81* (6), 1495–1505. https://doi.org/10.1099/0022-1317-81-6-1495.

(37) Corti, D.; Voss, J.; Gamblin, S. J.; Codoni, G.; Macagno, A.; Jarrossay, D.; Vachieri, S. G.; Pinna, D.; Minola, A.; Vanzetta, F.; Silacci, C.; Fernandez-Rodriguez, B. M.; Agatic, G.; Bianchi, S.; Giacchetto-Sasselli, I.; Calder, L.; Sallusto, F.; Collins, P.; Haire, L. F.; Temperton, N.; Langedijk, J. P. M.; Skehel, J. J.; Lanzavecchia, A. A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins. *Science* **2011**, *333* (6044), 850–856. https://doi.org/10.1126/science.1205669.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(38) Pejchal, R.; Doores, K. J.; Walker, L. M.; Khayat, R.; Huang, P.-S.; Wang, S.-K.; Stanfield, R. L.; Julien, J.-P.; Ramos, A.; Crispin, M.; Depetris, R.; Katpally, U.; Marozsan, A.; Cupo, A.; Maloveste, S.; Liu, Y.; McBride, R.; Ito, Y.; Sanders, R. W.; Ogohara, C.; Paulson, J. C.; Feizi, T.; Scanlan, C. N.; Wong, C.-H.; Moore, J. P.; Olson, W. C.; Ward, A. B.; Poignard, P.; Schief, W. R.; Burton, D. R.; Wilson, I. A. A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the HIV Glycan Shield. *Science* **2011**, *334* (6059), 1097–1103. https://doi.org/10.1126/science.1213256.

(39) Dias, J. M.; Kuehne, A. I.; Abelson, D. M.; Bale, S.; Wong, A. C.; Halfmann, P.; Muhammad, M. A.; Fusco, M. L.; Zak, S. E.; Kang, E.; Kawaoka, Y.; Chandran, K.; Dye, J. M.; Saphire, E. O. A Shared Structural Solution for Neutralizing Ebolaviruses. *Nat. Struct. Mol. Biol.* **2011**, *18* (12), 1424–1427. https://doi.org/10.1038/nsmb.2150.

(40) Barbey-Martin, C.; Gigant, B.; Bizebard, T.; Calder, L. J.; Wharton, S. A.; Skehel, J. J.; Knossow, M. An Antibody That Prevents the Hemagglutinin Low PH Fusogenic Transition. *Virology* **2002**, *294* (1), 70–74. https://doi.org/10.1006/viro.2001.1320.

(41) Venkatramani, L.; Bochkareva, E.; Lee, J. T.; Gulati, U.; Graeme Laver, W.; Bochkarev, A.; Air, G. M. An Epidemiologically Significant Epitope of a 1998 Human Influenza Virus Neuraminidase Forms a Highly Hydrated Interface in the NA–Antibody Complex. *J. Mol. Biol.* **2006**, *356* (3), 651–663. https://doi.org/10.1016/j.jmb.2005.11.061.

(42) Ekiert, D. C.; Bhabha, G.; Elsliger, M.-A.; Friesen, R. H. E.; Jongeneelen, M.; Throsby, M.; Goudsmit, J.; Wilson, I. A. Antibody Recognition of a Highly Conserved Influenza Virus Epitope. *Science* **2009**, *324* (5924), 246–251. https://doi.org/10.1126/science.1171491.

(43) Fleury, D.; Wharton, S. A.; Skehel, J. J.; Knossow, M.; Bizebard, T. Antigen Distortion Allows Influenza Virus to Escape Neutralization. *Nat. Struct. Biol.* **1998**, *5* (2), 119–123. https://doi.org/10.1038/nsb0298-119.

(44) Lok, S.-M.; Kostyuchenko, V.; Nybakken, G. E.; Holdaway, H. A.; Battisti, A. J.; Sukupolvi-Petty, S.; Sedlak, D.; Fremont, D. H.; Chipman, P. R.; Roehrig, J. T.; Diamond, M. S.; Kuhn, R. J.; Rossmann, M. G. Binding of a Neutralizing Antibody to Dengue Virus Alters the Arrangement of Surface Glycoproteins. *Nat. Struct. Mol. Biol.* **2008**, *15* (3), 312–317. https://doi.org/10.1038/nsmb.1382.

(45) Chi, S.-W.; Maeng, C.-Y.; Kim, S. J.; Oh, M. S.; Ryu, C. J.; Kim, S. J.; Han, K.-H.; Hong, H. J.; Ryu, S. E. Broadly Neutralizing Anti-Hepatitis B Virus Antibody Reveals a Complementarity Determining Region H3 Lid-Opening Mechanism. *Proc. Natl. Acad. Sci.* **2007**, *104* (22), 9230–9235. https://doi.org/10.1073/pnas.0701279104.

(46) Lee, J. E.; Kuehne, A.; Abelson, D. M.; Fusco, M. L.; Hart, M. K.; Saphire, E. O. Complex of a Protective Antibody with Its Ebola Virus GP Peptide Epitope: Unusual Features of a Vλx Light Chain. *J. Mol. Biol.* **2008**, *375* (1), 202–216. https://doi.org/10.1016/j.jmb.2007.10.017.

(47) Azoitei, M. L.; Correia, B. E.; Ban, Y.-E. A.; Carrico, C.; Kalyuzhniy, O.; Chen, L.; Schroeter, A.; Huang, P.-S.; McLellan, J. S.; Kwong, P. D.; Baker, D.; Strong, R. K.; Schief, W. R. Computation-Guided Backbone Grafting of a Discontinuous Motif onto a Protein Scaffold. *Science* **2011**, *334* (6054), 373–376. https://doi.org/10.1126/science.1209368.

