

Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization

Sisi Shan^{a,1}, Shitong Luo^{b,1}, Ziqing Yang^{a,1}, Junxian Hong^{a,1}, Yufeng Su^c, Fan Ding^b, Lili Fu^a, Chenyu Li^a, Peng Chen^a, Jianzhu Ma^b, Xuanling Shi^a, Qi Zhang^a, Bonnie Berger^{d,e,2}, Linqi Zhang^{a,2}, and Jian Peng^{b,c,f,2}

^aNexVac Research Center, Comprehensive AIDS Research Center, Center for Infectious Disease Research, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China; ^bHelixon Limited, Beijing 100084, China; ^cDepartment of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL 61822; ^dComputer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; ^eDepartment of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^fInstitute for Industry AI Research, Tsinghua University, Beijing 100084, China

Contributed by Bonnie Berger; received December 20, 2021; accepted January 29, 2022; reviewed by Nir Ben-Tal and Victoria Popic

The ability of viruses to mutate and evade the human immune system and neutralizing antibodies remains an obstacle to antiviral and vaccine development. Many neutralizing antibodies, including some approved for emergency use authorization (EUA), reduced or lost activity against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants. Here, we introduce a geometric deep learning algorithm that efficiently enhances antibody affinity to achieve broader and more potent neutralizing activity against such variants. We demonstrate the utility of our approach on a human antibody P36-5D2, which is effective against SARS-CoV-2 Alpha, Beta, and Gamma but not Delta. We show that our geometric neural network model optimizes this antibody's complementarity-determining region (CDR) sequences to improve its binding affinity against multiple SARS-CoV-2 variants. Through iterative optimization of the CDR regions and experimental measurements, we enable expanded antibody breadth and improved potency by ~10- to 600-fold against SARS-CoV-2 variants, including Delta. We have also demonstrated that our approach can identify CDR changes that alleviate the impact of two Omicron mutations on the epitope. These results highlight the power of our deep learning approach in antibody optimization and its potential application to engineering other protein molecules. Our optimized antibodies can potentially be developed into antibody drug candidates for current and emerging SARS-CoV-2 variants.

computational biology | deep learning | geometric neural networks | SARS-CoV-2 variants | broadly neutralizing antibodies

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread worldwide over the past 2 y, causing hundreds of millions of confirmed infections and millions of deaths (1). The receptor-binding domain (RBD) of the SARS-CoV-2 virus spike protein initiates binding to the host receptor, angiotensin converting enzyme 2 (ACE2) (2–6), and serves as an initial essential step in viral-cell membrane fusion, as well as a potential target for neutralizing antibodies (7–10). Neutralizing antibodies that target RBD have already shown therapeutic and clinical value (11–17).

However, reduced sensitivity of SARS-CoV-2 variants to antibody and serum neutralization has been widely observed (18–21). For example, the B.1.617 lineage, also known as the Delta variant, contains two mutations (L452R and T478K) in the RBD that facilitate viral escape—the ability of viruses to evade the immune system and cause disease (22). The L452R mutation is located at the periphery of the receptor binding motif (RBM) and is found to reduce neutralizing activity by antibodies. The T478K mutation in the RBD, located within the epitope region in the RBM, is also associated with antibody escape. There has been striking evidence of antibodies that have been greatly affected, or even have lost their neutralizing activity altogether, by viral escape (23–26).

Experimental methods to improve antibody binding and neutralization have been developed. In vitro affinity maturation methods, such as random mutagenesis with display technologies, has been shown to improve antibody binding against target proteins, but such approaches are time consuming and labor intensive (27–32). Targeted optimization toward one particular variant may also result in loss of neutralizing activity against other variants. Efficient optimization of antibodies that confer broad and potent neutralizing activity against diverse variants is therefore urgently needed.

Here, we develop and apply a deep learning framework to efficiently optimize antibodies to achieve broader and more potent neutralizing activity against SARS-CoV-2 variants. Based on a large collection of antibody-antigen complex structures and binding affinity data, we trained a geometric neural network model, recently developed in computer vision, that effectively extracts interresidue interaction features and makes predictions of changes in binding affinity due to single or multiple amino acid substitutions to the antigen. To search for favorable complementarity-determining region (CDR) mutations that potentially improve antibody binding, we also simulate an in silico ensemble of predicted complex structures with CDR

Significance

SARS-CoV-2 continues to evolve through emerging variants, more frequently observed with higher transmissibility. Despite the wide application of vaccines and antibodies, the selection pressure on the Spike protein may lead to further evolution of variants that include mutations that can evade immune response. To catch up with the virus's evolution, we introduced a deep learning approach to redesign the complementarity-determining regions (CDRs) to target multiple virus variants and obtained an antibody that broadly neutralizes SARS-CoV-2 variants.

Author contributions: B.B., L.Z., and J.P. designed research; S.S., S.L., Z.Y., J.H., Y.S., F.D., L.F., C.L., P.C., J.M., X.S., Q.Z., and J.P. performed research; S.S., L.F., J.M., and Q.Z. contributed new reagents/analytic tools; S.S., S.L., Z.Y., J.H., Y.S., F.D., B.B., L.Z., and J.P. analyzed data; and S.S., S.L., Z.Y., B.B., L.Z., and J.P. wrote the paper.

Reviewers: N.B.-T., Tel Aviv University; and V.P., Broad Institute.

Competing interest statement: A patent has been filed for the optimized antibodies targeting SARS-CoV-2. J.P., J.M., and S.L. are employees of Helixon.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹S.S., S.L., Z.Y., and J.H. contributed equally to this work.

²To whom correspondence may be addressed. Email: bab@mit.edu, zhanglinqi@tsinghua.edu.cn, or jianpeng@helixon.com.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2122954119/-DCSupplemental>.

Published March 1, 2022.

mutations to obtain a robust estimation of the free energy change, also known as $\Delta\Delta G$. Compared to traditional approaches, the deep learning search space is theoretically much larger and is also easily applicable in targeting multiple variants simultaneously via multiobjective optimization.

To demonstrate the utility of our approach, we sought to optimize a human neutralizing antibody P36-5D2, which was initially isolated from a convalescent patient, and demonstrated reasonably strong potency and breadth against Alpha, Beta, and Gamma (33) but not Delta, due to Delta's L452R but not T478K mutation through computational structure analysis. We applied our deep learning model to predict CDR sequences that potentially improved binding affinity against the Delta variant while maintaining activity against Alpha, Beta, and Gamma. Through an iterative process of modeling and experimental validation, we were able to obtain six optimized antibodies with substantially improved potency of about 10- to 600-fold against multiple variants, including Delta. We also provide initial promising studies on Omicron. These results highlight the power of deep learning approaches for antibody optimization and their potential application to a wide range of other protein molecules. The optimized antibodies presented here also have the potential to be further developed as antibody drug candidates against SARS-CoV-2 variants.

