

Mathematical artificial intelligence design of mutation-proof COVID-19 monoclonal antibodies*

JIAHUI CHEN[†] AND GUO-WEI WEI^{†‡}

Emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants have compromised existing vaccines and posed a grand challenge to coronavirus disease 2019 (COVID-19) prevention, control, and global economic recovery. For COVID-19 patients, one of the most effective COVID-19 medications is monoclonal antibody (mAb) therapies. The United States Food and Drug Administration (U.S. FDA) has given the emergency use authorization (EUA) to a few mAbs, including those from Regeneron, Eli Lilly, etc. However, they are also undermined by SARS-CoV-2 mutations. It is imperative to develop effective mutation-proof mAbs for treating COVID-19 patients infected by all emerging variants and/or the original SARS-CoV-2. We carry out a deep mutational scanning to present the blueprint of such mAbs using algebraic topology and artificial intelligence (AI). To reduce the risk of clinical trial-related failure, we select five mAbs either with FDA EUA or in clinical trials as our starting point. We demonstrate that topological AI-designed mAbs are effective for variants of concerns and variants of interest designated by the World Health Organization (WHO), as well as the original SARS-CoV-2. Our topological AI methodologies have been validated by tens of thousands of deep mutational data and their predictions have been confirmed by results from tens of experimental laboratories and population-level statistics of genome isolates from hundreds of thousands of patients.

*Research supported in part by NIH grants GM126189 and AI164266, NSF grants DMS-2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, Michigan Economic Development Corporation, MSU Foundation, Bristol-Myers Squibb 65109, and Pfizer.

[†]The authors thank The IBM TJ Watson Research Center, The COVID-19 High Performance Computing Consortium, NVIDIA, and MSU HPCC for computational assistance. The authors thank Drs. Kaifu Gao and Changchuan Yin, and Ms. Rui Wang for useful discussions.

[‡]Corresponding author.

1. Introduction

In combating the coronavirus disease 2019 (COVID-19) pandemic, there has been exigency to develop effective antiviral treatments i.e., vaccines, antiviral drugs, and antibody therapies. The developments in these treatments are some of the most paramount scientific accomplishments in the battle against COVID-19. However, emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants, particularly variants of concern (VOCs), impact transmission, virulence, and immunity and pose a threat to existing vaccines and antibody drugs.

SARS-CoV-2 is an enveloped, unsegmented positive-sense single-strand ribonucleic acid (RNA) virus, which enters cells depending on the binding of its spike (S) protein receptor-binding domain (RBD) to host angiotensin-converting enzyme 2 (ACE2) receptor [1]. The binding free energy (BFE) between the S protein and ACE2, according to epidemiological and biochemical analysis, is proportional to the infectivity of SARS-CoV-2 in the host cells [2, 3]. In July 2020, it was shown that driven by natural selection [4], mutations strengthen RBD-ACE2 binding and thus make the virus more infectious. The high-frequency RBD mutations were shown to be undoubtedly governed by natural selection [4, 5]. Additionally, natural selection also creates new SARS-CoV-2 variants easily escaping antibodies induced by either infection or vaccination [6]. By comparing to the first SARS-CoV-2 strain deposited to GenBank (Access number: NC 045512.2), the mutation-induced BFE changes ($\Delta\Delta G$) of the binding of S protein and ACE2 provide a way to measure the infectivity changes of a SARS-CoV-2 variant. Positive BFE changes induced by mutations of RBD binding to ACE2 reveal that mutations potentially improve the binding, while negative BFE changes indicate mutations weaken the transmissibility and infectivity. Thus, the impact of SARS-CoV-2 RBD variants on infectivity can be evaluated according to their BFE changes [7, 8, 4, 9].

Currently, except for antiviral drugs which are proved more efficacious than placebo such as Pfizer's Paxlovid (nirmatrelvir), COVID-19 vaccines are considered as the game-changer and SARS-CoV-2 monoclonal antibody (mAb) therapies are shown to reduce the risk of disease progression. Both approaches rely on antibodies in different mechanisms. Specifically, vaccines are designed to stimulate an effective host immune response triggering the host adaptive immune system to produce antibodies against future infection [10], while antibody therapies are obtained from patients convalescing from COVID-19 or other diseases, which block viral entry by binding to

the viral S protein. Various vaccines, including two mRNA vaccines designed by Pfizer-BioNTech and Moderna, have been granted authorization for emergency use as well as antibody therapies (such as casirivimab [11], imdevimab [11], bamlanivimab [12], etesevimab [13], regdanvimab [14], et al.) in many countries. However, RBD mutations simultaneously strengthen SARS-CoV-2 infectious [4], escape existing vaccines [6], and attenuate antibodies [15].

Genetic mutations of SARS-CoV-2 provide a mechanism for viruses to adapt to and evade host immune responses, COVID-19 vaccines, and antibody therapies. Although SARS-CoV-2 has higher fidelity and a slower evolutionary rate than other RNA viruses [16], over 5,000 unique mutations were found on SARS-CoV-2 S protein [5, 8]. This situation awakes the question of the impacts of existing mutations on vaccines and antibodies. According to the WHO tracking SARS-CoV-2 variants [17], variants are characterized as Variants of Interest (VOIs) and Variants of Concern (VOCs) and prioritized global monitoring and research. Other variants of local interest/concern are designated by national authorities. There are more than ten designated VOCs, including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), etc. It is interesting to note that RBD residues 452 and 501 were predicted to “have high changes to mutate into significantly more infectious COVID-19 strains” in early 2020 [4]. As predicted, variants Alpha, Beta, Gamma, Delta, Kappa, Theta, Lambda, Mu, etc. all have at least one of these two mutations.

Evidence shows VOCs have high transmissibility and dominate the spreading of SARS-CoV-2 on multiple countries [15, 18, 19, 19] (see Fig. 4a). Studies show VOCs are resistant to antibody neutralization. For example, Alpha and Beta variants are reported as antibody resistance to neutralization by some anti-N-terminal domain (NTD) and anti-BRD mAbs, including casirivimab and bamlanivimab for Beta variants [15]. Gamma variant is also shown refractory to neutralization by some mAbs, including emergency use authorization (EUA) antibody therapies casirivimab, imdevimab, and etesevimab [20, 21], and similar results are for Delta variant as well [19]. Additionally, according to WHO [17], VOIs including Eta, Iota, Kappa, and Lambda variants have genetic changes impacting virus characteristics of transmissibility, disease severity, immune escape, and diagnostic escape and lead to significant community transmission. VOIs may share some mutations with VOCs on RBD. Thus, single mutation experiments on L452R, S477N, and E484K, can be used to analyze their effects on antibody neutralization [22, 21, 23]. For example, the mutation L452R

on the S protein RBD increases 20% of the transmissibility of SARS-CoV-2 [23], and has mild negative impacts on the neutralization by EUA antibody therapies according to Food and Drug Administration (FDA) [24, 25].

