Supporting information for Emerging vaccine-breakthrough SARS-CoV-2 variants

Rui Wang[1], Jiahui Chen[1], Yuta Hozumi[1], Changchuan Yin[2], and Guo-Wei Wei[1,3,4*]
[1] Department of Mathematics,
Michigan State University, MI 48824, USA.
[2] Department of Mathematics, Statistics, and Computer Science,
University of Illinois at Chicago, Chicago, IL 60607, USA
[3] Department of Electrical and Computer Engineering,
Michigan State University, MI 48824, USA.
[4] Department of Biochemistry and Molecular Biology,
Michigan State University, MI 48824, USA.

January 22, 2022

Contents

*Corresponding author. Email: weig@msu.edu

S1 Overview of SARS-CoV-2 prevailing and emerging variants

Starting in early January 2021, several SARS-CoV-2 vaccines have been authorized for full or emergence use. The majority of the vaccines are target to S protein of SARS-CoV-2, which aims to induce neutralizing antibodies against the viral S protein, alleviate COVID-19 symptoms and severity, and block infection. However, the emergence of new variants has assistant SARS-CoV-2 to fight back the human immune systems built by vaccines. It is noteworthy that the alarming rise in the number of new cases due to several viral variants threatens the efficacy of existing vaccines in late June 2021, which poses significant challenges to public health. Therefore, tracking the specific emerging SARS-CoV-2 variants of interest (VOIs) and variants of concern (VOCs), and understanding their potential effects on vaccines and antibody therapies is urgently needed. WHO has been established a naming system using the letters of the Greek alphabet for VOIs and VOCs, aiming to create easy-to-pronounce and non-stigmatizing labels that will be easier and more practical to be discussed. At this present, the potential effects on existing vaccines and antibody therapies of ten variants have been widely discussed: Alpha, Beta, Gamma, Delta, Eta, Iota, Kappa, Lambda, Delta plus, and Beta plus.

The S protein of SARS-CoV-2 is the majority target for vaccines due to the receptor-binding domain (RBD) located on the S protein. RBD on SARS-CoV-2 binds strongly to human ACE2 and exhibits significantly high binding affinity to ACE2, which makes S protein RBD an appealing and dominant antigenic target for inducing SARS-CoV-2 neutralizing antibodies [24]. Therefore, prioritizing the analysis of mutations on S protein RBD for different variants is essential for studying the viral infectivity, mAb efficacy, vaccine efficacy and protection rate. Studies have shown that the binding free energy (BFE) between the S RBD and the ACE2 is relevant to the infectivity [11, 13, 20, 22, 25]. In this work, we will apply our TopNet-Tree model to predict the RBD mutation-induced BFE changes of S protein with RBD and 130 existing antibodies based on 1,489,884 complete SARS-CoV-2 genome sequences from the GISAID database [21] up to August 5, 2021. At this stage, a total of 683 non-degenerate RBD mutations are detected, which can be found on Mutation Analyzer. Here, the positive BFE change of a specific mutation means it will strengthen the binding between S protein and ACE2, suggesting a more infectious variant. Notably, the count of antibody disruption represents the number of antibodies and S protein complexes disrupted by a specific mutation. If a binding strength between S protein and antibody is reduced by more than 0.3 kcal/mol by a specific mutation, the binding will be considered as disrupted. From Mutation Analyzer, it can be seen that the antibody disruption ratios of mutation L452R, E484K, K417T, K417N, F490S, S494P are greater than 30%, and the disruption ratios of mutation E484K, K417T are nearly 30%, suggesting disruptive effects on the efficacy and reliability of antibody therapies and vaccines. Moreover, all of the top 10 most observed RBD mutations are ACE2 binding-strengthening mutations, and the BFE changes of S protein and ACE2 for mutation T478K is nearly 1.00 kcal/mol, which results in a strong binding performance and enhance the stabilization of the RBDACE2 complex [9].