Results

Neural Network Architecture. We tapped neural networks because they are known as universal approximators for learning complex nonlinear mappings from high-dimensional data with generalizability. Among many successful neural network architectures, attention-based networks are extremely powerful for modeling complicated interactions between entities, making them suitable for modeling proteins whose functions are largely determined by interactions between amino acids (34–37).

We developed an attention-based geometric neural network architecture to learn the mutational effect on protein–protein interactions from three-dimensional protein complex structures (Fig. 1). The geometric part of the model learns a vector embedding for each residue by considering the proximity of its surrounding atoms. Based on these learned geometric embeddings, the attention network learns to identify key residue pairs near the protein interface contributing to binding affinity. Specifically, for each residue in a protein complex, the network first identifies the importance of other residues via the attention mechanism and aggregates information including spatial proximity and physicochemical properties from them. The aggregated information thus can encode the environment as well as interaction features for each residue. To estimate the effect of mutation(s), we first predict the structure of mutated protein complexes by repacking sidechains around mutation sites and encode both wild-type (WT) and mutant complexes using the network to obtain both WT and mutant embeddings. Then, additional neural network layers compare the two embeddings to predict the effect of mutation measured by $\Delta\Delta G$. This model was evaluated by split-by-complex fivefold cross-validation over the Structural Kinetic and Energetic database of Mutant Protein Interactions (SKEMPI) V2.0 dataset (38). A subset consisting of 1,131 single-point mutations (S1131) (39) was used to benchmark the model and other baselines. The Pearson correlations between the predicted $\Delta\Delta G$ values by different methods and real $\Delta\Delta G$ are reported in *SI Appendix, Table S1*. An additional subset consisting of multipoint mutations (M1707) (40) was also used as a benchmark. We found that our model is able to make predictions with moderate to high correlation with experimental binding data and also outperforms GeoPPI (41), the current state-of-the-art method, as well as a few other recently proposed methods for predicting single-mutation effects. To further evaluate mutated antibody–antigen complexes, we additionally

used two other methods, Rosetta and GeoPPI, that also predict mutational effects on stability and binding (41–43), in order to build an ensemble method to evaluate single and higher-order mutations. Our ensemble method is thus able to identify modified antibody sequences that potentially improve antigen binding.

Antibody of Choice. We sought to experiment with an antibody that has an epitope with highly conserved regions, which makes it a good candidate to become potent against multiple SARS-CoV-2 variants. We focused on a broad neutralizing antibody P36-5D2, isolated from convalescent patients, known to neutralize Alpha, Beta, and Gamma (33). However, P36-5D2 activity against Delta is reduced by ~ 40 -fold, relative to that of WT SARS-CoV-2 (Fig. 1D). Structural information on the epitope of P36-5D2 indicated that such a reduction was likely attributed to L452R and T478K mutations found in the RBD of Delta. The longer sidechain of the L452R mutation might have exerted steric interference for antibody recognition.

Generalization to Delta Variant. To improve the neutralizing activity of P36-5D2 against the Delta variant, we applied our deep learning method to predict and identify optimized antibody mutations. While the initial structure of P36-5D2 bound with the RBD, the neutralizing activity of P36-5D2 against SARS-CoV-2 WT, Beta, Gamma, and Delta was then utilized to analyze and select CDR mutations. Based on those data, our method generated an in silico mutation library of antibody CDRs, ranked by trained geometric neural networks in such a way that they should not only improve antibody binding to the Delta RBD but also maintain binding to the RBD of other variants of concern (VOC). The predicted top-ranked mutations were selected for experimental verification via pseudotyped virus neutralization assays (*Methods*). A total of four rounds of optimization for P36-5D2 were performed to select the optimized antibodies with best neutralizing potency and breadth. During experimental evaluation, we also included REGN-10987 (44), an ultrapotent antibody approved for emergency use authorization, as the positive control.

First-round computational results. We first sought to narrow down which optimized antibodies to experimentally test, the number determined by our experimental capacity. According to the model's predictions, 12 top-ranked single mutations were selected and introduced into the original antibody P36-5D2. These mutations included T30P, T31W, and Y32P in HCDR1; N52D, A53F, N55L, and N57L in HCDR2; G100P, R103M, L104F, and Q105F in HCDR3; and S30Y in LCDR1. Out of the 12 single-mutation sites, only 4 (T30, T31, R103, and Q105) were positioned on the paratope of P36-5D2 bound to the RBD, while the remaining 8 were outside and did not directly interact with the RBD. To evaluate the neutralizing breadth and potency of the optimized antibodies, we used an established panel of 10 pseudotyped viruses, including SARS-CoV-2 WT 614G, four VOCs—Alpha, Beta, Gamma, Delta—as well as Delta Plus, two VOIs—Kappa, Epsilon which carried L452R, and Eta which carried E484K—and the variant N439K which effected reference antibody REGN10987.

First-round experimental results. In the first-round neutralization shown in Fig. 2, the original antibody P36-5D2 neutralizes SARS-CoV-2 WT, Alpha, Beta, and Gamma, but is affected by Delta and other viruses carrying L452R, consistent with our structural analysis and predictions. R103M in HCDR3, which was the most surprising mutation, significantly improved neutralizing activity of P36-5D2 against all 10 pseudotyped viruses with average half-maximal inhibitory concentration (IC_{50}) reaching 0.038 $\mu\text{g/mL}$. Four mutations, T31W in HCDR1, A53F, and N57L in HCDR2, and L104F in HCDR3, also improved the overall neutralizing activity, with average IC_{50} of 0.479 $\mu\text{g/mL}$, 0.547 $\mu\text{g/mL}$, 0.025 $\mu\text{g/mL}$, and 0.104 $\mu\text{g/mL}$, respectively. Three