Experimental studies of mutational impacts on the existing antibodies and antibody drugs are time-consuming and are limited to a small fraction of known viral mutations. It is difficult to accurately determine whether a mutation will evade a vaccine in general populations of various races, genders, ages, and existing health conditions. Based on the molecular mechanism of host cells infected by SARS-CoV-2 virus and immune system responses, quantitative assessment of mutational impacts on SARS-CoV-2 infectivity and antibody drugs can be achieved by computing BFE changes following mutations of the RBD-ACE2 complex and RBD-antibody complexes. In our earlier work, we applied a topology-based deep learning model to predict the binding free energy (BFE) changes of the RBD-ACE2 complex and 106 RBD-antibody complexes induced by RBD mutations [4, 8, 5, 9]. These predictions were validated by experimental results [26, 27, 28, 29, 30]. For example, our predictions of mutation-induced BFE changes on CTC-445.2 binding to RBD were shown to be highly correlated with the experimental data [28, 8]. In recent work, we validated our predictions of BFE changes on the RBD-ACE2 complex with deep mutational scanning data, achieving the Pearson correlation of 70% [28, 8]. Moreover, in a comparison with experimental data, the predicted BFE changes have an 80% correlation with the escape fraction [8]. A high prediction accuracy with experimental data was found in predicting emerging variant impacts on clinical trial antibodies [8].

The objective of this work is to introduce a mathematical artificial intelligence (AI)-based computational strategy for the rational design of mutation-proof mAbs. As examples, we consider high-frequency RBD mutations on 5 mAb therapies, namely casirivimab, imdevimab, etesevimab, bamlanivimab, and regdanvimab. Among them, casirivimab and imdevimab are authorized for the treatment of COVID-19 by the U.S. Food and Drug Administration (FDA). Etesevimab and bamlanivimab also obtained FDA's emergency use authorization (EUA). Regdanvimab is issued advice on use for treating COVID-19 by European Medicines Agency (EMA). We use our intensively-validated algebraic topology-based deep learning model to estimate the mutation-induced BFE changes of antibody-RBD complexes. This study also offers an important strategy for the design of mutation-proof mAbs for other viruses.

Table 1: The statistics of BFE changes ($\Delta\Delta G$) induced by AI-based deep mutations on antibody variable domains. The number of AI-based deep mutations on each chain is denoted as “Total No”. Three categories of the numbers of topological AI-based mutations are given to BFE changes greater than 0 kcal/mol, 0.5 kcal/mol, and 1 kcal/mol, respectively

Antibody	Chain	Total No	$\Delta\Delta G > 0$ kcal/mol		$\Delta\Delta G > 0.5$ kcal/mol		$\Delta\Delta G > 1$ kcal/mol	
			No	Ratio (%)	No	Ratio (%)	No	Ratio (%)
REGN10933	Heavy	2223	742	33.38	46	2.07	19	0.85
	Light	1995	858	43.01	11	0.55	1	0.05
REGN10987	Heavy	2223	675	30.36	24	1.08	11	0.49
	Light	1995	734	36.79	7	0.35	1	0.05
LY-CoV016	Heavy	2242	220	9.81	8	0.36	2	0.09
	Light	2090	168	8.04	2	0.10	1	0.05
LY-CoV555	Heavy	2337	480	20.54	35	1.50	5	0.21
	Light	2014	518	25.72	11	0.55	3	0.15
CT-P59	Heavy	2394	514	21.47	18	0.75	8	0.33
	Light	2090	542	25.93	9	0.43	0	0.00
Average		2160	545	25.51	17	0.77	5	0.23

2. Results

2.1. AI-based deep mutational screening of five RBD-binding antibodies

We first carry out a topological AI-based deep mutational scanning on the antibody variable domains that bind to the RBD for five mAbs. These mutations are conducted systematically such that each residue in each mAb’s light and heavy chains are mutated to all 19 other possible amino acids. Then, the BFE change for the antibody-RBD complex induced by each mutation is computed by the topological AI model. Most mutations on the antibody variable domain tend to have negative BFE changes or mild positive BFE changes (see the Appendix), indicating that mAbs have been optimized for their RBD binding. Table 1 shows the statistical results for five mAbs involving about 21,600 AI-based deep mutations on antibody variable domains. An average of 25.51% mutations cause the strengthening of antibody-RBD binding (or having positive BFE changes). In fact, only 0.77% and 0.23% mutations have BFE changes greater than 0.5 kcal/mol and 1 kcal/mol, respectively. The dramatic decrease in the number of mutations having BFE changes greater than 0.5 kcal/mol indicates that these antibodies have a small number of residue sites for improving the mAb neutralization effect against SARS-CoV-2. Among the five antibodies, LY-CoV016 has the least number of antibody mutations for strengthening its binding with RBD, while

REGN10933 has a relatively large number of residues that can be improved. The heap map of complete virtual mutational scans on the antibody variable domains is provided in the Appendix.

In Figure 4c, the residues with at least one mutation having BFE changes greater than 1 kcal/mol are presented according to Table 1. For REGN10933, two residues A75 and T102 on the heavy chain have four mutations (A75Y/W/F/M) and seven mutations (T102D/E/Q/W/I/L/V) with BFE changes greater than 1 kcal/mol. For the heavy chain of REGN10987, A33 has eight candidates (A33K/D/E/Q/T/I/L/M) for strengthening the binding of REGN10987 and RBD. For the rest of the selected residues, none of them have more than three effective mutants. These small numbers of candidates also indicate that these antibody therapies were optimized. However, their optimizations were with respect to the original SARS-CoV-2 virus and these mAbs are prone to emerging RBD mutations.

2.2. AI-based rational design of mutation-proof antibodies

SARS-CoV-2 variants have been evolving to increase their capability to evade vaccine and antibody protections [6]. With the threat of emerging SARS-CoV-2 variants, it is important to design mutation-proof antibody therapies. Our essential idea is to systematically mutate each residue of an antibody into 19 possible other amino acids to search for mutation-proof new designs of antibodies. Variants Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Lambda (C.37), Epsilon (B.1.427), and Kappa (B.1.427) encode spike proteins with mutations K417N/T, L452R/Q, T478K, E484K/Q, F490S, and N501Y in the spike protein RBD that provide a degree of resistance to neutralization by our previous modeling prediction [9] and experimental analysis [31, 32, 33, 34, 35, 36, 37] (see Fig. 4b). In addition to WHO designated variants, the 10 most observed RBD mutations in terms of their frequencies are more infectious and increase the virus transmissibility [9], which include seven mutations appearing in the WHO designated variants plus S477N, N439K, and S494P. Mutation S477N, N439K, and S494K rank 5th, 7th, and 9th in terms of frequencies. Mutations L452Q and E484Q of Lambda and Kappa variants, respectively, where E484Q ranks 11th, are not in the top ten observed RBD mutations [5]. Thus, we focus on these twelve mutations for the antibody redesigning and provide the 100 most observed RBD mutation results in the Appendix.