Although a variant is defined by its mutations on the whole SARS-CoV-2 genome, in this work, we will only focus on the S RBD co-mutations of different variants, due to the fact that RBD contains the majority of neutralizing epitopes and is the most dominant antigenic domain for inducing SARS-CoV-2 neutralizing antibodies [12]. Table S1 demonstrates the co-mutations on S protein RBD, the first detected country, and the earliest detection date of the current VOCs and VOIs, and their corresponding upgraded variants. The alpha variant was first identified in the United Kingdom (UK) in September 2020, which carries a single mutation N501Y on S RBD. The alpha variant was dominant in the UK in mid-March 2021, and since August 23, 2021, 268,099 patients were infected by the alpha variant in the UK according to GISAID. Beta variant (Lineage B.1.3514) was first found in South Africa and was characterized by three mutations on

S RBD: K417N, E484K, and N501Y. Up to August 23, 2021, a total of 6,368 patients carry beta variants in South Africa. In addition, an additional mutation on RBD P438L on Beta variants is called Beta plus variants. Gamma variant is characterized by its RBD mutations K417T, E484K, and N501Y, which were first found in Brazil in November 2020. At this stage, the gamma variant is still one of the prevailing variants in Brazil. Delta variant was characterized by its RBD co-mutations L452R and T478K, which was first detected in India in October 2020, and then rapidly spread out around the world, which is a preponderance of SARS-CoV-2 variants. In the past four weeks, 91.9% of new cases in the world are delta variants. Some evidence has shown that delta variant will cause immune escape and has adverse effects on neutralizing antibodies [16]. Notably, Delta plus variants that are identified by an additional K417N substitution on RBD also draw people's attention. Moreover, two VOIs, Eta and Iota, were identified with E484K mutation on S RBD in late 2020. Kappa variants were characterized by L452R, E484Q substitutions on RBD, which was first detected in India in October 2020. Furthermore, in late 2020, a new variant Lambda was found in Peru, which carries two mutations on S RBD that are different from other COVID-19 variants: L452Q and F490S. Some preliminary studies have reported that the S protein of Lambda variant may make SARS-CoV-2 easier to latch onto cells and evade vaccines, which may be a potential threat to the human society [18].

Table S1: The RBD co-mutations, the first detected country, and the earliest detection date of ten VOCs, VOIs, and their corresponding upgraded variants: Alpha, Beta, Gamma, Delta, Eta, Iota, Kappa, Lambda, Delta plus, and Beta plus.

Variants	Categories	Co-mutations on RBD	Country	Date
Alpha	VOC	[N501Y]	United Kingdom	2020-09
Beta	VOC	[K417N, E484K, N501Y]	South Africa	2020-05
Gamma	VOC	[K417T, E484K, N501Y]	Brazil	2020-11
Delta	VOC	[L452R, T478K]	India	2020-10
Eta	VOI	[E484K]	Multiple countries	2020-12
Iota	VOI	[E484K]	United States	2020-11
Kappa	VOI	[L452R, E484Q]	India	2020-10
Lambda	VOI	[L452Q, F490S]	Peru	2020-12
Delta plus	Upgraded variants	[K417N, L452R, T478K]	/	/
Beta plus	Upgraded variants	[P384L, K417N, E484K, N501Y]	/	/

S2 Supplementary data

The Supplementary_Data.zip contains two folders. One is a folder with 4 csv/xlsx files and the other is a folder with 29 HTML files.

S2.1 Excel files

A total of 4 Excel files are in this folder.

S2.1.1 Disrupted antibodies

File antibodies_disruptmutation.csv shows the name of antibodies disrupted by mutations.

S2.1.2 List of antibodies

File antibodies.csv lists the Protein Data Bank (PDB) IDs for all of the 130 SARS-CoV-2 antibodies.

S2.1.3 SNPs

File RBD_comutation_residue_08052021.csv lists all of the SNPs of RBD co-mutations.

S2.1.4 Non-degenerate RBD co-mutations

File Track_Comutation_08052021.xlsx records all of the non-degenerate RBD co-mutations with their frequencies, antibody disruption counts, total BFE changes, and the first detection dates and countries.

S2.2 HTML files

A total of 29 HTML files are given.

S2.2.1 HTML files

Twenty HTML files for the time evolution of 2, 3, and 4 co-mutations on the S protein RBD of SARS-CoV-2 from January 01, 2021 to July 31, 2021, in 20COVID-19 devastated countries: the United Kingdom (UK), the United States (US), Denmark (DK), Brazil (BR), Germany (DE), Netherlands (NL), Sweden (SE), Italy (IT), Canada (CA), France (FR), India (IN), Belgium (BE), Ireland (IE), Spain (ES), Chile (CL), Portugal (PT), Mexico (MX), Singapore (SG), Turkey (TR), and Finland (FL).