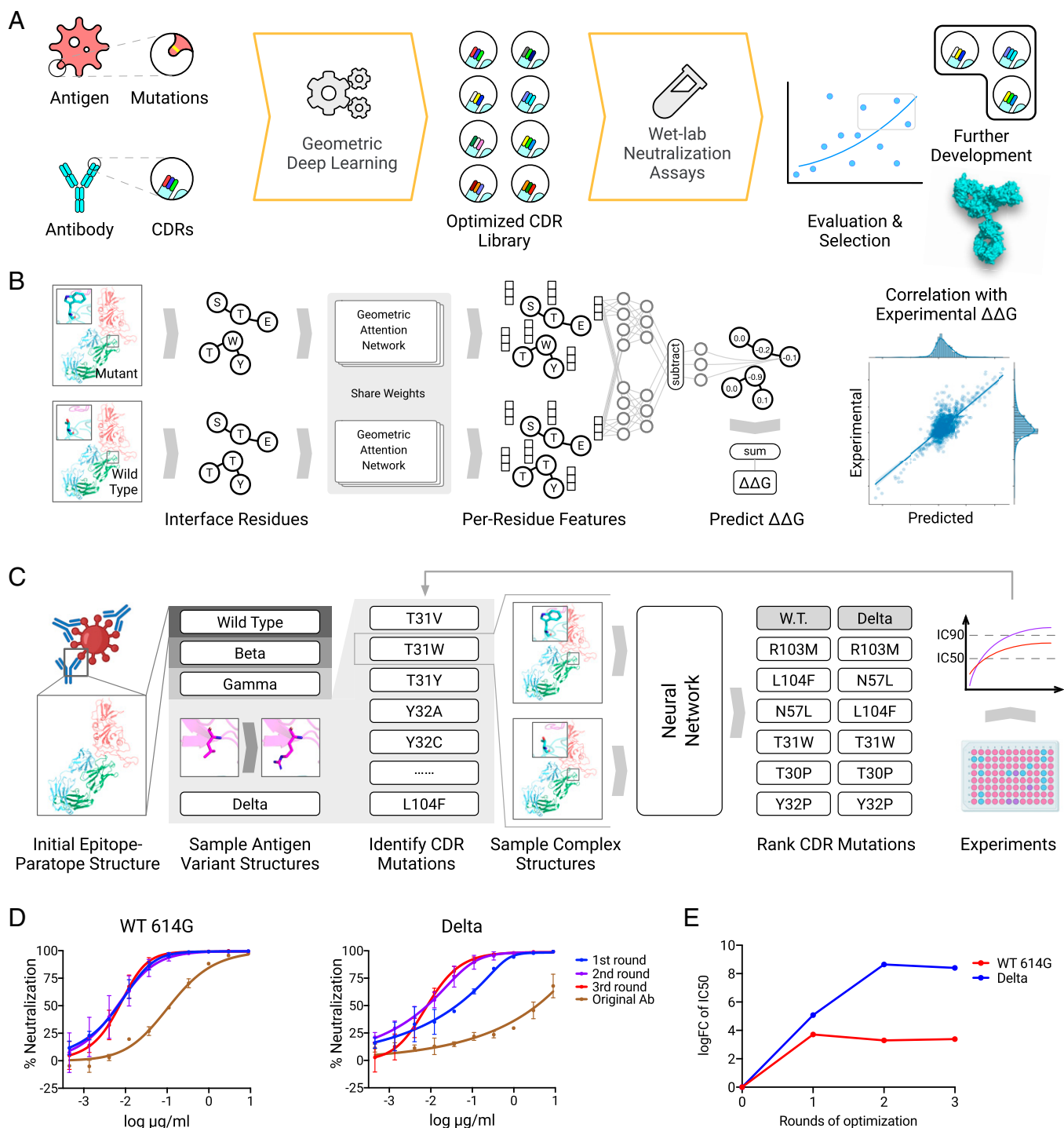


Fig. 1. Deep learning guided antibody optimization platform. (A) Overview of the pipeline. It demonstrates the computational/experimental feedback loop to refine antibody design. (B) Geometric deep learning model. The WT complex and the mutated complex structures are encoded using a shared geometric attention network. The effect of mutation measured by $\Delta\Delta G$ is then predicted by a network that compares features of the two complexes. (C) P36-5D2 antibody optimization. Given the complex structure, we first simulate different variants and then evaluate potential CDR mutations that will improve binding by predicted $\Delta\Delta G$ values. Mutants with top $\Delta\Delta G$ scores are examined in laboratory experiments, and those with neutralizing potency are combined for the next round of optimization. (D) Optimization improves neutralization ability against SARS-CoV-2 and Delta variant. (E) The log fold changes of IC_{50} relative to the original antibody.

mutations, N52D, N55L in HCDR2, and G100P in HCDR3, obtained enhanced neutralization against Delta but reduced neutralization against WT. The four mutations T30P, Y32P in HCDR1, Q105F in HCDR3, and S30Y in LCDR1 had negligible impact on the neutralizing activity of P36-5D2.

Second round: From single to double mutations improves potency. Based on the first-round results, we combined single mutations to create double mutants and selected them also according to $\Delta\Delta G$ predictions. As R103M was the best single mutation in the first round, the double mutations were designed to pair

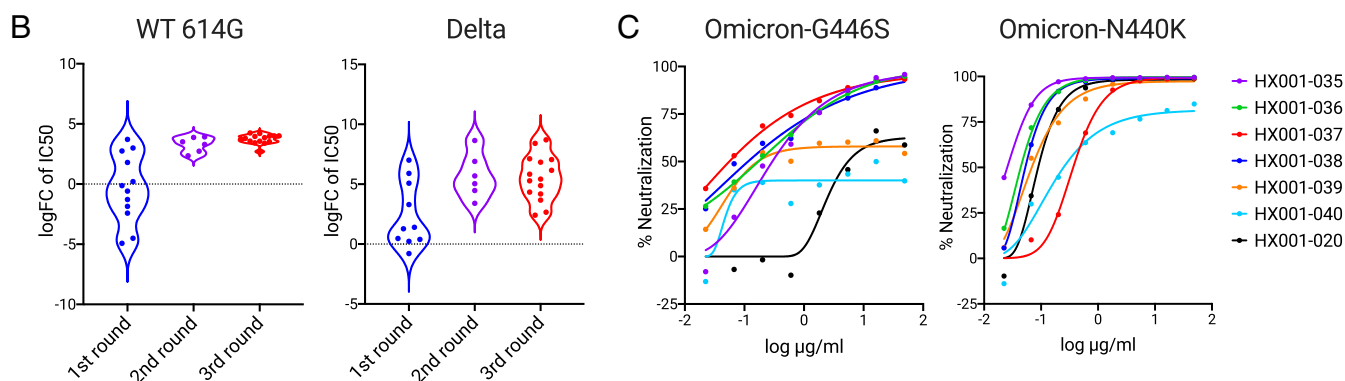
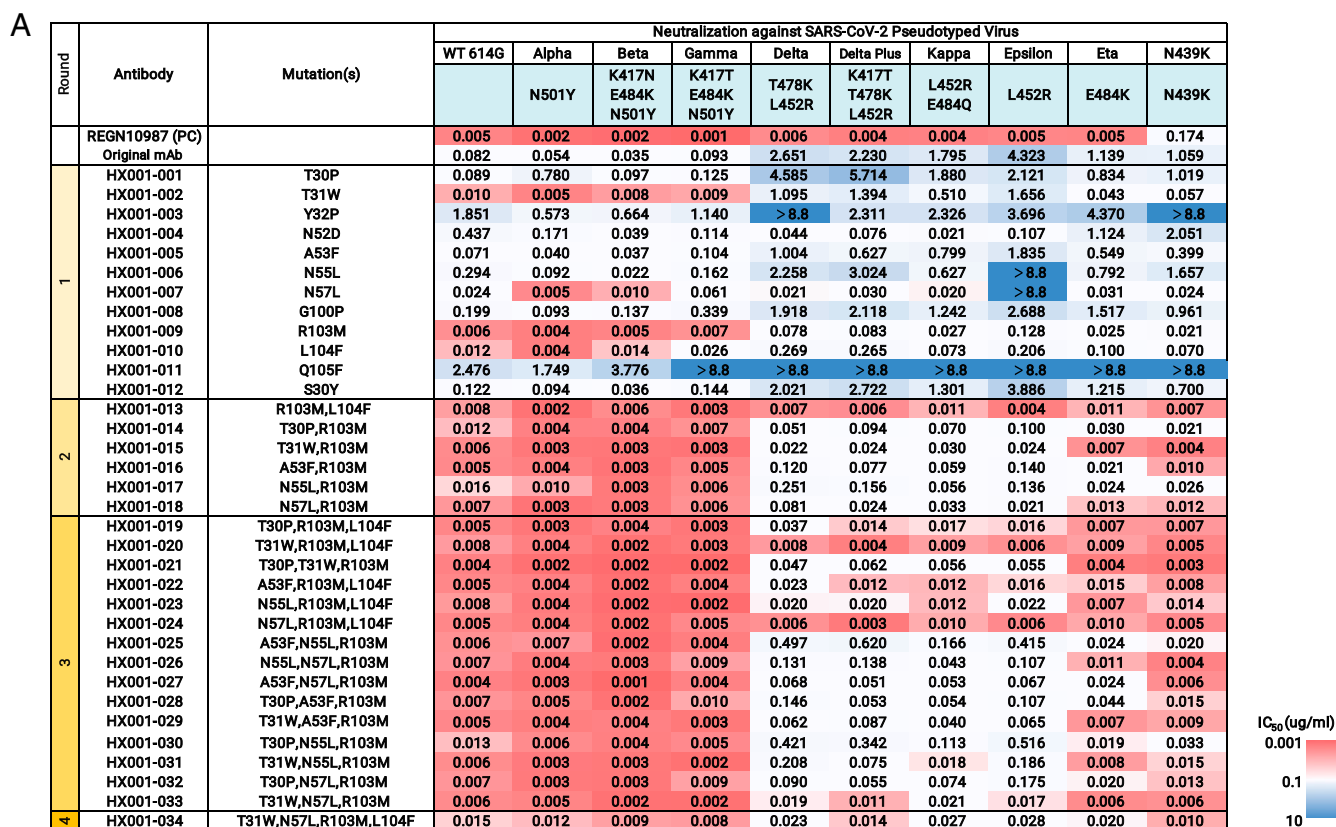