2.2.1. REGN10933 and REGN10987 As shown in Figures 1a and 1d, the analysis of antibodies REGN10933 and REGN10987 are given for the

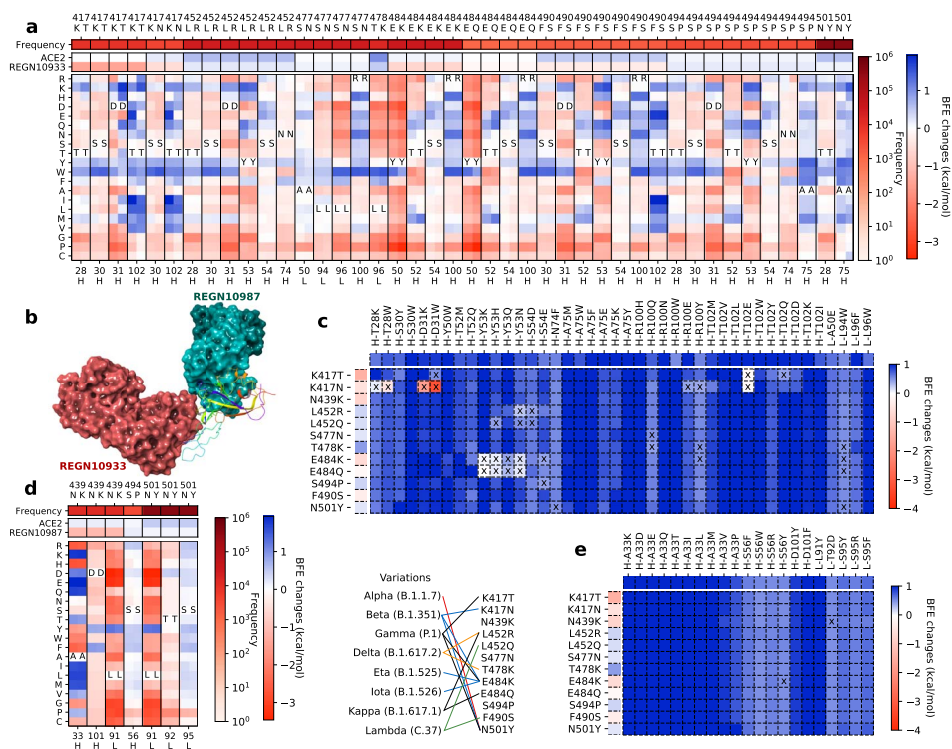


Figure 1: The deep mutational analysis on antibodies REGN10933 and REGN10987. **a** The mutational scanning on antibody REGN10933 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of REGN10933 and L indicates the light chain. BFE change range is from -3.46kcal/mol to 1.94kcal/mol. **b** 3D structure of the complex (PDB: 6XDG) [11]. **c** Illustration of BFE changes of effective antibody mutations. The first column indicates the BFE changes induced by RBD mutations of the binding between RBD and antibody. The first row indicates the BFE changes induced by antibody mutations of the binding between RBD and antibody. **d** The mutation scanning on antibody REGN10987 binding to S protein RBD and mutated RBD. **e** Illustration of BFE changes of effective antibody mutations.

deep mutational scanning on antibody variable domains that bind to the original S protein RBD and mutated RBD of variants. The mutations on antibodies are considered if the distances between C α s of antibody residues and RBD residues are less than 15 Å and selected when antibody mutations

have positive BFE changes greater than 0.5 kcal/mol both for binding to the original RBD and the RBD of variants. Figure 1a shows ten mutations on the S protein RBD with effective mutations on antibody REGN10933. For unselected RBD mutations N439K and L452Q, the deep mutational scanning on antibody variable domains within 15 Å to the RBD shows no BFE changes greater than 0.5 kcal/mol. The first row of Figure 1a gives the frequency information of each RBD mutation. The following two rows give the BFE changes following the RBD mutations of the binding between S protein RBD and ACE2 or between the RBD and antibodies, where RBD mutations are more favorable of the binding to ACE2 than to antibody REGN10933. The rest rows demonstrate the deep mutational scanning on REGN10933 binding to S protein RBD on odd columns and RBD mutations on even columns. For a pair of the RBD and its mutation, there are multiple residues on the antibody REGN10933 having mutations that increase the binding affinity for both. Notice that not all mutations on residues have positive BFE changes, and in total, there are 42 candidates on antibody REGN10933.

Once the 42 candidates of antibody REGN10933 are selected, their BFE changes of the binding to the RBD with the 12 mutations induced by antibody mutations are displayed in Figure 1c. The cross marks indicate that the BFE changes are less than 0.5 kcal/mol. Meanwhile, the first column of the heatmap gives the BFE changes of the RBD binding to antibody induced by RBD mutations, and the first row of the heatmap gives the BFE changes of the binding complex induced by antibody mutations. Here, 17 of 42 mutations on REGN10933 have one or more BFE changes less than 0.5 kcal/mol. Especially, D31W on the heavy chain of REGN10933 causes a negative BFE change of -3.16 kcal/mol. Y53N on the heavy chain and L94W on the light chain induce four BFE changes less than 0.5 kcal/mol. For antibody REGN10933, mutations H-S30Y/W, H-Y50W, H-T52Q, H-A75M/W/F/E/K/Y, H-R100H/N/W, H-T102M/V/L/W/Y/D/K/I, L-A50E, and L-L96F/W can be the effective candidates for improving the neutralization of antibody REGN10933 against the S protein RBD and its variants.

For antibody REGN10987, there are 21 candidates on the variable domain and 10 candidates are on the heavy chain A33 in Figure 1e. Considering REGN10987 does not directly connect S protein RBD to the receptor the binding motif of ACE2, three RBD mutations are selected with distances less than 15 Å to REGN10987 (see Fig. 1b) and having BFE changes greater than 0.5 kcal/mol for both cases (see Fig. 1d). Except for the antibody mutations with BFE changes less than 0.5 kcal/mol, there are 19