S2.2.2 2D histograms

Three 2D histograms are given for antibody disruption counts and total BFE changes for RBD 2 co-mutations, 3 co-mutations, and 4 co-mutations.

S2.2.3 Histograms

Three histograms of total BFE changes, antibody disruption count, and natural log of frequencies for RBD 2 co-mutations, 3 co-mutations, and 4 co-mutations.

S2.2.4 Barplots

Three barplots for RBD 2, 3, and 4 co-mutations with a frequency greater than 90, 30, and 20, respectively.

S3 Supplementary figures

Figure S1 illustrates the the time evolution of 2, 3, and 4 co-mutations on the S protein RBD of SARS-CoV-2 from January 01, 2021, to July 31, 2021, in 8 COVID-19 devastated countries: Ireland (IE), Spain (ES), Chile (CL), Portugal (PT), Mexico (MX), Singapore (SG), Turkey (TR), and Finland (FL). The Delta variant is the most popular one in these countries, except for Chile and Mexico. Gamma has the hishest population in Chile and Mexico, which may change soon. It worthy to note that Chile has a relatively large population of [L452Q, F490S], the key RBD co-mutation in the Lambda variant.

Figure S1: Illustration of the time evolution of 2, 3, and 4 co-mutations on the S protein RBD of SARS-CoV-2 from January 01, 2021, to July 31, 2021, in 8 COVID-19 devastated countries: Ireland (IE), Spain (ES), Chile (CL), Portugal (PT), Mexico (MX), Singapore (SG), Turkey (TR), and Finland (FL). The y-axis represents the natural log frequency of each RBD co-mutation. The top 5 high-frequency co-mutations in each country are marked by red, blue, green, yellow, and pink lines. The cyan line is for the RBD co-mutation [L452Q, F490S] on the Lambda variant, and the other co-mutations are marked by light grey lines. (Please check the interactive HTML files in the Supplementary Data for a better view of these plots.)

S4 Supplementary feature generation

The feature generation of protein-protein interactions is crucial for BFE change predictions induced by mutations. In the main content, we briefly discuss the topology theory and its implementation in this model. There are other features, named as auxiliary features, which also play important roles in the training process. The auxiliary features includes chemical and physical information of the complexes, such as molecular surface areas, partial charges, Coulomb interactions, van der Waals interaction, mutation site neighborhood amino acid composition, pKa shifts, electrostatic solvation free energy, and secondary structure informa-

tion [7, 26]. Two different levels of features which are residue-level and atom-level and will be discussed following.

S4.1 Residue-level features

Mutation site neighborhood amino acid composition Neighbor residues are the residues within 10 Å of the mutation site. Distances between residues are calculated based on residue C_α atoms. Six categories of amino acid residues are counted, which are hydrophobic, polar, positively charged, negatively charged, special cases, and pharmacophore changes. The count and percentage of the 6 amino acid groups in the neighbor site are regrading as the environment composition features of the mutation site. The sum, average, and variance of residue volumes, surface areas, weights, and hydropathy scores are used but only the sum of charges is included.

pKa shifts The pKa values are calculated by the PROPKA software [3], namely the values of 7 ionizable amino acids, namely, ASP, GLU, ARG, LYS, HIS, CYS, and TYR. The maximum, minimum, sum, the sum of absolute values, and the minimum of the absolute value of total pKa shifts are calculated. We also consider the difference of pKa values between a wild type and its mutant. Additionally, the sum and the sum of the absolute value of pKa shifts based on ionizable amino acid groups are included.

Position-specific scoring matrix (PSSM) Features are computed from the conservation scores in the position-specific scoring matrix of the mutation site for the wild type and the mutant as well as their difference. The conservation scores are generated by PSI-BLAST [2].

Secondary structure The SPIDER2 software is used to compute the probability scores for residue torsion angle and residues being in a coil, alpha helix, and beta strand based on the sequences for the wild type and the mutant [29].