Fig. 2. Evaluation of the neutralization level of optimized antibodies. (A) IC₅₀ values of optimized antibodies against pseudotyped SARS-CoV-2 variants. Mutations of each optimized antibody and mutations of each variant on RBD are indicated. Results were calculated from three independent experiments. (B) The log fold changes of IC₅₀ relative to the original antibody against SARS-CoV-2 and Delta variants are calculated. (C) Neutralization curves of optimized against pseudotyped SARS-CoV-2 carrying Omicron mutations N440K or G446S.

R103M with another mutation, including L104F, T30P, T31W, A53F, N55L, or N57L in HCDRs. Each of those six double mutations achieved over an order of magnitude improvement in neutralizing activity over the best single mutation R103M in the first round, especially for HX001-013 (R103M and L104F) and HX001-015 (R103M and T31W), for which average IC₅₀ values increased to 0.007μg/mL and 0.013μg/mL, respectively (Fig. 2). HX001-013 has an IC₅₀ similar to the reference antibody REGN10987 and is more potent than REGN10987 in neutralizing the N439K variant.

Third-round optimization identifies potent triple mutations. Following a similar strategy, single mutations proven to improve potency and breadth in the first and second rounds of optimization were combined to create 15 antibodies with triple mutations.

They included R103M and an additional two selected from the following six residues: T30P, T31W, A53F, A55L, N57L, and L104F. As expected, optimized antibodies with three mutations demonstrated improved potent neutralizing activity against the pseudotyped viruses, especially HX001-020 (R103M, T31W, and L104F), HX001-024 (R103M, N57L, and L104F), and HX001-033 (R103M, T31W, and N57L) with average IC₅₀ of 0.006 μg/mL, 0.006μg/mL, and 0.010μg/mL, respectively (Fig. 2). HX001-020 and HX001-024 displayed IC₅₀ values similar to the reference antibody REGN10987 in general and were better than REGN10987 in neutralizing the N439K mutant.

Fourth round investigates quadruple mutants. Since optimized antibodies with triple mutations such as T31W/N57L/R103M, R103M/T31W/N57L, and R103M/N57L/L104F all demonstrated

strong neutralizing activity in round three, we constructed HX001-034, a quadruple mutant, to include all four single mutations. HX001-034 had potent neutralizing activity, with an IC_{50} of 0.017 μ g/mL, yet was slightly weaker than three mutation combinations: HX001-020, HX001-024, and HX001-033 (Fig. 2).

Optimized Antibodies Neutralizing Delta. To experimentally test our predicted mutations, we compared the binding affinity of the original antibody and optimized ones against the SARS-CoV-2 WT RBD monomer and Delta RBD monomer by surface plasmon resonance (SPR) (Table 1). The dissociation constant (K_D) of the optimized antibodies to the SARS-CoV-2 WT RBD was 1.2 nM to 0.42 nM, which is 20- to 50-fold stronger than the original P36-5D2 antibody. While the off-rate (k_d) value of the original antibody against WT RBD was around 10^{-2} , the off-rate (k_d) values of the six optimized antibodies all reached 10^{-3} , signifying a longer half-life binding period and higher binding stability. Antibodies HX001-020, HX001-024, HX001-033, and HX001-034 with three or four mutations were also stronger than HX001-013 with only two mutations. The increase in binding affinity may contribute to the increased neutralizing activity of these antibodies against SARS-CoV-2 WT and variants. Notably, the dissociation constant of optimized antibodies with the SARS-CoV-2 Delta RBD was 6.22 nM to 26.4 nM, which is twofold stronger than or similar to P36-5D2, while the K_D of optimized antibodies with the SARS-CoV-2 Delta RBD is still slightly weaker than with WT RBD.

Structural Investigations of Antibody Neutralization. To help explain why our predicted mutations could improve the neutralizing activity from a structural perspective, we used Rosetta to predict the complex structures of WT/Delta RBD with antibodies carrying four critical single mutations: T31W, N57L, R103M, or L104F. As shown in Fig. 3A, T31 on the original antibody HCDR1 might cause steric clash with the sidechain of R452 on the Delta RBD because the minimal distance between the heavy atoms on two sidechains is within 4 Å. After substituting T31 with W31, which does not have a large sidechain, we no longer see the steric clash with R452 on the Delta RBD. As shown in Fig. 3B, N57 on HCDR2 is not on the Delta RBD interaction interface and has no direct contact with Y449 or R452. After substituting N57 with L57, we observed a new interaction between L57's sidechain and Y449, which potentially explains the enhanced binding affinity of this mutation. As shown in Fig. 3C, R103 on the original antibody HCDR3 is spatially very close to the Delta RBD, which may also induce a steric clash with R346. Since both R103 and R346 have very long sidechains and carry positive charges, proximity between the two may introduce strong repulsion which may greatly reduce the binding affinity between the antibody and the antigen. After substituting M103, which has a much smaller sidechain, we no longer observe a direct interaction with R346 on the Delta RBD. This factor might explain the greatly improved neutralization against the Delta variant. Similarly, we can also see that the substitution of L104 for F104 on the HCDR3 might

also remove a potential steric clash with R452 on the Delta RBD and thus improve binding and neutralization (Fig. 3D).