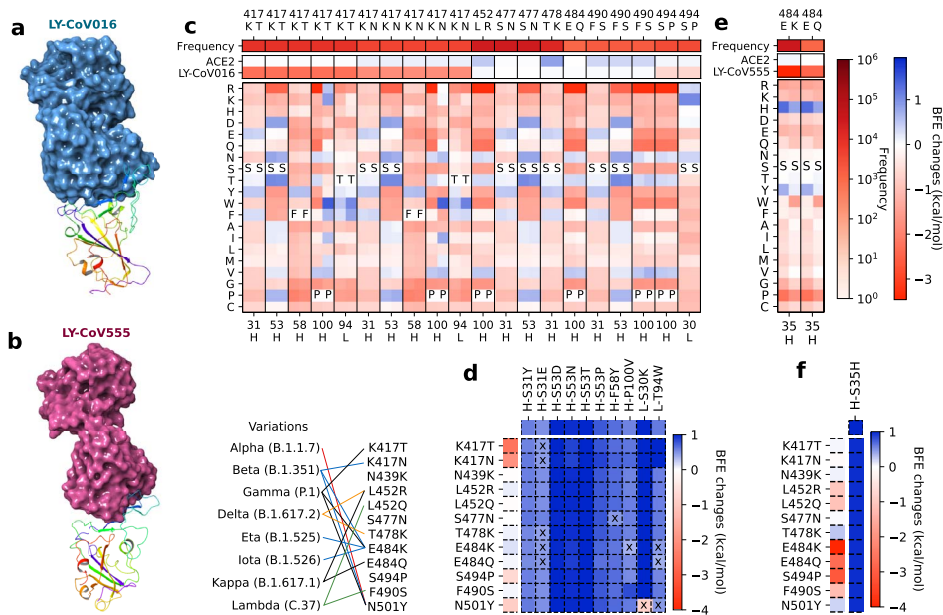


Figure 2: The deep mutational analysis on antibodies LY-CoV016 and LY-CoV555. **a** 3D structure of LY-CoV016 and RBD (PDB: 7C01) [13]. **b** 3D structure of LY-CoV555 and RBD (PDB: 7KMG) [12]. **c** The deep mutational scanning on antibody LY-CoV016 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of LY-CoV016 and L indicates the light chain. BFE change range is from -3.46kcal/mol to 1.94kcal/mol. **d** Illustration of BFE changes of effective antibody mutations. The first column indicates the BFE changes induced by RBD mutations of the binding between RBD and antibody. The first row indicates the BFE changes induced by antibody mutations of the binding between RBD and antibody. **e** The deep mutational scanning on antibody LY-CoV555 binding to S protein RBD and mutated RBD. **f** Illustration of BFE changes of effective antibody LY-CoV55 mutations.

mutations H-A33K/D/E/Q/T/I/L/M/V/P, H-S56F/W/R/Y, H-D101Y/F, L-L91Y, L-T92D, and L-S95Y/R/F improve the neutralizing efficacy against SARS-CoV-2 and variants.

2.2.2. LY-CoV016 and LY-CoV555 With a similar analysis on antibodies LY-CoV016 (see Fig. 2a) and LY-CoV555 (see Fig. 2b), we collected

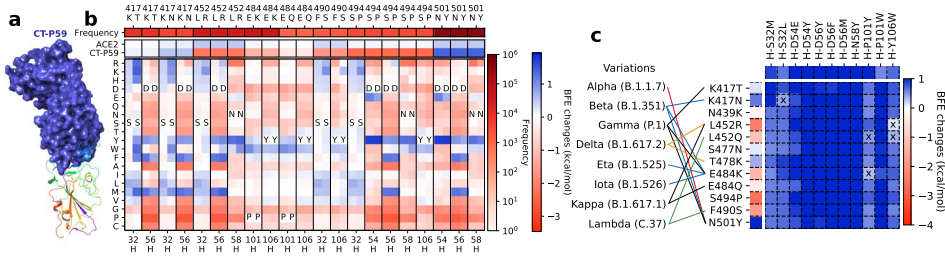


Figure 3: The deep mutational analysis on antibody CT-P59. **a** 3D structure of CT-P59 and RBD (PDB: 7CM4) [14]. **b** The mutational scanning on antibody CT-P59 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of CT-P59. BFE change range is from -3.46kcal/mol to 1.94kcal/mol. **c** Illustration of BFE changes of effective antibody mutations. The first column indicates the BFE changes induced by RBD mutations of the binding between RBD and antibody. The first row indicates the BFE changes induced by antibody mutations of the binding between RBD and antibody.

10 candidates and one candidate for LY-CoV016 and LY-CoV555, respectively, in Figure 2. In Figure 2c, seven RBD mutations are evaluated on 6 antibody residues H-S31, H-S53, H-F58, H-P100, L-S30, and L-T94. Interestingly, half of the residues are serine, which has a small polar uncharged side chain. From the second and third columns of Figure 2c, it is noticed that RBD mutations are more favorable to the RBD binding to ACE2 than to LY-CoV016. Eliminating the mutations with BFE changes less than 0.5 kcal/mol, there are five effective mutations H-S31Y and H-S53D/N/T/P for LY-CoV016 for improving its competitiveness (see Fig. 2d). As for LY-CoV555, only one candidate H-S35H is selected as shown in Figures 2e and f.

2.2.3. CT-P59 Finally, we analyze antibody CT-P59 with 11 effective mutations on 6 residues on the heavy chain, which are H-S32M/L, H-D54E/Y, H-D56Y/F/M, H-N58Y, H-P101Y/W, and H-Y106W (see Fig. 3). There are seven RBD mutations K417T/N, L452R, E484K/Q, F490S, S494P, and N501Y. For example, mutation L452R, a mutation of the Delta variant, is favorable to the neutralization of binding to ACE2, but disrupts the neutralization of the binding to CT-P59. Six candidates H-S32M/L, H-N56Y/F/M, and H-N58Y on CT-P59 can counteract the disrupting effect by mutation L452R. In addition, the RBD mutation T478K, another mutation of the Delta variant, has a mild positive BFE change of its binding to CT-P59 and

positive BFE changes in its binding to CT-P59 with mutations as shown in Figure 3c. Although the Delta variant reduces susceptibility against CT-P59, the selected candidates H-S32M/L, H-D54E/Y, H-D56Y/F/M, H-N58Y, and H-P101W can improve the neutralizing efficacy of CT-P59 against the Delta variant. Notably, our early predictions of BFE changes induced by mutations L452R and T478K binding to CT-P59 [9] are proved by later experimental results on the neutralization of CT-P59 binding to the single mutants of the RBD [38]. Overall, mutations H-S32M/L, H-D54E/Y, H-D56Y/F/M, H-N58Y, and H-P101W on CT-P59 can improve the neutralization ability of CT-P59 binding to the S protein RBD.

3. Discussion

Emerging variants have dominated the spreading of SARS-CoV-2 worldwide and have been shown to reduce the neutralization efficacy of antibodies and degrade the protection of SARS-CoV-2 vaccine and antibody treatments. Especially, more attention should be paid to RBD mutations as the S protein RBD and ACE2 binding is the key to SARS-CoV-2 virus-host cell entry. The SARS-CoV-2 RBD mutations can strengthen the RBD-ACE2 binding to make the virus more infectious and meanwhile, weaken the RBD-antibody binding to breakthrough vaccines and mAbs. Consequently, the efficacy of vaccines and antibody therapies are compromised and viral transmissibility is enhanced. Twelve mutations on RBD are observed from variants of concern (VOCs) and variants of interest (VOIs) (see Fig. 4). Interestingly, according to our prediction on BFE changes induced by RBD mutations for RBD binding to human ACE2, all these VOCs have at least one mutation on RBD has a BFE change greater than 0.5 kcal/mol. For instance, Alpha, Beta, and Gamma variants have the mutation N501Y with a BFE change 0.55 kcal/mol, and the Delta variant has mutations L452R and T478K with BFE changes 0.58 kcal/mol and 1.00 kcal/mol, respectively. The BFE strengthening mutations on the Delta variant RBD enhance the infectivity of the Delta variant, creating a dominant strain. For VOIs, none of them have a mutation on RBD with BFE changes greater than 0.5 kcal/mol. The highest BFE change is induced by L452Q from the Lambda variant which is 0.44 kcal/mol. Thereafter, the hypothesis is that emerging SARS-CoV-2 variants have at least one mutation with BFE changes greater than 0.5 kcal/mol. Based on our previous findings in [4], 606 out of 1149 RBD mutations that we predicted as “most likely” mutations have been observed, while the rest mutations 1912 “likely” and 625 “unlikely” mutations are rarely found on the S protein RBD. In Figure 4, we list 61 most likely mutations on RBD whose