S4.2 Atom-level features

Seven groups of atom types, including C, N, O, S, H, all heavy atoms, and all atoms, are considered when generating the element-type features. Meanwhile, other three atom types, i.e., mutation site atoms, all heavy atoms, and all atoms, are used when generating the general atom-level features.

Surface areas Atom-level solvent excluded surface areas are computed by ESES [15].

Partial changes Partial change of each atom is generated by pdb2pqr software [10] using the Amber force field [5] for wild type and CHARMM force field [4] for mutant. The sum of the partial charges and the sum of absolute values of partial charges for each atomic group are collected.

Atomic pairwise interaction interactions Coulomb energy of the ith single atom is calculated as the sum of pairwise coulomb energy with every other atom as

$$C_i = \sum_{j,j\neq i} k_e \frac{q_i q_j}{r_{ij}}, \tag{1}$$

where k_e is the Coulomb's constant, r_{ij} is the distance of ith atom to jth atom, and q_i is the charge of ith atom. The van der Waals energy of the ith atom is modeled as the sum of pairwise Lennard-Jones potentials with other atoms as

$$V_i = \sum_{j,j\neq i} \epsilon\Big[\big(\frac{r_i + r_j}{r_{ij}}\big)^{12} - 2\big(\frac{r_i + r_j}{r_{ij}}\big)^{6}\Big], \tag{2}$$

where ϵ is the depth of the potential well, and r_i is van der Waals radii.

In atomic pairwise interaction, 5 groups (C, N, O, S, and all heavy atoms) are counted both for Coulomb interaction energy and van der Waals interaction energy.

Electrostatic solvation free energy Electrostatic solvation free energy of each atom is calculated using the Poisson-Boltzmann equation via MIBPB [6] and are summed up by atom groups.

S5 Supplementary machine learning methods

The topology-based network model for BFE change predictions induced mutations on SARS-CoV-2 studying applies a deep neural network structure. Similar approaches have been widely implemented in the energy prediction of protein-ligand binding [17] and protein-protein interactions [26]. The neural network method maps an input feature layer to output layer and mimics biological brains for solving problems where numerous neuron units are involved and weights of neurons are updated by backpropagation methods. To make more complicated structure in order to extract abstract properties, more layers and more neurons in each layer can be constructed. In the training process, optimization methods are applied to avoid overfitting issue, such as dropout methods [23] that a partial of computed neurons of each layer is dropped. For the model cross validations, the Pearson correlations of 10-fold cross validations is 0.864 and root mean square error is 1.019 kcal/mol.

S5.1 Deep learning algorithms

A deep neural network is a neural network methods with multi-layers (hidden layer) of neurons between the input and output layers. In each layer, the single neuron gets fully connecting with the neurons in next leyer. It should be preserve the consistency of all labels when applying the model for mutation-induced BFE change predictions. The loss function is constructed as following:

$$\underset{W,b}{\operatorname{argmin}} L(W,b) = \underset{W,b}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{N}\left(y_i - f(x_i;\{W,b\})\right)^2 + \lambda\|W\|^2 \tag{3}$$

where N is the number of samples, f is a function of the feature vector x_i parameterized by a weight vector W and bias term b, and λ represents a penalty constant.

S5.2 Optimization

The backpropagation is applied to evaluated the loss function start from the output layer and propagates backward through the network structure to update the weight vector W and bias term b. According to that the gradient calculation is required, we apply the stochastic gradient descent method with momentum which only evaluates a small part of training data and can be considered as calculating exponentially weighted averages, which is given as

$$\begin{aligned} V_i &= \beta V_{i-1} + \eta \nabla_{W_i} L(W_i, b_i) \\ W_{i+1} &= W_i - V_i, \end{aligned} \tag{4}$$

where W_i is the parameters in the network, $L(W_i, b_i)$ is the objective function, η is the learning rate, X and y are the input and target of the training set, and $\beta \in [0, 1]$ is a scalar coefficient for the momentum term. The momentum term involved accelerates the converging speed.