(48) Ekiert, D. C.; Kashyap, A. K.; Steel, J.; Rubrum, A.; Bhabha, G.; Khayat, R.; Lee, J. H.; Dillon, M. A.; O'Neil, R. E.; Faynboym, A. M.; Horowitz, M.; Horowitz, L.; Ward, A. B.; Palese, P.; Webby, R.; Lerner, R. A.; Bhatt, R. R.; Wilson, I. A. Cross-Neutralization of Influenza A Viruses Mediated by a Single Antibody Loop. *Nature* **2012**, *489* (7417), 526–532. https://doi.org/10.1038/nature11414.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(49) Lyumkis, D.; Julien, J.-P.; Val, N. de; Cupo, A.; Potter, C. S.; Klasse, P.-J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Carragher, B.; Wilson, I. A.; Ward, A. B. Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. *Science* **2013**, *342* (6165), 1484–1490. https://doi.org/10.1126/science.1245627.

(50) Ménez, R.; Bossus, M.; Muller, B. H.; Sibaï, G.; Dalbon, P.; Ducancel, F.; Jolivet-Reynaud, C.; Stura, E. A. Crystal Structure of a Hydrophobic Immunodominant Antigenic Site on Hepatitis C Virus Core Protein Complexed to Monoclonal Antibody 19D9D6. *J. Immunol.* **2003**, *170* (4), 1917–1924. https://doi.org/10.4049/jimmunol.170.4.1917.

(51) Rini, J. M.; Stanfield, R. L.; Stura, E. A.; Salinas, P. A.; Profy, A. T.; Wilson, I. A. Crystal Structure of a Human Immunodeficiency Virus Type 1 Neutralizing Antibody, 50.1, in Complex with Its V3 Loop Peptide Antigen. *Proc. Natl. Acad. Sci.* **1993**, *90* (13), 6325–6329. https://doi.org/10.1073/pnas.90.13.6325.

(52) Momany, C.; Kovari, L. C.; Prongay, A. J.; Keller, W.; Gitti, R. K.; Lee, B. M.; Gorbalenya, A. E.; Tong, L.; McClure, J.; Ehrlich, L. S.; Summers, M. F.; Carter, C.; Rossmann, M. G. Crystal Structure of Dimeric HIV-1 Capsid Protein. *Nat. Struct. Biol.* **1996**, *3* (9), 763–770. https://doi.org/10.1038/nsb0996-763.

(53) Stanfield, R. L.; Gorny, M. K.; Zolla-Pazner, S.; Wilson, I. A. Crystal Structures of Human Immunodeficiency Virus Type 1 (HIV-1) Neutralizing Antibody 2219 in Complex with Three Different V3 Peptides Reveal a New Binding Mode for HIV-1 Cross-Reactivity. *J. Virol.* **2006**, *80* (12), 6093–6105. https://doi.org/10.1128/JVI.00205-06.

(54) Bryson, S.; Julien, J.-P.; Hynes, R. C.; Pai, E. F. Crystallographic Definition of the Epitope Promiscuity of the Broadly Neutralizing Anti-Human Immunodeficiency Virus Type 1 Antibody 2F5: Vaccine Design Implications. *J. Virol.* **2009**, *83* (22), 11862–11875. https://doi.org/10.1128/JVI.01604-09.

(55) McLellan, J. S.; Correia, B. E.; Chen, M.; Yang, Y.; Graham, B. S.; Schief, W. R.; Kwong, P. D. Design and Characterization of Epitope-Scaffold Immunogens That Present the Motavizumab Epitope from Respiratory Syncytial Virus. *J. Mol. Biol.* **2011**, *409* (5), 853–866. https://doi.org/10.1016/j.jmb.2011.04.044.

(56) Frey, G.; Chen, J.; Rits-Volloch, S.; Freeman, M. M.; Zolla-Pazner, S.; Chen, B. Distinct Conformational States of HIV-1 Gp41 Are Recognized by Neutralizing and Non-Neutralizing Antibodies. *Nat. Struct. Mol. Biol.* **2010**, *17* (12), 1486–1491. https://doi.org/10.1038/nsmb.1950.

(57) Wu, X.; Zhou, T.; Zhu, J.; Zhang, B.; Georgiev, I.; Wang, C.; Chen, X.; Longo, N. S.; Louder, M.; McKee, K.; O'Dell, S.; Perfetto, S.; Schmidt, S. D.; Shi, W.; Wu, L.; Yang, Y.; Yang, Z.-Y.; Yang, Z.; Zhang, Z.; Bonsignori, M.; Crump, J. A.; Kapiga, S. H.; Sam, N. E.; Haynes, B. F.; Simek, M.; Burton, D. R.; Koff, W. C.; Doria-Rose, N. A.; Connors, M.; Program, N. C. S.; Mullikin, J. C.; Nabel, G. J.; Roederer, M.; Shapiro, L.; Kwong, P. D.; Mascola, J. R. Focused Evolution of HIV-1 Neutralizing Antibodies Revealed by Structures and Deep Sequencing. *Science* **2011**, *333* (6049), 1593–1602. https://doi.org/10.1126/science.1207532.

(58) Thomson, C. A.; Bryson, S.; McLean, G. R.; Creagh, A. L.; Pai, E. F.; Schrader, J. W. Germline V-Genes Sculpt the Binding Site of a Family of Antibodies Neutralizing Human Cytomegalovirus. *EMBO J.* **2008**, *27* (19), 2592–2602. https://doi.org/10.1038/emboj.2008.179.

(59) Diskin, R.; Scheid, J. F.; Marcovecchio, P. M.; West, A. P.; Klein, F.; Gao, H.; Gnanapragasam, P. N. P.; Abadir, A.; Seaman, M. S.; Nussenzweig, M. C.; Bjorkman, P. J.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Increasing the Potency and Breadth of an HIV Antibody by Using Structure-Based Rational Design. *Science* **2011**, *334* (6060), 1289–1293. https://doi.org/10.1126/science.1213782.

(60) Cockburn, J. J. B.; Navarro Sanchez, M. E.; Fretes, N.; Urvoas, A.; Staropoli, I.; Kikuti, C. M.; Coffey, L. L.; Arenzana Seisdedos, F.; Bedouelle, H.; Rey, F. A. Mechanism of Dengue Virus Broad Cross-Neutralization by a Monoclonal Antibody. *Structure* **2012**, *20* (2), 303–314. https://doi.org/10.1016/j.str.2012.01.001.