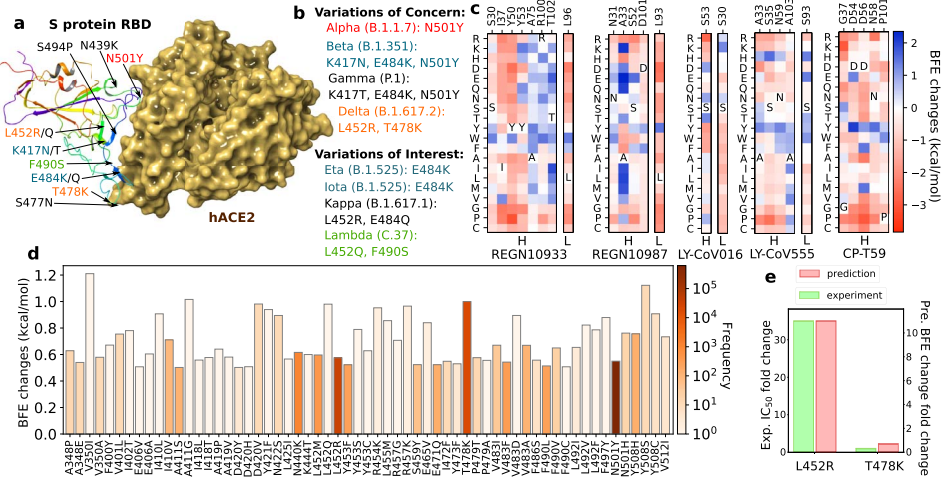


Figure 4: **a** 3D structure of host ACE2 (ACE2) and RBD (PDB: 6M0J) [39]. **b** The lists of variants of concerns (VOCs) and variants of interest (VOIs). **c** Residues on antibodies selected with at least one mutation having BFE changes greater than 1 kcal/mol. BFE change range is from -3.84kcal/mol to 2.35kcal/mol. **d** Illustration of SARS-CoV-2 mutation-induced BFE changes for the complexes of S protein and ACE2. Here, mutations with BFE changes greater than 0.5 kcal/mol are presented. **e** Comparison of experimental fold change and predicted fold change in IC₅₀ of CT-P59 and S protein RBD complex induced by mutations L452R and T478K.

BFE changes are greater than 0.5 kcal/mol, 38 of 61 mutations have been observed and 17 mutations V350I, I410L, A411G, D420V, Y421F, N422S, L452Q, R454K, L455M, R457K, E465V, T478K, V483D, L492V, F497Y, Y508S/C have BFE changes from 0.82 kcal/mol to 1.21 kcal/mol, which could be effective mutations for VOCs. Note that a high BFE change of the binding between S protein RBD and ACE2 indicates the strengthening of SARS-CoV-2 infectivity. Potentially, this mutation could be a vaccine escape mutation if it weakens the binding of RBD to antibodies.

4. Validation

The validation of our topological AI model predictions for mutation-induced BFE changes has been demonstrated by comparison with experimental data in recent publications [8, 9]. Firstly, we showed high correlations of experimental deep mutation enrichment data and predictions for SARS-CoV-2 S

protein RBD and CTC-445.2 complex [8] and SARS-CoV-2 RBD and ACE2 complex [9]. In the comparison with experimental data on clinical trial antibody therapies for high-frequency mutations, our predictions achieve a Pearson correlation of 0.80 [9]. Considering the BFE changes induced by mutations on the RBD of the ACE2-RBD complex, predictions on mutations L452R and N501Y have a highly similar trend with experimental data [9]. Meanwhile, as we presented early [5], high-frequency mutations are associated with positive BFE changes. Moreover, for multi-mutation tests, our BFE change predictions have the same pattern as experimental data on the impact of SARS-CoV-2 variants on major antibody therapeutic candidates [9].

Recent studies on the potency of CT-P59 *in vitro* and *in vivo* against Delta variants [38] show that the neutralization of CT-P59 is reduced by the effects of L452R (13.22 ng/mL) and is retained against T478K (0.213 ng/mL). In our predictions [9], L452R induces a negative BFE change (-2.39 kcal/mol) and T478K induces a positive BFE change (0.36 kcal/mol). In Figure 4e, the fold changes are presented for experimental and prediction values.

Further validation on the Alpha variant RBD mutation was discussed elsewhere [40]. Our predictions of Omicron BA.1 and BA.2 infectivity, vaccine breakthrough, and antibody resistance, which was made when there were no experimental results available, were later nearly perfectly confirmed by experimental data [41, 42].

5. Methods

The development of our deep learning model for BFE change predictions on protein-protein interactions for SARS-CoV-2 problems can be summarized in four steps. First, preparing genome sequence data from the GISAID database [43] (<https://www.gisaid.org/>). By taking the first complete SARS-CoV-2 genome from the GenBank (NC_045512.2) as the referencing [44], a set of single nucleotide polymorphism (SNP) profiles is generated, i.e., residues 329 to 530 on the S protein RBD have 606 non-degenerate mutations are found. Then, 100 most observed mutations have been collected with frequency more than 40 times. Next, collecting SARS-CoV-2 data and related data is the key step, which makes the model reliable and accurate. Massive data of BFE changes of SARS-CoV-2 are rarely reported, while the enrichment ratio data via high-throughput deep mutations are relatively easy to obtain. With the fundamental dataset of BFE changes upon mutations the SKEMPI 2.0 dataset [45], deep mutational enrichment ratio

data is added as another database for our machine learning training [9]. After the database preparation, the third step is the feature generations of protein-protein interaction complexes. We implemented the element-specific algebraic topological analysis on point cloud samples consisting of complex atoms [46, 9, 47]. This topological approach is based on persistent homology [48, 49], a powerful method for protein structure representation [47, 50] and drug discovery [51]. Additionally, biophysics and biochemistry features such as surface areas, partial charges, Coulomb interactions, et al., are combined with topological features [8]. Lastly, deep neural networks are constructed for the BFE change prediction of protein-protein interactions involving mutations [9].

In the third step, obtaining the mutant protein structure requires using Scap utility from Jackal software package [52], which replaces the side chain of the mutation site with min option being set to 4 with additional conformers obtained by perturbing conformers in a rotamer library. The mutant protein structures of RBD variants are constructed by Scap and, then, are used as primal structures for the calculation of antibody mutation impact on RBD variants.

The detailed descriptions of datasets and machine learning model are given in literature [4, 53, 9] and are available at [TopNetmAb](#). In addition, the SARS-CoV-2 single nucleotide polymorphism data in the world is available at [Mutation Tracker](#). The analysis of RBD mutations is available at [Mutation Analyzer](#).