Generalization to Omicron Mutations. A new SARS-CoV-2 variant Omicron emerged in South Africa at the end of November 2021. Omicron carries 36 total mutations in the spike protein and 15 mutations in the RBD, which greatly changed the immunogenicity of the RBD and caused many neutralizing antibodies to lose their neutralization (45–47). Through structure analysis, we found that the G446S mutation was located on the epitope of antibody P36-5D2, and N440K was near (within 5 Å) the epitope. Thus, we applied our deep learning approach to further optimize the antibody sequence HX001-020. The best mutations identified include N92F (HX001-035), G93M (HX001-036), Y94G (HX001-037), and Y74D (HX001-038) in LCDR3. We also constructed a double mutant, HX001-039 (N92F/Y94G), and a triple mutant (N92F/G93M/Y94G). HX001-041 was not produced. It can be seen from Fig. 2C that the optimized antibodies were one-fold to threefold higher than HX-020 in neutralizing against N440K, and 20- to 100-fold stronger in the pseudotyped virus carrying the G446S mutation. These results demonstrate that adaptive deep learning optimization holds promise for designing potent antibodies against Omicron and future emerging variants.

Summary. Collectively, the deep learning platform presented here efficiently optimized the original antibody to obtain broader and more potent neutralizing activity against SARS-CoV-2 variants in only 2 wk for each single-round optimization. Optimized antibodies with single mutations or their combinations can achieve 10- to 600-fold improvement in neutralizing activity against SARS-CoV-2 variants, including Delta. In all optimized antibodies, six antibodies (HX001-013, HX001-015, HX001-020, HX001-024, HX001-033, and HX001-034) had the most potent and broad neutralizing activity, with an IC_{50} between 0.006 μ g/mL and 0.017 μ g/mL, indicating the capability of our deep learning method to predict mutations with improved neutralizing ability of antibodies against WT or other variants.

Discussion

In this study, we have reported the efficient optimization of a single original antibody P36-5D2 to obtain broader and more potent neutralizing activity against multiple SARS-CoV-2 variants via a cutting-edge, deep learning guided approach. Based on the crystal structure of the interaction of the RBD and antibodies, and neutralizing data, our geometric neural network model created libraries of CDR mutations and ranked each mutation according to its contribution to binding affinity and structural stability. Through an iterative optimization procedure, we found that optimized antibodies exhibited broader and much more potent neutralizing activity compared to the original antibody. More detailed neutralization assays and binding affinity analyses by SPR further proved that optimized antibodies had hundredfold improvements against specific viral strains.

Table 1. Binding kinetics of original Ab and optimized antibodies with SARS-CoV-2 RBD measured by SPR

mAb name	WT RBD			Delta RBD		
	k_a (1/Ms)	k_d (1/s)	K_D (nM)	k_a (1/Ms)	k_d (1/s)	K_D (nM)
Original antibody	1.37×10^6	3.14×10^{-2}	22.9	2.58×10^6	2.94×10^{-2}	11.4
HX001-013	1.49×10^6	1.79×10^{-3}	1.20	2.42×10^6	1.89×10^{-2}	7.81
HX001-015	3.94×10^6	2.53×10^{-3}	0.64	1.50×10^6	3.97×10^{-2}	26.4
HX001-020	4.50×10^6	2.51×10^{-3}	0.56	2.73×10^6	2.03×10^{-2}	7.44
HX001-024	2.77×10^6	2.28×10^{-3}	0.82	2.16×10^6	1.34×10^{-2}	6.22
HX001-033	3.10×10^6	1.29×10^{-3}	0.42	2.83×10^6	4.37×10^{-2}	15.5
HX001-034	5.10×10^6	3.07×10^{-3}	0.60	2.64×10^6	1.36×10^{-2}	5.16