(61) Tugarinov, V.; Zvi, A.; Levy, R.; Hayek, Y.; Matsushita, S.; Anglister, J. NMR Structure of an Anti-Gp120 Antibody Complex with a V3 Peptide Reveals a Surface Important for Co-Receptor Binding. *Struct. Lond. Engl. 1993* **2000**, *8* (4), 385–395. https://doi.org/10.1016/s0969-2126(00)00119-2.

(62) Midgley, C. M.; Flanagan, A.; Tran, H. B.; Dejnirattisai, W.; Chawansuntati, K.; Jumnainsong, A.; Wongwiwat, W.; Duangchinda, T.; Mongkolsapaya, J.; Grimes, J. M.; Screaton, G. R. Structural Analysis of a Dengue Cross-Reactive Antibody Complexed with Envelope Domain III Reveals the Molecular Basis of Cross-Reactivity. *J. Immunol.* **2012**, *188* (10), 4971–4979. https://doi.org/10.4049/jimmunol.1200227.

(63) Reguera, J.; Santiago, C.; Mudgal, G.; Ordoño, D.; Enjuanes, L.; Casasnovas, J. M. Structural Bases of Coronavirus Attachment to Host Aminopeptidase N and Its Inhibition by Neutralizing Antibodies. *PLOS Pathog.* **2012**, *8* (8), e1002859. https://doi.org/10.1371/journal.ppat.1002859.

(64) Zhou, T.; Georgiev, I.; Wu, X.; Yang, Z.-Y.; Dai, K.; Finzi, A.; Kwon, Y. D.; Scheid, J. F.; Shi, W.; Xu, L.; Yang, Y.; Zhu, J.; Nussenzweig, M. C.; Sodroski, J.; Shapiro, L.; Nabel, G. J.; Mascola, J. R.; Kwong, P. D. Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science* **2010**, *329* (5993), 811–817. https://doi.org/10.1126/science.1192819.

(65) Pancera, M.; Shahzad-ul-Hussan, S.; Doria-Rose, N. A.; McLellan, J. S.; Bailer, R. T.; Dai, K.; Loesgen, S.; Louder, M. K.; Staupe, R. P.; Yang, Y.; Zhang, B.; Parks, R.; Eudailey, J.; Lloyd, K. E.; Blinn, J.; Alam, S. M.; Haynes, B. F.; Amin, M. N.; Wang, L.-X.; Burton, D. R.; Koff, W. C.; Nabel, G. J.; Mascola, J. R.; Bewley, C. A.; Kwong, P. D. Structural Basis for Diverse N-Glycan Recognition by HIV-1–Neutralizing V1–V2–Directed Antibody PG16. *Nat. Struct. Mol. Biol.* **2013**, *20* (7), 804–813. https://doi.org/10.1038/nsmb.2600.

(66) Luftig, M. A.; Mattu, M.; Giovine, P. D.; Geleziunas, R.; Hrin, R.; Barbato, G.; Bianchi, E.; Miller, M. D.; Pessi, A.; Carfí, A. Structural Basis for HIV-1 Neutralization by a Gp41 Fusion Intermediate–Directed Antibody. *Nat. Struct. Mol. Biol.* **2006**, *13* (8), 740–747. https://doi.org/10.1038/nsmb1127.

(67) Lee, C.-C.; Lin, L.-L.; Chan, W.-E.; Ko, T.-P.; Lai, J.-S.; Wang, A. H.-J. Structural Basis for the Antibody Neutralization of Herpes Simplex Virus. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (10), 1935–1945. https://doi.org/10.1107/S0907444913016776.

(68) Su, H.-P.; Golden, J. W.; Gittis, A. G.; Hooper, J. W.; Garboczi, D. N. Structural Basis for the Binding of the Neutralizing Antibody, 7D11, to the Poxvirus L1 Protein. *Virology* **2007**, *368* (2), 331–341. https://doi.org/10.1016/j.virol.2007.06.042.

(69) Tang, X.; Yang, C.; Gu, Y.; Song, C.; Zhang, X.; Wang, Y.; Zhang, J.; Hew, C. L.; Li, S.; Xia, N.; Sivaraman, J. Structural Basis for the Neutralization and Genotype Specificity of Hepatitis E Virus. *Proc. Natl. Acad. Sci.* **2011**, *108* (25), 10266–10271. https://doi.org/10.1073/pnas.1101309108.

(70) Cherrier, M. V.; Kaufmann, B.; Nybakken, G. E.; Lok, S.-M.; Warren, J. T.; Chen, B. R.; Nelson, C. A.; Kostyuchenko, V. A.; Holdaway, H. A.; Chipman, P. R.; Kuhn, R. J.; Diamond, M. S.; Rossmann, M. G.; Fremont, D. H. Structural Basis for the Preferential

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Recognition of Immature Flaviviruses by a Fusion-Loop Antibody. *EMBO J.* **2009**, *28* (20), 3269–3276. https://doi.org/10.1038/emboj.2009.245.

(71) Austin, S. K.; Dowd, K. A.; Shrestha, B.; Nelson, C. A.; Edeling, M. A.; Johnson, S.; Pierson, T. C.; Diamond, M. S.; Fremont, D. H. Structural Basis of Differential Neutralization of DENV-1 Genotypes by an Antibody That Recognizes a Cryptic Epitope. *PLOS Pathog.* **2012**, *8* (10), e1002930. https://doi.org/10.1371/journal.ppat.1002930.

(72) Kong, L.; Giang, E.; Robbins, J. B.; Stanfield, R. L.; Burton, D. R.; Wilson, I. A.; Law, M. Structural Basis of Hepatitis C Virus Neutralization by Broadly Neutralizing Antibody HCV1. *Proc. Natl. Acad. Sci.* **2012**, *109* (24), 9499–9504. https://doi.org/10.1073/pnas.1202924109.

(73) Nybakken, G. E.; Oliphant, T.; Johnson, S.; Burke, S.; Diamond, M. S.; Fremont, D. H. Structural Basis of West Nile Virus Neutralization by a Therapeutic Antibody. *Nature* **2005**, *437* (7059), 764–769. https://doi.org/10.1038/nature03956.