6. Conclusion

Driven by natural selection [4], severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been evolving towards increasingly more infectious, vaccine escape, and antibody resistance [6]. Interestingly, this evolution can be achieved through mutations at the viral spike protein receptor-binding domain (RBD), which binds to the human angiotensin-converting enzyme 2 (ACE2) to facilitate the viral cell entry. Meanwhile, the RBD is also a target of most monoclonal antibodies (mAbs) for direct neutralization of the virus. As a result, natural selection-driven virus evolution gives rise to variants of concerns (VOCs), such as Variants Alpha, Beta, Gamma, Delta, etc. VOCs fuel the waves of widespread infections, evade vaccines, and attenuate the efficacy of existing mAbs. This work provides a mathematical artificial intelligence (AI)-based strategy to design mutation-proof antibodies.

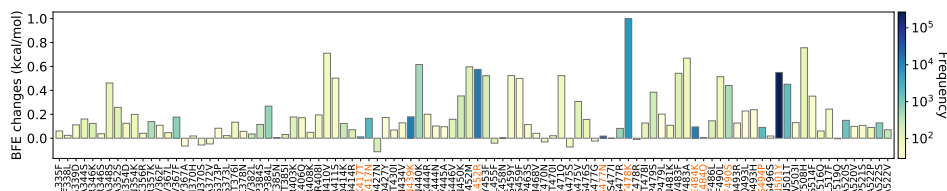


Figure 5: BFE changes of the 100 most observed RBD mutations with their frequency. The variants' mutations are colored in orange.

Our mathematical AI model utilizes persistent homology and deep learning and was trained with tens of thousands of experimental data, including SARS-CoV-2 related deep mutational data. We carry out an AI-based deep mutational screen of five existing mAbs, including those approved by the U.S. Food and Drug Administration (FDA) for emergency use authorization (EUA). Our deep mutational screen indicates that most mAbs have been optimized against the original SARS-CoV-2 but are prone to the RBD mutations. By considering high-frequency RBD mutations, including those from VOCs, we systematically mutate each residue of the five selected mAbs to 19 possible variants to search for potentially mutation-proof new mAbs. Our study offers many alternative designs of mutation-proof mAbs.

Appendix

The 100 most observed RBD mutations are collected with their BFE changes and frequency correspondingly. In Figure 5, variants' mutations are colored in orange. Mutations L452R, T478K, and N501Y are predicted with high BFE changes.

Second, this appendix provides the full results of the mutational scanning on antibodies REGN10933 (see Figure 6), REGN10987 (see Figure 7), LY-CoV016 (see Figure 8), LY-CoV555 (see Figure 9), and CT-P59 (see Figure 10) binding to S protein RBD and mutated RBD. Note that some of these results do not show a good option in designing mutation-proof antibodies. For example, Figure 6 shows that the deep mutational scanning is on antibody REGN10933 of the binding to S protein RBD and mutated RBD. The mutations on S protein RBD are selected from the twelve-selected mutations. As a global observation, an antibody mutation from others to residue tryptophan (denoted as W) favors the binding of S protein RBD and mutated RBD than the others. This provides potential suggestions on antibody design that increasing tryptophan populations on the binding interface will enhance the binding affinity of REGN10933 to S protein RBD.

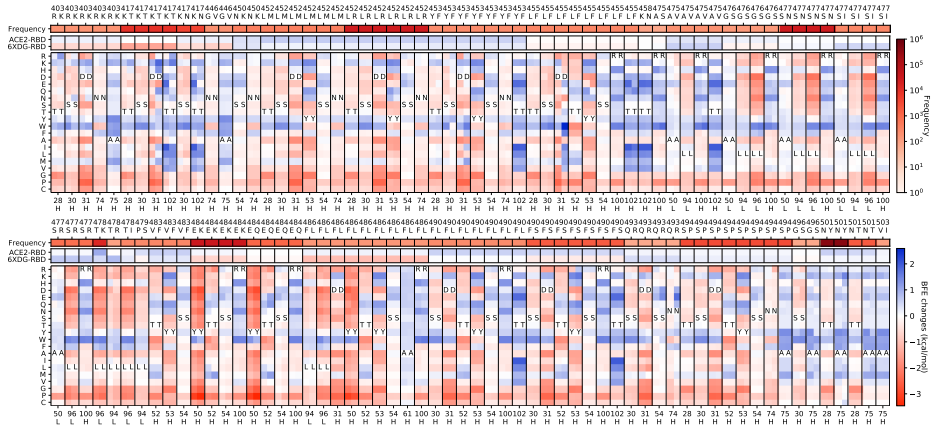


Figure 6: The mutational scanning on antibody REGN10933 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of REGN10933. L indicates the light chain of RENG10933. BFE change range is from -3.28kcal/mol to 2.41kcal/mol.

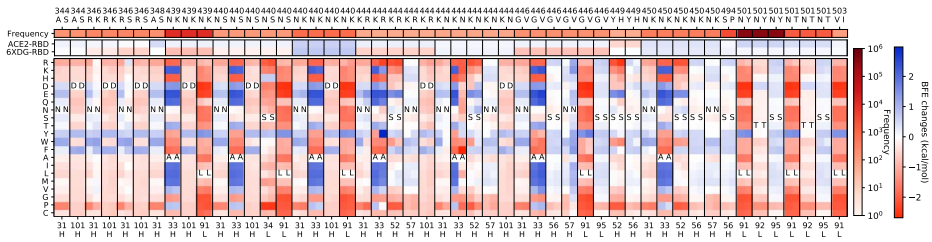


Figure 7: The mutational scanning on antibody REGN10987 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of REGN10987. L indicates the light chain of RENG10933. BFE change range is from -2.47kcal/mol to 2.98kcal/mol.

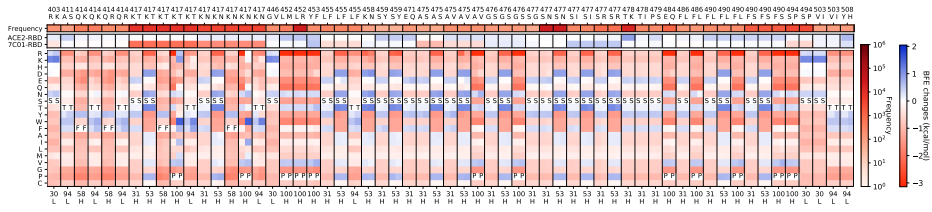


Figure 8: The mutational scanning on antibody LY-CoV016 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of LY-CoV016. L indicates the light chain of LY-CoV016. BFE change range is from -3.08kcal/mol to 2.03kcal/mol.

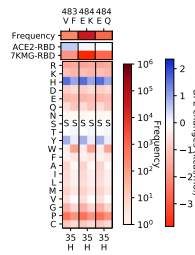


Figure 9: The mutational scanning on antibody LY-CoV555 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of REGN10987. L indicates the light chain of LY-CoV555. BFE change range is from -3.79kcal/mol to 2.23kcal/mol.