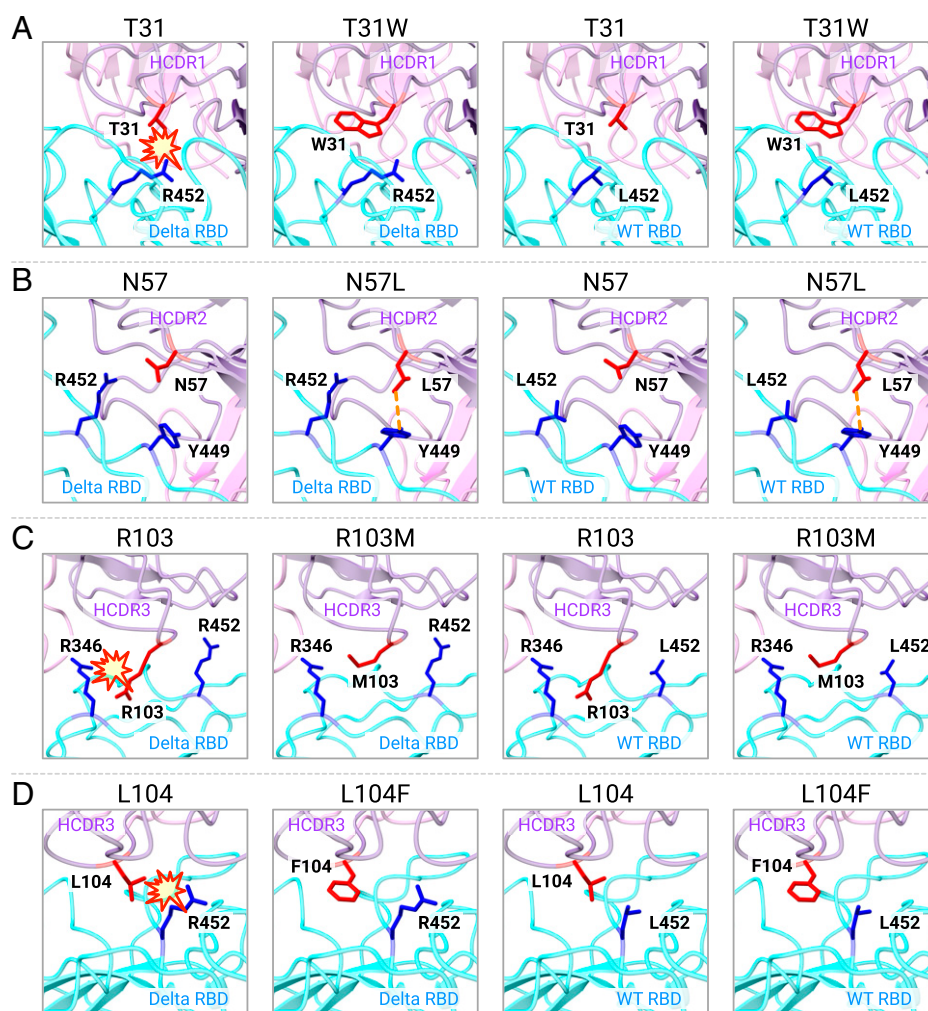


Fig. 3. Predicted structure of important mutations on optimized antibodies. (A) Predicted structure of interactions between original antibody carrying T31 (red) or optimized antibody carrying W31 (red) with related residue R452/L452 (blue) on Delta/WT RBD (cyan); heavy chain is labeled in purple, and light chain is labeled in pink. (B) Predicted structure of interactions between original antibody carrying N57 (red) or optimized antibody carrying L57 (red) with Y449 (blue) and R452/L452 (blue) on Delta/WT RBD. (C) Predicted structure of interactions between original antibody carrying R103 (red) or optimized antibody carrying M103 (red) with related residues R346 (blue) and R452/L452 (blue) on Delta/WT RBD. (D) Predicted structure of interactions between original antibody carrying L104 (red) or optimized antibody carrying F104 (red) with R452/L452 (blue) on Delta/WT RBD.

By deep learning and computational modeling of optimized antibodies of SARS-CoV-2 RBD variants, we observed that predicted optimized antibody candidates could successfully abrogate preexisting steric clashes with RBD at the interface. Some other mutations significantly increase the number of interaction sites, expanding antibody epitopes to more regions on the surface of RBD. Interestingly, a few good mutations were located away from the antibody–RBD interface, which we speculate improve antibody stability and increase CDRs conformational exposure to enhance the neutralization. These results strongly suggested that our model can successfully predict potential modifications of antibodies that are critical to bind to an antigen, with increased structural stability and outstanding neutralizing activity. Surprisingly, peripheral sites not on the binding interface have also been picked up by our deep learning model which seem to be a challenge for conventional structural analysis. Our deep learning model exhibits superior advantages in antibody prediction and optimization for more broader and potent neutralization.

As the virus continues to evolve, more and more diverse variants will emerge in the infected population with greater immune pressure. Many such variants, including Delta and the

most recent Omicron, have adopted different mutation strategies to evade immune recognition, leading to the failure of neutralizing antibodies and vaccines (18, 23, 45–48). Undoubtedly, it is urgent to develop and optimize antibodies to overcome emerging or future variants with broad neutralization activities. The conventional *in vitro* affinity maturation usually takes a long time with poor efficiency due to its small library content, incompatible optimization for multiple goals, and other technical obstacles (32, 49). Directed maturation toward a new variant may also cause the antibody to lose neutralization against previous variants. One important aspect of our approach is that our integrated approach is able to simultaneously optimize binding or neutralization against multiple RBD variants. Based on a deep learning method, we are able to select promising CDR mutation candidates and effectively combine them to achieve efficient antibody optimization. The time spent for each round of the optimization is about 2 wk, most of which is for experimental mutant construction and neutralization assays, much faster than conventional affinity maturation procedures.

Based on our platform, it would even be feasible to train a pool of antibodies that target different potential variants. It is also critical to develop a system to predict virus variants with

potential immune escape capabilities such as in Hie et al. (22), even before they emerge. Altogether, we believe that further development of these methods will allow us to predict and keep pace with viral evolution.

Methods

Building In Silico Mutation Libraries and Predicted Structures. Variant-specific amino acid substitutions in the RBD were first created to build complexes with different antigen variants. For each antigen variant, single-point mutations in CDR of the antibody were enumerated. The structures of mutated complexes were sampled using the Rosetta relax and Cartesian $\Delta\Delta G$ programs (43). Point mutations with better predicted binding affinity and experimental neutralization were combined to double and triple mutations, which go through the next round of selection.

Extraction of Interresidue Interaction Features by Geometric Neural Networks. Each residue in the complex structure was encoded into a vector representation using a geometric attention network (34) designed to capture interresidue interactions. A residue was represented by its $C\alpha$ coordinate denoted as t_i and an orientation matrix R_i defined as

$$R_i = \begin{bmatrix} u_{i1} & u_{i2} \\ \|u_{i1}\| & \|u_{i2}\| \end{bmatrix} \times \begin{bmatrix} u_{i1} & u_{i2} \\ \|u_{i1}\| & \|u_{i2}\| \end{bmatrix}^{-1}$$

where $p_i^{(N)}$ is the position of N, $p_i^{(C)}$ is the position of C, and "Gram-Schmidt" denotes the Gram-Schmidt process for orthogonalization, where, here, u_{i1} , u_{i2} are the orthogonal vectors. Each residue was also assigned an initial feature $h_i^{(0)}$ encoding its amino acid type and the local geometry of its atoms,

$$h_i^{(0)} = MLP_{\text{amino acid}}(R_i^T(p_i^{(a)} - t_i) \mid a \in \{N, C\alpha, C, O, C\beta, \dots\}),$$

where each amino acid type was encoded by a separate multilayer perceptron (MLP) with local coordinates of atoms as input. Each residue pair was represented as a vector z_{ij} indicating the relative position of the two residues. The feature vector of a residue was updated iteratively by the multihead attention network (the computational components, or heads, of the transformer network that are repeated multiple times in parallel). The attention weight was formulated by considering residue features, pairwise features, and distances,

$$\begin{aligned} \beta_{ij}^h &= \frac{1}{d} \text{Linear}(h_i^{(l)}) \cdot \text{Linear}(h_j^{(l)}), \\ \gamma_{ij}^h &= \text{Linear}(z_{ij}), \\ \delta_{ij}^h &= w_h \|p_i^{(C\beta)} - p_j^{(C\beta)}\|, \\ \alpha_{ij}^h &= \text{Softmax}_j(\beta_{ij}^h + \gamma_{ij}^h + \delta_{ij}^h), \end{aligned}$$

where h denotes the number of the attention head, d is the dimension of $h_i^{(l)}$, and w_h is a learnable factor. The message passed from residue j to residue i was formulated as

$$\begin{aligned} u_i^h &= \sum_j \alpha_{ij}^h \text{Linear}(h_j^l), \\ v_i^h &= \sum_j \alpha_{ij}^h \text{Linear}(z_{ij}^l), \\ w_i^h &= \sum_j \alpha_{ij}^h R_j^T(p_j^{(C\beta)} - p_i^{(C\beta)}), \\ m_i^h &= \text{Concat}(u_i^h, v_i^h, w_i^h). \end{aligned}$$

Finally, message vectors m_i^h from different heads were concatenated to update residue features with a residual connection and a layer normalization step (which directly estimates the normalization statistics from the summed inputs to the neurons within a hidden layer so the normalization does not introduce any new dependencies between training cases). By exchanging information among residues, the final feature vector of each residue $h_i^{(L)}$ encodes its interaction with surrounding residues and will subsequently be used to predict the change in binding affinity.