(74) Zhou, T.; Xu, L.; Dey, B.; Hessell, A. J.; Ryk, D. V.; Xiang, S.-H.; Yang, X.; Zhang, M.-Y.; Zwick, M. B.; Arthos, J.; Burton, D. R.; Dimitrov, D. S.; Sodroski, J.; Wyatt, R.; Nabel, G. J.; Kwong, P. D. Structural Definition of a Conserved Neutralization Epitope on HIV-1 Gp120. *Nature* **2007**, *445* (7129), 732–737. https://doi.org/10.1038/nature05580.

(75) Fleury, D.; Daniels, R. S.; Skehel, J. J.; Knossow, M.; Bizebard, T. Structural Evidence for Recognition of a Single Epitope by Two Distinct Antibodies. *Proteins* **2000**, *40* (4), 572–578.

(76) Pak, J. E.; Sharon, C.; Satkunarajah, M.; Auperin, T. C.; Cameron, C. M.; Kelvin, D. J.; Seetharaman, J.; Cochrane, A.; Plummer, F. A.; Berry, J. D.; Rini, J. M. Structural Insights into Immune Recognition of the Severe Acute Respiratory Syndrome Coronavirus S Protein Receptor Binding Domain. *J. Mol. Biol.* **2009**, *388* (4), 815–823. https://doi.org/10.1016/j.jmb.2009.03.042.

(77) Wu, Y.; Eigenbrot, C.; Liang, W.-C.; Stawicki, S.; Shia, S.; Fan, B.; Ganesan, R.; Lipari, M. T.; Kirchhofer, D. Structural Insight into Distinct Mechanisms of Protease Inhibition by Antibodies. *Proc. Natl. Acad. Sci.* **2007**, *104* (50), 19784–19789. https://doi.org/10.1073/pnas.0708251104.

(78) Cockburn, J. J.; Navarro Sanchez, M. E.; Goncalvez, A. P.; Zaitseva, E.; Stura, E. A.; Kikuti, C. M.; Duquerroy, S.; Dussart, P.; Chernomordik, L. V.; Lai, C.-J.; Rey, F. A. Structural Insights into the Neutralization Mechanism of a Higher Primate Antibody against Dengue Virus. *EMBO J.* **2012**, *31* (3), 767–779. https://doi.org/10.1038/emboj.2011.439.

(79) Backovic, M.; DuBois, R. M.; Cockburn, J. J.; Sharff, A. J.; Vaney, M.-C.; Granzow, H.; Klupp, B. G.; Bricogne, G.; Mettenleiter, T. C.; Rey, F. A. Structure of a Core Fragment of Glycoprotein H from Pseudorabies Virus in Complex with Antibody. *Proc. Natl. Acad. Sci.* **2010**, *107* (52), 22635–22640. https://doi.org/10.1073/pnas.1011507107.

(80) McLellan, J. S.; Chen, M.; Chang, J.-S.; Yang, Y.; Kim, A.; Graham, B. S.; Kwong, P. D. Structure of a Major Antigenic Site on the Respiratory Syncytial Virus Fusion Glycoprotein in Complex with Neutralizing Antibody 101F. *J. Virol.* **2010**, *84* (23), 12236–12244. https://doi.org/10.1128/JVI.01579-10.

(81) Prabakaran, P.; Gan, J.; Feng, Y.; Zhu, Z.; Choudhry, V.; Xiao, X.; Ji, X.; Dimitrov, D. S. Structure of Severe Acute Respiratory Syndrome Coronavirus Receptor-Binding Domain Complexed with Neutralizing Antibody. *J. Biol. Chem.* **2006**, *281* (23), 15829–15836. https://doi.org/10.1074/jbc.M600697200.

(82) Lee, J. E.; Fusco, M. L.; Hessell, A. J.; Oswald, W. B.; Burton, D. R.; Saphire, E. O. Structure of the Ebola Virus Glycoprotein Bound to an Antibody from a Human Survivor. *Nature* **2008**, *454* (7201), 177–182. https://doi.org/10.1038/nature07082.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(83) Hu, G.; Liu, J.; Roux, K. H.; Taylor, K. A. Structure of Simian Immunodeficiency Virus Envelope Spikes Bound with CD4 and Monoclonal Antibody 36D5. *J. Virol.* **2017**, *91* (16). https://doi.org/10.1128/JVI.00134-17.

(84) Huang, C.; Lam, S. N.; Acharya, P.; Tang, M.; Xiang, S.-H.; Hussan, S. S.; Stanfield, R. L.; Robinson, J.; Sodroski, J.; Wilson, I. A.; Wyatt, R.; Bewley, C. A.; Kwong, P. D. Structures of the CCR5 N Terminus and of a Tyrosine-Sulfated Antibody with HIV-1 Gp120 and CD4. *Science* **2007**, *317* (5846), 1930–1934. https://doi.org/10.1126/science.1145373.

(85) Malby, R. L.; Tulip, W. R.; Harley, V. R.; McKimm-Breschkin, J. L.; Laver, W. G.; Webster, R. G.; Colman, P. M. The Structure of a Complex between the NC10 Antibody and Influenza Virus Neuraminidase and Comparison with the Overlapping Binding Site of the NC41 Antibody. *Structure* **1994**, *2* (8), 733–746. https://doi.org/10.1016/S0969-2126(00)00074-5.

(86) Walls, A. C.; Xiong, X.; Park, Y.-J.; Tortorici, M. A.; Snijder, J.; Quispe, J.; Cameroni, E.; Gopal, R.; Dai, M.; Lanzavecchia, A.; Zambon, M.; Rey, F. A.; Corti, D.; Veesler, D. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **2019**, *176* (5), 1026-1039.e15. https://doi.org/10.1016/j.cell.2018.12.028.

(87) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.

(88) Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Advances in Neural Information Processing Systems 28*; Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc., 2015; pp 2224–2232.

(89) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. https://doi.org/10.1021/ci100050t.

(90) RDKit https://www.rdkit.org/ (accessed Mar 14, 2020).