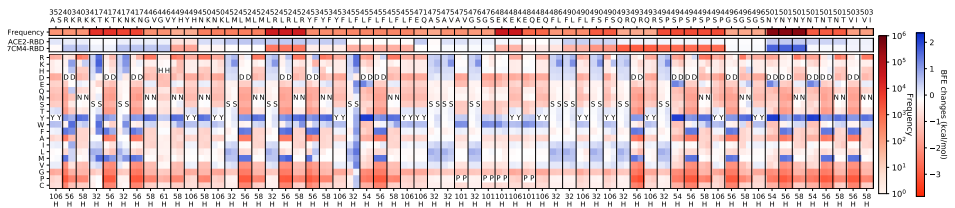


Figure 10: The mutational scanning on antibody CT-P59 binding to S protein RBD and mutated RBD. In the bottom column labels, H indicates the heavy chain of CT-P59. L indicates the light chain of LY-CoV555. BFE change range is from -3.80kcal/mol to 2.26kcal/mol.

Competing interests

The authors declare no competing interests.

References

- [1] Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2):271–280, 2020.
- [2] Xiu-Xia Qu, Pei Hao, Xi-Jun Song, Si-Ming Jiang, Yan-Xia Liu, Pei-Gang Wang, Xi Rao, Huai-Dong Song, Sheng-Yue Wang, Yu Zuo, et al.

- 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):29588–29595, 2005.
- [3] Rui Wang, Jiahui Chen, Kaifu Gao, Yuta Hozumi, Changchuan Yin, and Guo-Wei Wei. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Communications Biology*, 4(1):1–14, 2021.
- [4] Jiahui Chen, Rui Wang, Menglun Wang, and Guo-Wei Wei. Mutations strengthened SARS-CoV-2 infectivity. *Journal of Molecular Biology*, 432(19):5212–5226, 2020.
- [5] Rui Wang, Jiahui Chen, Kaifu Gao, and Guo-Wei Wei. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other covid-19-devastated countries. *Genomics*, 113(4):2158–2170, 2021.
- [6] Rui Wang, Jiahui Chen, and Guo-Wei Wei. Mechanisms of SARS-CoV-2 evolution revealing vaccine-resistant mutations in Europe and America. *The Journal of Physical Chemistry Letters*, 12(49):11850–11857, 2021.
- [7] Wendong Li, Zhengli Shi, Meng Yu, Wuze Ren, Craig Smith, Jonathan H Epstein, Hanzhong Wang, Gary Crameri, Zhihong Hu, Huajun Zhang, et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science*, 310(5748):676–679, 2005.
- [8] Jiahui Chen, Kaifu Gao, Rui Wang, and Guo-Wei Wei. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chemical Science*, 12(20):6929–6948, 2021.
- [9] Jiahui Chen, Kaifu Gao, Rui Wang, and Guo-Wei Wei. Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies. *Journal of Molecular Biology*, 433(18):167155, 2021.
- [10] Fatima Amanat and Florian Krammer. SARS-CoV-2 vaccines: status report. *Immunity*, 52(4):583–589, 2020.
- [11] Johanna Hansen, Alina Baum, Kristen E Pascal, Vincenzo Russo, Stephanie Giordano, Elzbieta Wloga, Benjamin O Fulton, Ying Yan, Katrina Koon, Krunal Patel, et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science*, 369(6506):1010–1014, 2020.

- [12] Bryan E Jones, Patricia L Brown-Augsburger, Kizzmekia S Corbett, Kathryn Westerndorf, Julian Davies, Thomas P Cujec, Christopher M Wiethoff, Jamie L Blackbourne, Beverly A Heinz, Denisa Foster, et al. Ly-cov555, a rapidly isolated potent neutralizing antibody, provides protection in a non-human primate model of SARS-CoV-2 infection. *BioRxiv*, 2020.
- [13] Rui Shi, Chao Shan, Xiaomin Duan, Zhihai Chen, Peipei Liu, Jinwen Song, Tao Song, Xiaoshan Bi, Chao Han, Lianao Wu, et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature*, 584(7819):120–124, 2020.
- [14] Cheolmin Kim, Dong-Kyun Ryu, Jihun Lee, Young-Il Kim, Ji-Min Seo, Yeon-Gil Kim, Jae-Hee Jeong, Minsoo Kim, Jong-In Kim, Pankyeom Kim, et al. A therapeutic neutralizing antibody targeting receptor binding domain of SARS-CoV-2 spike protein. *Nature Communications*, 12(1):1–10, 2021.
- [15] Pengfei Wang, Manoj S Nair, Lihong Liu, Sho Iketani, Yang Luo, Yicheng Guo, Maple Wang, Jian Yu, Baoshan Zhang, Peter D Kwong, et al. Antibody resistance of SARS-CoV-2 variants b. 1.351 and b. 1.1.7. *Nature*, 593(7857):130–135, 2021.
- [16] Troy Day, Sylvain Gandon, Sébastien Lion, and Sarah P Otto. On the evolutionary epidemiology of SARS-CoV-2. *Current Biology*, 30(15):R849–R857, 2020.
- [17] Tracking SARS-CoV-2 variants, WHO. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>.
- [18] Nicholas G Davies, Sam Abbott, Rosanna C Barnard, Christopher I Jarvis, Adam J Kucharski, James Munday, Carl AB Pearson, Timothy W Russell, Damien C Tully, Alex D Washburne, et al. Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. *MedRxiv*, pages 2020–12, 2021.
- [19] Delphine Planas, David Veyer, Artem Baidaliuk, Isabelle Staropoli, Florence Guivel-Benhassine, Maaran Michael Rajah, Cyril Planchais, Françoise Porrot, Nicolas Robillard, Julien Puech, et al. Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization. *Nature*, pages 1–7, 2021.
- [20] Pengfei Wang, Ryan G Casner, Manoj S Nair, Maple Wang, Jian Yu, Gabriele Cerutti, Lihong Liu, Peter D Kwong, Yaoxing Huang,