Prediction of the Change in Binding Affinity upon Mutation. Residues in the WT complex and the mutant complex were encoded using the aforementioned geometric neural network. As a complement to the features output by the network, physical energy values of a complex calculated by Rosetta were also included as a feature shared across residues in the complex. Let h_i^{wt} denote the feature of the i th residue in the WT complex, and let h_i^{mut} denote the feature of its counterpart in the mutant complex. These features were used as input to the following antisymmetric network so as to predict the difference in binding affinity between the two complexes:

$$\begin{aligned} x_i &= MLP_1(h_i^{\text{wt}}, h_i^{\text{mut}}) - MLP_1(h_i^{\text{mut}}, h_i^{\text{wt}}), \\ D &= W \sum_i x_i, \end{aligned}$$

where, MLP_1 is a standard MLP network, and W is a trainable weight matrix.

Model Training. The training and validation datasets were constructed from the SKEMPI V2.0 dataset (38), the largest antibody-antigen binding affinity dataset, which contains 342 complexes and 5,318 mutations in total. Mutant structures were sampled by the Rosetta Cartesian $\Delta\Delta G$ program. The dataset was split evenly into five subsets without overlapping complexes. Accordingly, five models were trained, using each subset for validation. For each chain in a complex, we chose only the 48 residues spatially closest to the other chain to train the model, because it was sufficient to consider only residues on or close to the interface. The loss function we used for training was the mean-squared error (MSE) between the predicted $\Delta\Delta G$ values and the experimentally measured values. The model was trained for 100,000 iterations using the Adam optimizer at a learning rate of $1e-4$ until convergence. Validation was performed every thousand training steps to monitor the optimization procedure. The learning rate was reduced by half if the lowest validation loss did not drop for 10 validation iterations.

Evaluating Model Performance. The model was evaluated by split-by-complex fivefold cross-validation over the entire SKEMPI V2.0 dataset. A subset consisting of 1,131 single-point mutations (S1131) (39) was used to benchmark the model and other baselines (50, 51). An ensemble model that averages the prediction of our model, GeoPPI, and Rosetta was also evaluated. Pearson correlation between the predicted $\Delta\Delta G$ values by different methods and real $\Delta\Delta G$ is reported in [SI Appendix, Table S1](#). Figure plots of $\Delta\Delta G$ predicted by the model and experimental $\Delta\Delta G$ are reported in Fig. 1B. An additional subset consisting of multipoint mutations (M1707) (40) was also used as a benchmark; Pearson correlation on this subset is reported in [SI Appendix, Table S1](#). Overall, our model's predictions were highly correlated with experimental results.

Ranking Mutations by Model Ensemble. Mutants in CDRs were scored using the average predicted $\Delta\Delta G$ from an ensemble of five trained models, as the models were trained in a fivefold cross-validation manner. Mutations were grouped by antigen variants and sorted in ascending order; those with lower $\Delta\Delta G$ values were more favorable. We also used GeoPPI as another $\Delta\Delta G$ predictor for ranking mutations in PPI complexes based on a pretrained graph neural network and a random forest regressor (41). The $\Delta\Delta G$ values reported by the Rosetta Cartesian $\Delta\Delta G$ program were also employed to rank mutations. The average of three rankings serves as the criteria for selecting candidate mutations.

The initial single-point mutation library was built as follows: 1) All single-point mutations in HCDRs were enumerated and scored; 2) all double-point mutations from the top single-point mutations were enumerated; and 3) all single and double mutations were ranked, and the top mutations on three CDR-H loops were selected as the initial mutations, including T30P, T31W, Y32O, N52D, N53F, N55L, N57L, G100P, R103M, L104F, and Q105F.

Cell Lines. HEK293T cells (ATCC), HeLa cells expressing hACE2 orthologs (kindly provided by Qiang Ding, Center for Infectious Disease Research, School of Medicine, Tsinghua University, Beijing, China) were maintained at 37°C in 5% CO₂ in Dulbecco's minimal essential medium containing 10% (vol/vol) heat-inactivated fetal bovine serum and 100 U/mL of penicillin-streptomycin. FreeStyle 293F cells (Thermo Fisher Scientific, R79007) were maintained at 37°C in 5% CO₂.

Antibody and Fab Production. Antibodies and Fab production were conducted as previously described (33, 52). Reference REGN10987 were synthesized according to the sequences released in the Protein Data Bank (44, 53–56). Antibodies were produced by transient transfection of HEK 293F cells (Life Technologies) using plasmids expressing the heavy and light chain by 1 mg/mL polyetherimide (Sigma). After 5 d, the supernatant was collected and purified with protein A. Antibodies were eluted from protein A with elution buffer (0.3 M glycine, pH 2.0), followed by dialyzing into phosphate-buffered saline (PBS).

Production of Pseudotyped SARS-CoV-2. The WT pseudotyped virus used throughout the analysis was the prototype strain (GenBank: MN908947.3) with a D614G mutation (WT D614G). Variants were extracted from the Global Initiative on Sharing All Influenza Data (GISAID) website. The Alpha variant (GISAID: EPI_ISL_601443) was constructed with a total of nine mutations (69-70del, 144del, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H). The Beta variant (GISAID: EPI_ISL_700450) was constructed with 10 mutations: L18F, D80A, D215G, 242-244del, S305T, K417N, E484K, N501Y, D614G, and

A701V. The Gamma variant (GISAID: EPI_ISL_792681) was constructed with 12 mutations: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, and V1176F. The Delta variant (GISAID: EPI_ISL_1534938) was constructed with 10 mutations: T19R, G142D, 156-157del, R158G, A222V, L452R, T478K, D614G, P681R, and D950N. The variant Delta Plus (GISAID: EPI_ISL_3019629) was constructed with 11 mutations: 19R, G142D, 156-157del, R158G, A222V, K417N, L452R, T478K, D614G, P681R, and D950N. The Kappa variant (GISAID: EPI_ISL_1384866) was constructed with eight mutations: T95I, G142D, E154L, L452R, E484Q, D614G, P681R, and N1071H. The Eta variant (GISAID: EPI_ISL_2885901) was constructed with eight mutations: Q52R, A67V, 69-70del, 144del, E484K, D614G, Q677H, and F888L. The Epsilon variant (GISAID: EPI_ISL_2922315) was constructed with four mutations: S13I, W152C, L452R, and D614G. The single mutation N439K was introduced into the pcDNA3.1 vector encoding WT D614G using QuikChange site-directed mutagenesis (Agilent 210519). The gene of variants were synthesized in Genewiz, Inc. WT, and mutated SARS-CoV-2 pseudotyped viruses were generated by cotransfection of human immunodeficiency virus backbones expressing firefly luciferase (pNL43R-E-luciferase) and pcDNA3.1 (Invitrogen) expression vectors encoding the respective spike proteins into 293T cells (ATCC) (20). Viral supernatants were collected 48 h later. Viral titers were measured as the luciferase activity in relative light units (Bright-Glo Luciferase Assay Vector System; Promega Biosciences).