(91) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *ArXiv170300564 Phys. Stat* **2018**.

(92) Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z. *Deep Learning for the Life Sciences*; O'Reilly Media, 2019.

(93) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; KDD '16; Association for Computing Machinery: San Francisco, California, USA, 2016; pp 785–794. https://doi.org/10.1145/2939672.2939785.

(94) Breiman, L. Random Forests. *Mach. Lang.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(95) Neffati, S.; Ben Abdellafou, K.; Taouali, O.; Bouzrara, K. Enhanced SVM–KPCA Method for Brain MR Image Classification. *Comput. J.* https://doi.org/10.1093/comjnl/bxz035.

(96) He, R.; Dobie, F.; Ballantine, M.; Leeson, A.; Li, Y.; Bastien, N.; Cutts, T.; Andonov, A.; Cao, J.; Booth, T. F.; Plummer, F. A.; Tyler, S.; Baker, L.; Li, X. Analysis of Multimerization of the SARS Coronavirus Nucleocapsid Protein. *Biochem. Biophys. Res. Commun.* **2004**, *316* (2), 476–483. https://doi.org/10.1016/j.bbrc.2004.02.074.

(97) Henikoff, S.; Henikoff, J. G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (22), 10915–10919.

(98) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.14.992156. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

(99) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.
(100) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(101) Lindahl; Abraham; Hess; van der Spoel. *GROMACS 2020.1 Source Code*; Zenodo, 2020. https://doi.org/10.5281/zenodo.3685919.
(102) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40* (7), 843. https://doi.org/10.1007/s00249-011-0700-9.
(103) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
(104) Darden, T.; York, D. M.; Pedersen, L. G. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems; 1993. https://doi.org/10.1063/1.464397.
(105) Lingenheil, M.; Denschlag, R.; Reichold, R.; Tavan, P. The "Hot-Solvent/Cold-Solute" Problem Revisited. *J. Chem. Theory Comput.* **2008**, *4* (8), 1293–1306. https://doi.org/10.1021/ct8000365.
(106) Simulations of proteins with inhomogeneous degrees of freedom: The effect of thermostats - Mor - 2008 - Journal of Computational Chemistry - Wiley Online Library https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20951 (accessed Mar 13, 2020).
(107) (IUCr) Coot: model-building tools for molecular graphics https://onlinelibrary.wiley.com/iucr/doi/10.1107/S0907444904019158 (accessed Mar 9, 2020).
(108) (IUCr) Features and development of Coot https://onlinelibrary.wiley.com/iucr/doi/10.1107/S0907444910007493 (accessed Mar 13, 2020).
(109) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.
(110) Native contacts determine protein folding mechanisms in atomistic simulations | PNAS https://www.pnas.org/content/110/44/17874 (accessed Mar 13, 2020).

Supporting Information
Potential Neutralizing Antibodies Discovered for Novel Corona Virus Using Machine Learning

Rishikesh Magar [¥], *Prakarsh Yadav* [‡] *and Amir Barati Farimani* [1¥§‡]

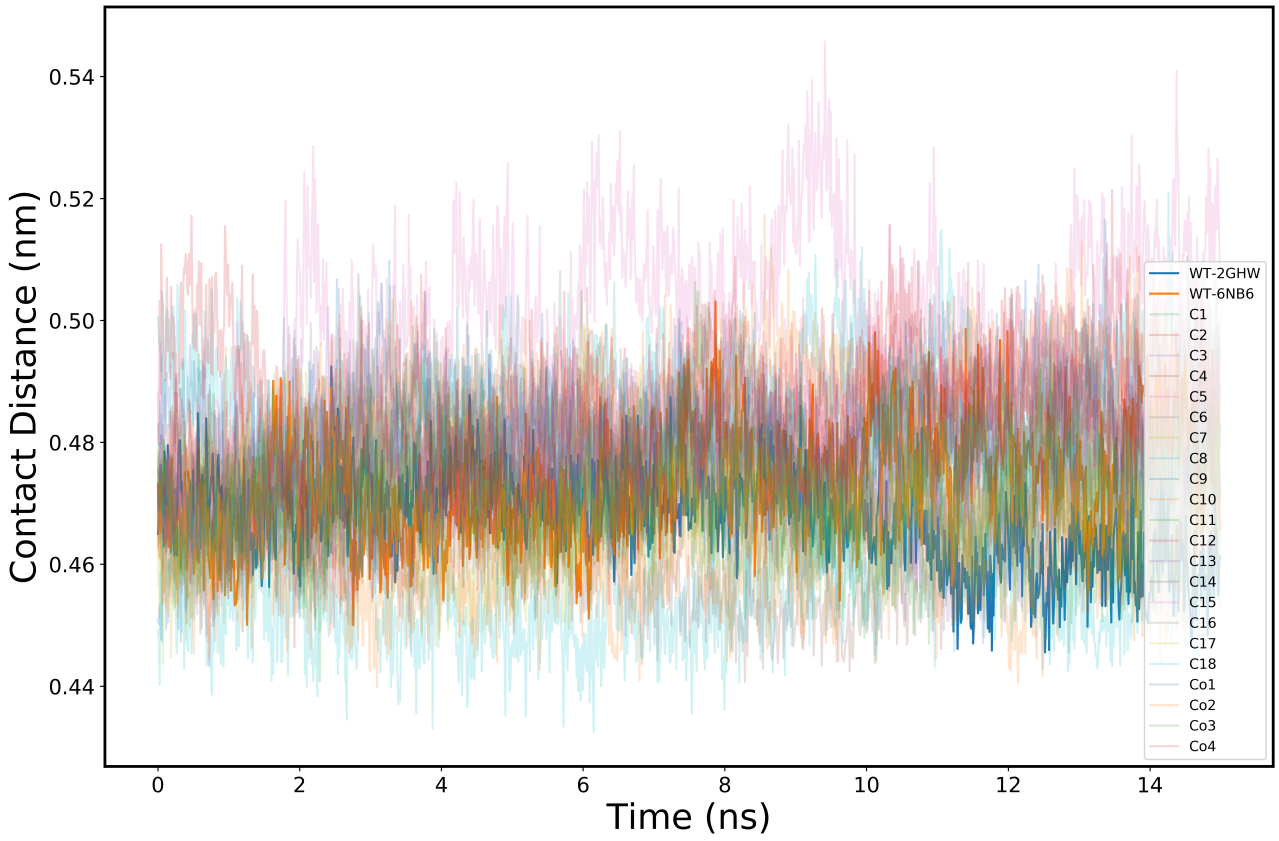
¥ Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