S6 Supplementary validation

In the main content, we briefly summarized validations of our machine learning predictions and experimental data. For large quantitative validations, we compared the BFE change prediction for mutations on S protein RBD to the experimental deep mutational enrichment data on RBD binding to human ACE2 and CTC-445.2 induced by RBD mutations [7,8,14]. To make these validations, we eliminated the experimental deep mutational enrichment data of RBD binding to human ACE2 and CTC-445.2 from the training sets and set them as testing sets, which have 1539 and 1500 samples, respectively. In the validation of RBD and CTC-445.2 complex, there is a very high correlation between the enrichment data and machine learning predictions, as well as the validation of RBD binding to ACE2, with Pearson correlations are 0.69 and 0.70, respectively. The deep mutational enrichment data can give a proportional descriptor of the affinity strength of protein-protein interactions induced by mutations. The machine learning methods, however, gives stable and equalized evaluations, while experimental data might be different dramatically due to conditions and environments.

In addition, we compared our machine learning results with other experimental data, which are escape fraction, pseudovirus infection changes, and IC_{50} fold changes [8]. In the comparison of 35 cases to experimental escape fractions on RBD binding to clinical trial antibodies induced by emerging mutations, our machine learning predictions have a Pearson correlation of 0.80. Especially, those high escaping mutations E484K and E484Q on LY-CoV555, and mutations K417T and K417N on LY-CoV016, are indicated by both our predictions and the experimental data [8]. We also use the pattern comparisons of our prediction to experimental data. Lastly, we collected experimental data from different literature [1,19,27,28]. According to variations from different research groups, they were summarized in increasing/decreasing patterns of emerging variant (including co-mutations) impacts on antibody therapies in clinical trials. In total there are 20 pattern comparisons with an excellent agreement between various experimental data and our predictions, except for a minor discrepancy [8].

References

[1] FACT SHEET FOR HEALTH CARE PROVIDERS EMERGENCY USE AUTHORIZATION (EUA) OF REGEN-COV (fda.gov).

[2] ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W., AND LIPMAN, D. J. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic acids research 25*, 17 (1997), 3389–3402.

[3] BAS, D. C., ROGERS, D. M., AND JENSEN, J. H. Very fast prediction and rationalization of pka values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics 73*, 3 (2008), 765–783.

[4] BROOKS, B. R., BROOKS III, C. L., MACKERELL JR, A. D., NILSSON, L., PETRELLA, R. J., ROUX, B., WON, Y., ARCHONTIS, G., BARTELS, C., AND BORESCH, S. Charmm: the biomolecular simulation program. *Journal of computational chemistry 30*, 10 (2009), 1545–1614.

[5] CASE, D. A., DARDEN, T. A., CHEATHAM, T. E., SIMMERLING, C. L., WANG, J., DUKE, R. E., LUO, R., CROWLEY, M., WALKER, R. C., AND ZHANG, W. Amber 10. Tech. rep., University of California, 2008.

[6] CHEN, D., CHEN, Z., CHEN, C., GENG, W., AND WEI, G.-W. Mibpb: a software package for electrostatic analysis. *Journal of computational chemistry 32*, 4 (2011), 756–770.

[7] CHEN, J., GAO, K., WANG, R., AND WEI, G.-W. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chemical Science 12*, 20 (2021), 6929–6948.

[8] CHEN, J., GAO, K., WANG, R., AND WEI, G.-W. Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies. *Journal of Molecular Biology 433*, 7744 (2021).

[9] CHERIAN, S., POTDAR, V., JADHAV, S., YADAV, P., GUPTA, N., DAS, M., RAKSHIT, P., SINGH, S., ABRAHAM, P., AND PANDA, S. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms 9*, 7 (2021), 1542.

[10] DOLINSKY, T. J., NIELSEN, J. E., MCCAMMON, J. A., AND BAKER, N. A. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research 32*, suppl_2 (2004), W665–W667.

[11] HOFFMANN, M., KLEINE-WEBER, H., SCHROEDER, S., KRÜGER, N., HERRLER, T., ERICHSEN, S., SCHIERGENS, T. S., HERRLER, G., WU, N.-H., AND NITSCHE, A. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell 181*, 2 (2020), 271–280.

[12] LI, C., TIAN, X., JIA, X., WAN, J., LU, L., JIANG, S., LAN, F., LU, Y., WU, Y., AND YING, T. The impact of receptor-binding domain natural mutations on antibody recognition of SARS-CoV-2. *Signal Transduction and Targeted Therapy 6*, 1 (2021), 1–3.