- Lawrence Shapiro, et al. Increased resistance of SARS-CoV-2 variant p. 1 to antibody neutralization. *Cell Host & Microbe*, 29(5):747–751, 2021.
- [21] Sonia Jangra, Chengjin Ye, Raveen Rathnasinghe, Daniel Stadlbauer, Hala Alshammary, Angela A Amoako, Mahmoud H Awawda, Katherine F Beach, Maria C Bermúdez-González, Rachel L Chernet, et al. SARS-CoV-2 spike e484k mutation reduces antibody neutralisation. *The Lancet Microbe*, 2021.
- [22] Zhuoming Liu, Laura A VanBlargan, Louis-Marie Bloyet, Paul W Rothlauf, Rita E Chen, Spencer Stumpf, Haiyan Zhao, John M Errico, Elitza S Theel, Mariel J Liebeskind, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe*, 29(3):477–488, 2021.
- [23] Xianding Deng, Miguel A Garcia-Knight, Mir M Khalid, Venice Servellita, Candace Wang, Mary Kate Morris, Alicia Sotomayor-González, Dustin R Glasner, Kevin R Reyes, Amelia S Gliwa, et al. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a 1452r spike protein mutation. *MedRxiv*, 2021.
- [24] Fact Sheet For Health Care Providers Emergency Use Authorization (EUA) of Bamlanivimab and Etesevimab 02092021 (fda.gov).
- [25] Fact Sheet For Health Care Providers Emergency Use Authorization (EUA) of Regen-Cov (fda.gov).
- [26] Kui K Chan, Danielle Dorosky, Preeti Sharma, Shawn A Abbasi, John M Dye, David M Kranz, Andrew S Herbert, and Erik Procko. Engineering human ACE2 to optimize binding to the spike protein of sars coronavirus 2. *Science*, 369(6508):1261–1265, 2020.
- [27] Tyler N Starr, Allison J Greaney, Sarah K Hilton, Daniel Ellis, Katharine HD Crawford, Adam S Dingens, Mary Jane Navarro, John E Bowen, M Alejandra Tortorici, Alexandra C Walls, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cell*, 182(5):1295–1310, 2020.
- [28] Thomas W Linsky, Renan Vergara, Nuria Codina, Jorgen W Nelson, Matthew J Walker, Wen Su, Christopher O Barnes, Tien-Ying Hsiang, Katharina Esser-Nobis, Kevin Yu, et al. De novo design of potent and resilient hace2 decoys to neutralize SARS-CoV-2. *Science*, 370(6521):1208–1214, 2020.

- [29] Tyler N Starr, Allison J Greaney, Amin Addetia, William W Hannon, Manish C Choudhary, Adam S Dingens, Jonathan Z Li, and Jesse D Bloom. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*, 371(6531):850–854, 2021.
- [30] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine HD Crawford, et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host & Microbe*, 29(1):44–57, 2021.
- [31] Takuya Tada, Belinda M Dcosta, Marie Samanovic-Golden, Ramin S Herati, Amber Cornelius, Mark J Mulligan, and Nathaniel R Landau. Neutralization of viruses with european, south african, and united states SARS-CoV-2 variant spike proteins by convalescent sera and bnt162b2 mrna vaccine-elicited antibodies. *BioRxiv*, 2021.
- [32] Lingfei Wu, Dashun Wang, and James A Evans. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382, 2019.
- [33] Nathan D Grubaugh, William P Hanage, and Angela L Rasmussen. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell*, 182(4):794–795, 2020.
- [34] Wilfredo F Garcia-Beltran, Evan C Lam, Kerri St Denis, Adam D Nittido, Zeidy H Garcia, Blake M Hauser, Jared Feldman, Maia N Pavlovic, David J Gregory, Mark C Poznansky, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*, 184(9):2372–2383, 2021.
- [35] Constantinos Kurt Wibmer, Frances Ayres, Tandile Hermanus, Mashudu Madzivhandila, Prudence Kgagudi, Brent Oosthuysen, Bronwen E Lambson, Tulio De Oliveira, Marion Vermeulen, Karin Van der Berg, et al. SARS-CoV-2 501y. v2 escapes neutralization by South African covid-19 donor plasma. *Nature Medicine*, 27(4):622–625, 2021.
- [36] Kai Wu, Anne P Werner, Matthew Koch, Angela Choi, Elisabeth Narayanan, Guillaume BE Stewart-Jones, Tonya Colpitts, Hamilton Bennett, Seyhan Boyoglu-Barnum, Wei Shi, et al. Serum neutralizing activity elicited by mrna-1273 vaccine. *New England Journal of Medicine*, 384(15):1468–1470, 2021.
- [37] Takuya Tada, Hao Zhou, Belinda M Dcosta, Marie I Samanovic, Mark J Mulligan, and Nathaniel R Landau. SARS-CoV-2 lambda variant re-

- mains susceptible to neutralization by mrna vaccine-elicited antibodies and convalescent serum. *BioRxiv*, 2021.
- [38] Soo-Young Lee, Dong-Kyun Ryu, Hanmi Noh, Jongin Kim, Ji-Min Seo, Cheolmin Kim, Carel van Baalen, Aloys SL Tijsma, Hyo-Young Chung, Min-Ho Lee, et al. Therapeutic efficacy of CT-p59 against P.1 variant of SARS-CoV-2. *BioRxiv*, 2021.
- [39] Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ace2 receptor. *Nature*, 581(7807):215–220, 2020.
- [40] Rui Wang, Jiahui Chen, Yuta Hozumi, Changchuan Yin, and Guo-Wei Wei. Emerging vaccine-breakthrough SARS-CoV-2 variants. *ACS Infectious Diseases*, 8(3):546–556, 2022.
- [41] Jiahui Chen, Rui Wang, Nancy Benovich Gilby, and Guo-Wei Wei. Omicron variant (b. 1.1. 529): Infectivity, vaccine breakthrough, and antibody resistance. *Journal of Chemical Information and Modeling*, 62(2):412–422, 2022.
- [42] Jiahui Chen and Guo-Wei Wei. Omicron BA. 2 (b. 1.1. 529.2): high potential to becoming the next dominating variant. *The Journal of Physical Chemistry Letters*, accepted.
- [43] Yuelong Shu and John McCauley. Gisaid: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance*, 22(13):30494, 2017.
- [44] Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, et al. A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798):265–269, 2020.
- [45] Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H Moal. Skempi 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019.
- [46] Menglun Wang, Zixuan Cang, and Guo-Wei Wei. A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation. *Nature Machine Intelligence*, 2(2):116–123, 2020.
- [47] Zixuan Cang, Lin Mu, and Guo-Wei Wei. Representability of algebraic

topology for biomolecules in machine learning based scoring and virtual screening. *PLoS Computational Biology*, 14(1):e1005929, 2018.

- [48] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005. [MR2121296](#)
- [49] Herbert Edelsbrunner, John Harer, et al. Persistent homology-a survey. *Contemporary Mathematics*, 453:257–282, 2008. [MR2405684](#)
- [50] Kelin Xia and Guo-Wei Wei. Persistent homology analysis of protein structure, flexibility, and folding. *International Journal for Numerical Methods in Biomedical Engineering*, 30(8):814–844, 2014. [MR3247713](#)
- [51] Duc Duy Nguyen, Zixuan Cang, and Guo-Wei Wei. A review of mathematical representations of biomolecular data. *Physical Chemistry Chemical Physics*, 22(8):4343–4367, 2020.
- [52] Zhixin Xiang and Barry Honig. Extending the accuracy limits of prediction for side-chain conformations. *Journal of Molecular Biology*, 311(2):421–430, 2001.
- [53] Rui Wang, Yuta Hozumi, Changchuan Yin, and Guo-Wei Wei. Mutations on COVID-19 diagnostic targets. *Genomics*, 112(6):5204–5213, 2020.

JIAHUI CHEN

DEPARTMENT OF MATHEMATICS

MICHIGAN STATE UNIVERSITY

EAST LANSING

MI 48823

USA

E-mail address: chenj159@msu.edu

GUO-WEI WEI

DEPARTMENT OF MATHEMATICS

MICHIGAN STATE UNIVERSITY

EAST LANSING

MI 48823

USA

E-mail address: weig@msu.edu

RECEIVED MARCH 4, 2022