‡ Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

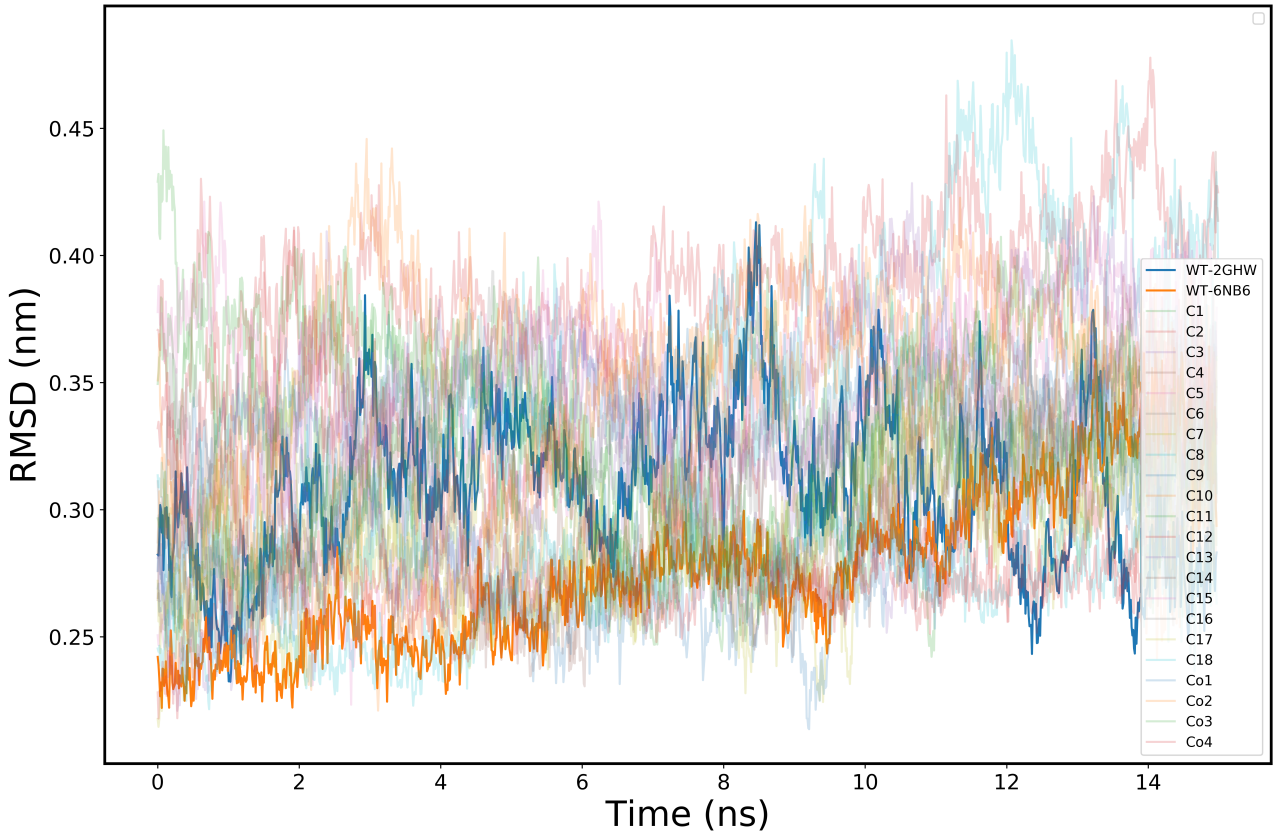
§Machine Learning Department, School of Computer Science, Carnegie Mellon University Pittsburgh, PA 15213

[1] Corresponding Author, e-mail: barati@cmu.edu

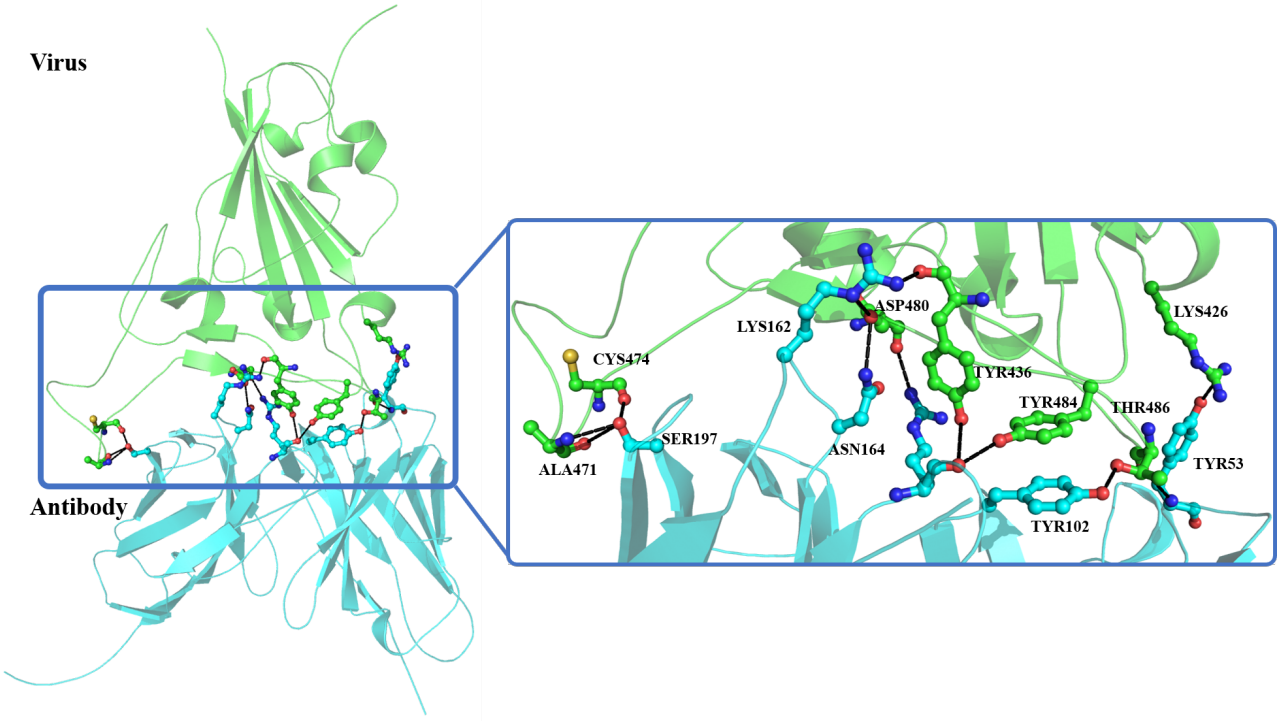
Supplementary Figure 1. Contact Distance plot of trajectories of all the 22 mutant structures and the crystal structures 2GHW and 6NB6. The contact distances were computed using MDTraj package. Instantaneous contact distances for each timestep are shown for the course of the simulation (15 ns). The contact distance values vary within a small range for all the structures, indicating that simulations did not have any abrupt changes.