[13] LI, W., SHI, Z., YU, M., REN, W., SMITH, C., EPSTEIN, J. H., WANG, H., CRAMERI, G., HU, Z., AND ZHANG, H. Bats are natural reservoirs of SARS-like coronaviruses. *Science 310*, 5748 (2005), 676–679.

[14] LINSKY, T. W., VERGARA, R., CODINA, N., NELSON, J. W., WALKER, M. J., SU, W., BARNES, C. O., HSIANG, T.-Y., ESSER-NOBIS, K., AND YU, K. De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science 370*, 6521 (2020), 1208–1214.

[15] LIU, B., WANG, B., ZHAO, R., TONG, Y., AND WEI, G.-W. Eses: software for e ulerian solvent excluded surface, 2017.

[16] LIU, C., GINN, H. M., DEJNIRATTISAI, W., SUPASA, P., WANG, B., TUEKPRAKHON, A., NUTALAI, R., ZHOU, D., MENTZER, A. J., AND ZHAO, Y. Reduced neutralization of SARS-CoV-2 B. 1.617 by vaccine and convalescent serum. *Cell* (2021).

[17] NGUYEN, D. D., AND WEI, G.-W. Agl-score: Algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. *Journal of chemical information and modeling 59*, 7 (2019), 3291–3304.

[18] PASCARELLA, S., CICCOZZI, M., BIANCHI, M., BENVENUTO, D., CAUDA, R., AND CASSONE, A. Cutting epitopes to survive: the case of lambda variant. *bioRxiv* (2021).

[19] PLANAS, D., VEYER, D., BAIDALIUK, A., STAROPOLI, I., GUIVEL-BENHASSINE, F., RAJAH, M. M., PLANCHAIS, C., PORROT, F., ROBILLARD, N., AND PUECH, J. Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization. *Nature* (2021), 1–7.

[20] QU, X.-X., HAO, P., SONG, X.-J., JIANG, S.-M., LIU, Y.-X., WANG, P.-G., RAO, X., SONG, H.-D., WANG, S.-Y., AND ZUO, Y. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *Journal of Biological Chemistry 280*, 33 (2005), 29588–29595.

[21] SHU, Y., AND MCCAULEY, J. GISAID: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance 22*, 13 (2017), 30494.

[22] SONG, H.-D., TU, C.-C., ZHANG, G.-W., WANG, S.-Y., ZHENG, K., LEI, L.-C., CHEN, Q.-X., GAO, Y.-W., ZHOU, H.-Q., AND XIANG, H. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proceedings of the National Academy of Sciences 102*, 7 (2005), 2430–2435.

[23] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I., AND SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research 15*, 1 (2014), 1929–1958.

[24] TAI, W., HE, L., ZHANG, X., PU, J., VORONIN, D., JIANG, S., ZHOU, Y., AND DU, L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & molecular immunology 17*, 6 (2020), 613–620.

[25] WALLS, A. C., PARK, Y.-J., TORTORICI, M. A., WALL, A., MCGUIRE, A. T., AND VEESLER, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* (2020).

[26] WANG, M., CANG, Z., AND WEI, G.-W. A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation. *Nature Machine Intelligence 2*, 2 (2020), 116–123.

[27] WANG, P., NAIR, M. S., LIU, L., IKETANI, S., LUO, Y., GUO, Y., WANG, M., YU, J., ZHANG, B., AND KWONG, P. D. Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1. 7. *Nature 10* (2021).

[28] WEISBLUM, Y., SCHMIDT, F., ZHANG, F., DASILVA, J., POSTON, D., LORENZI, J. C., MUECKSCH, F., RUTKOWSKA, M., HOFFMANN, H.-H., AND MICHAILIDIS, E. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife 9* (2020), e61312.

[29] YANG, Y., HEFFERNAN, R., PALIWAL, K., LYONS, J., DEHZANGI, A., SHARMA, A., WANG, J., SATTAR, A., AND ZHOU, Y. Spider2: A package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. In *Prediction of protein secondary structure*. Springer, 2017, pp. 55–63.