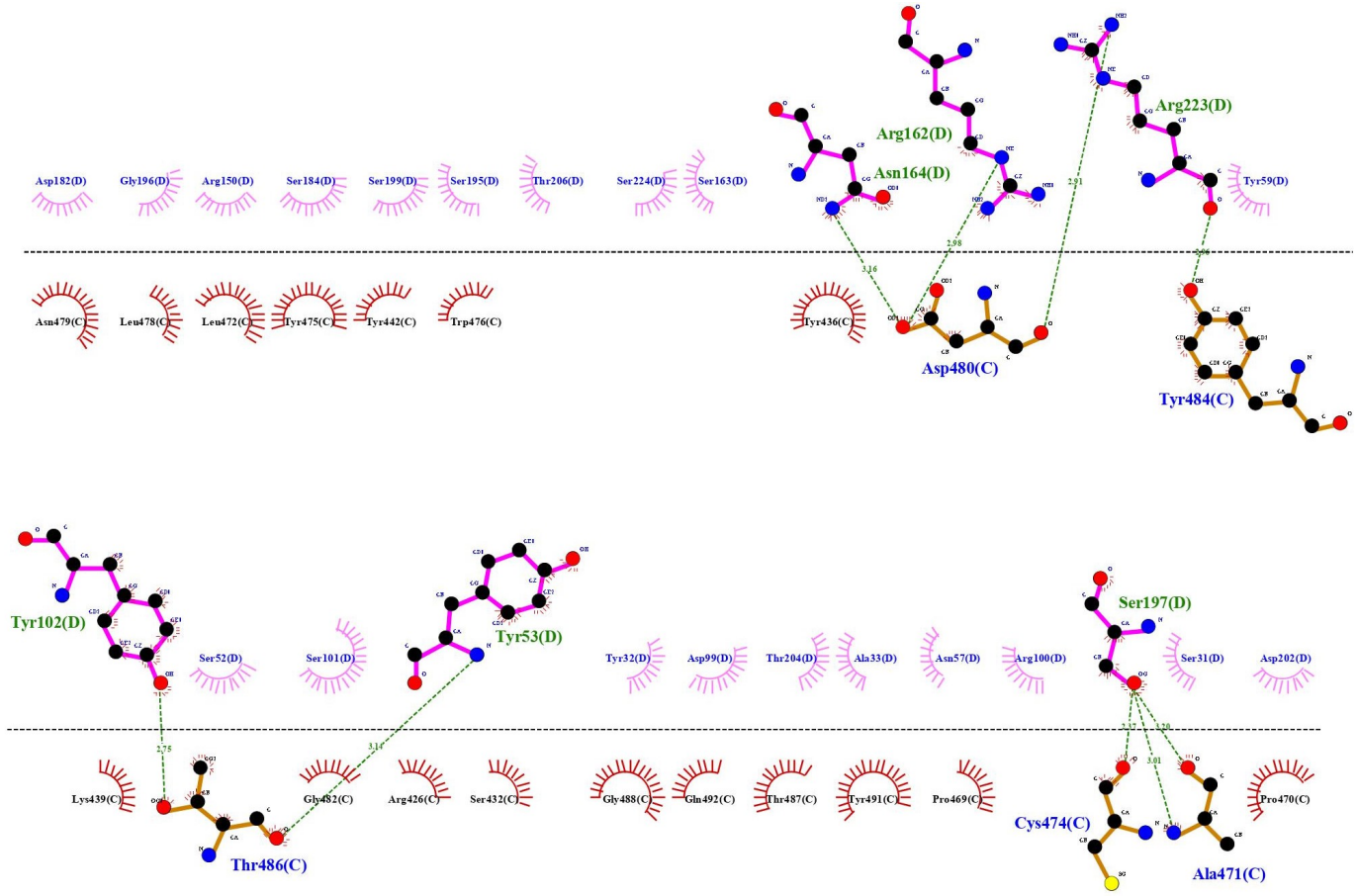
Supplementary Figure 2. Root Mean Square Deviation (RMSD) plot of trajectories of all the 22 structures and the crystal structures 2GHW and 6NB6. The deviations are within acceptable range indicating the strucutres were stable over the course of simulation. RMSD was averaged over the entire simultation and subsequently used for generation of Figure 4b.



Supplementary Figure 3. Structure of virus antigen and antibody along with the epitope interactions in the crystal structure PDB ID: 2GHW. The viral antigen is in green and antibody is in cyan. Black lines indicate electrostatic interactions between the antigen epitope and the antibody. These interactions play a significant role in the specificity of antigen recognition by antibody.



Supplementary Figure 4. Native contacts between the antibody and antigen in the crystal structure (PDB ID: 2GHW) shown pictographically. The amino acids above black dotted line represents the antibody (pink) and below it is the viral epitope (brown). Electrostatic interactions are represented by the whole amino acid ball-stick diagram and are connected by green dotted lines.



Supplementary Figure 5. Predicted contacts between the antibody and COVID-19 antigen. The COVID antigen is overlaid onto the crystal structure antigen (2GHW) to conserve the geometry and make predictions. The amino acids above black dotted line represents the antibody (pink) from PDB: 2GHW, and below is the viral epitope(brown) from COVID-19 sequence. Electrostatic interactions are represented by the whole amino acid ball-stick diagram and are connected by green dotted lines.

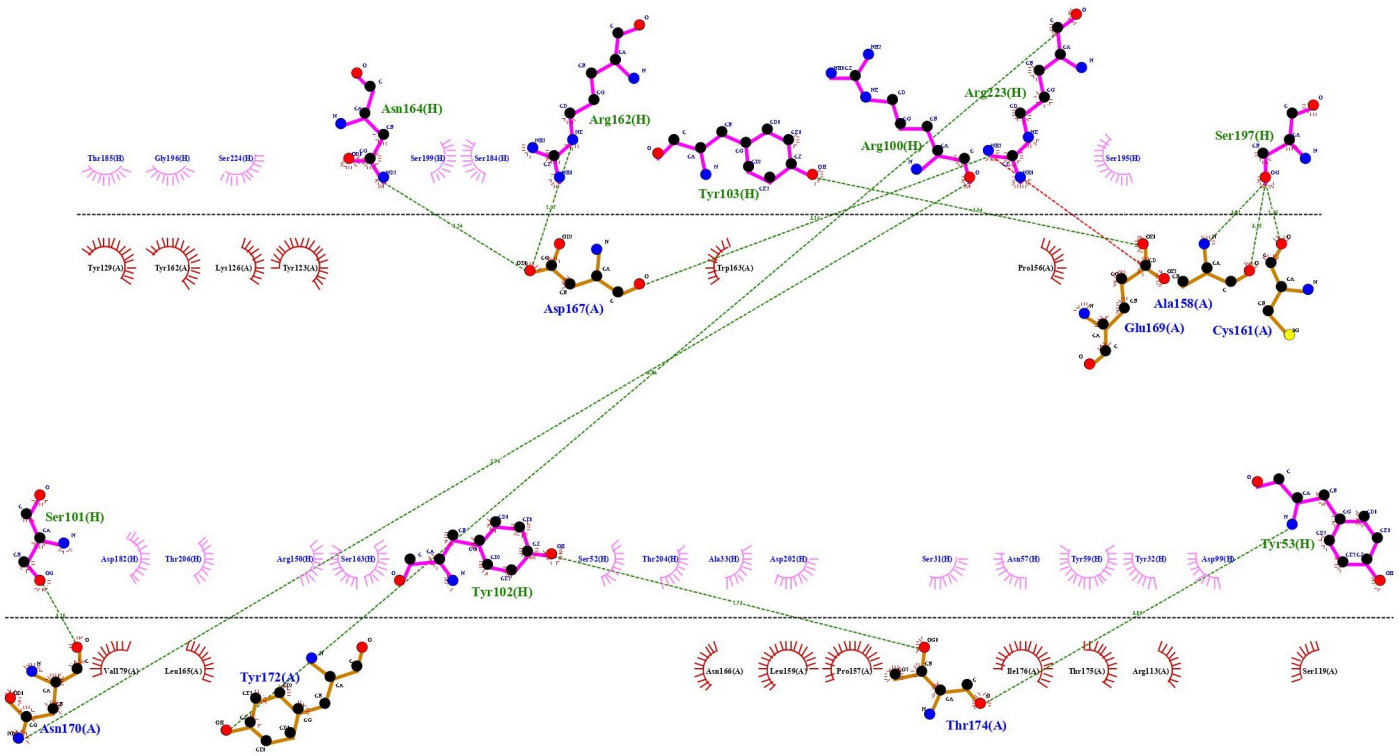


Table S1: The neutralizing candidates obtained before screening for stability, and after screening with ML and Bioinformatics. There are 18-point mutations in total which are predicted to have a neutralizing effect on COVID-19.

Structure	Mutation
C1	2GHW – A33C
C2	2GHW – V50M
C3	2GHW – I51M
C4	2GHW – I51V
C5	2GHW – N57H
C6	2GHW – R100H
C7	2GHW – R150H
C8	2GHW – V161M
C9	2GHW – R162H
C10	2GHW – N164H
C11	2GHW – T185N
C12	2GHW – R186H
C13	2GHW – F203M
C14	2GHW – T204N
C15	2GHW – T206N
C16	2GHW – R223H
C17	6NB6 – R56H
C18	6NB6 – K58S

Table S2: The possible co-mutations from 5 stable neutralizing candidates. Structures Co1 – Co4 were predicted as neutralizing, however; structure Co5 was predicted as non-neutralizing by the XGBoost model. All these co-mutant structures we MD validated for their stability.

Structure	Mutation	Neutralization Potential
Co1	2GHW – I51M, R150H, T204N	1
Co2	2GHW – I51M, R150H	1
Co3	2GHW – I51M, T204N	1
Co4	6NB6 – R56H, K58S	1
Co5	2GHW – R150H, T204N	0

Supplementary Data. Stable_str.zip. The zip file contains the PDB coordinates for the final frame of the simulation of each of the neutralizing candidate antibody. All the mutant antibodies are structurally intact and are stable.

Supplementary Methods

IC_{50} Value interpretation

The IC_{50} values in dataset varied from 0 to >50. For simplicity of analysis, we classified the IC_{50} values into two categories. Where, $IC_{50} > 10$ would indicate a non-neutralizing antibody and $IC_{50} < 10$ would indicate a neutralizing antibody. Many viruses had the same viral epitope and yet had very different IC_{50} values. As a result, the model had to predict two different labels for the same antibody-antigen combinations. We resolved this, by taking a majority vote among the conflicts and transformed the labels which do not concur with the majority. Employing this scheme improved the results of our model.