Neutralization Activity of Monoclonal Antibodies against Pseudotyped SARS-CoV-2. Neutralization assays were conducted as previously described (33). Serial dilutions of monoclonal antibodies (mAbs) were prepared with the highest concentration of 5 or 50 $\mu\text{g/mL}$. WT or mutated spike pseudotyped viruses were incubated with mAbs at 37 °C for 1 h. HeLa-hACE2 cells (1.5×10^4 per well) were then added into the mixture and incubated at 37 °C for 48 h to

60 h. Then the luciferase activity was measured after cell lysis. The percent of neutralization was determined by comparing with the virus control.

Antibody Binding Kinetics Measured by SPR. The binding kinetics of mAbs to SARS-CoV-2 WT or Delta RBD monomer were analyzed using SPR (Biacore 8K; GE Healthcare). Specifically, recombinant protein A (Sino Biological) was immobilized to a CM5 sensor chip. The mAbs (2 $\mu\text{g/mL}$) were captured by recombinant protein A, and then serial dilutions of SARS-CoV-2 WT/Delta RBD with highest concentration of 200 nM were run at a flow rate of 30 $\mu\text{L/min}$ in PBST buffer (1 $\times$ PBS and 0.05% [vol/vol] Tween-20). The resulting data were fitted to a 1:1 binding model using the Biacore 8K Evaluation software (GE Healthcare).

Quantification and Statistical Analysis. The technical and independent experiment replicates are indicated in the figure legends. IC_{50} of mAbs were calculated by the four-parameter dose inhibition equation in using Graphpad Prism 9.0.

Data Availability. All study data are included in the article and/or *SI Appendix*. The code can be found in GitHub at <https://github.com/HeliXonProtein/binding-ddg-predictor>.

ACKNOWLEDGMENTS. This research was supported by grants from National Key Plan for Scientific Research and Development of China (2020YFC0848800 and 2020YFC0849900), National Natural Science Foundation (81530065, 91442127, and 32000661), Beijing Municipal Science and Technology Commission (D17110000517 and Z201100005420019), and Beijing Advanced Innovation Center for Structural Biology, Tsinghua University Scientific Research Program (20201080053 and 2020Z99CFG004).

- P. Zhou et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- M. Hoffmann et al., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
- Q. Wang et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **181**, 894–904 (2020).
- R. Yan et al., Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
- J. Shang et al., Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
- J. Lan et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
- M. A. Tortorici, D. Veeler, Structural insights into coronavirus entry. *Adv. Virus Res.* **105**, 93–116 (2019).
- A. C. Walls et al., Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281–292.e6 (2020).
- R. de Assis et al., Analysis of SARS-CoV-2 antibodies in COVID-19 convalescent blood using a coronavirus antigen microarray. *Nat. Commun.* **12**, 1–9 (2021).
- J. Ge et al., Antibody neutralization of SARS-CoV-2 through ACE2 receptor mimicry. *Nat. Commun.* **12**, 250 (2021).
- A. Z. Wec et al., Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* **369**, 731–736 (2020).
- E. Seydoux et al., Analysis of a SARS-CoV-2-infected individual reveals development of potent neutralizing antibodies with limited somatic mutation. *Immunity* **53**, 98–105.e5 (2020).
- C. Wang et al., A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **11**, 225 (2020).
- B. Ju et al., Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115–119 (2020).
- D. M. Weinreich et al., Trial Investigators, REGN-COV2, a neutralizing antibody cocktail, in outpatients with Covid-19. *N. Engl. J. Med.* **384**, 238–251 (2021).
- M. Dougan et al., BLAZE-1 Investigators, Bamlanivimab plus Etesevimab in mild or moderate Covid-19. *N. Engl. J. Med.* **385**, 1382–1392 (2021).
- A. Gupta et al., COMET-ICE Investigators, Early treatment for Covid-19 with SARS-CoV-2 neutralizing antibody Sotrovimab. *N. Engl. J. Med.* **385**, 1941–1950 (2021).
- D. Corti, L. A. Purcell, G. Snell, D. Veeler, Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell* **184**, 4593–4595 (2021).
- W. T. Harvey et al., COVID-19 Genomics UK (COG-UK) Consortium, SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
- R. Wang et al., Analysis of SARS-CoV-2 variant mutations reveals neutralization escape mechanisms and the ability to use ACE2 receptors from additional species. *Immunity* **54**, 1611–1621 (2021).
- P. Wang et al., Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).
- B. Hie, E. D. Zhong, B. Berger, B. Bryson, Learning the language of viral evolution and escape. *Science* **371**, 284–288 (2021).
- D. Planas et al., Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021).
- J. Zhang et al., Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science* **374**, 1353–1360 (2021).
- W. F. Garcia-Beltran et al., Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383 (2021).
- J. Prévost, A. Finzi, The great escape? SARS-CoV-2 variants evading neutralizing responses. *Cell Host Microbe* **29**, 322–324 (2021).
- H. Gram et al., In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3576–3580 (1992).
- M. J. Hammerling et al., In vitro ribosome synthesis and evolution through ribosome display. *Nat. Commun.* **11**, 1108 (2020).
- J. R. Jackson, G. Sathe, M. Rosenberg, R. Sweet, In vitro antibody maturation. Improvement of a high affinity, neutralizing antibody against IL-1 beta. *J. Immunol.* **154**, 3310–3319 (1995).
- M. Ho, I. Pastan, In vitro antibody affinity maturation targeting germline hotspots. *Methods Mol. Biol.* **525**, 293–308, (2009).
- B. Li et al., In vitro affinity maturation of a natural human antibody overcomes a barrier to in vivo affinity maturation. *MAbs* **6**, 437–445 (2014).
- J. Wang et al., In vitro affinity maturation of antibody against membrane-bound GPCR molecules. *Appl. Microbiol. Biotechnol.* **103**, 7703–7717 (2019).
- S. Shan et al., A potent and protective human neutralizing antibody against SARS-CoV-2 variants. *Front. Immunol.* **12**, 766821 (2021).
- J. Jumper et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- A. Vaswani et al., “Attention is all you need” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, U. von Luxburg, I. Guyon, Eds. (Curran Associates Inc., Long Beach, CA, 2017), pp 6000–6010.
- R. Rao et al., MSA Transformer. *bioRxiv* [Preprint] (2021). <https://doi.org/10.1101/2021.02.12.430858> (Accessed 15 February 2022).
- R. Rao, J. Meier, T. Sercu, S. Ovchinnikov, A. Rives, Transformer protein language models are unsupervised structure learners. *bioRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.12.15.422761> (Accessed 15 February 2022).
- J. Jankauskaite, B. Jiménez-García, J. Dapkunas, J. Fernández-Reco, I. H. Moal, SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **35**, 462–469 (2019).
- P. Xiong, C. Zhang, W. Zheng, Y. Zhang, BindProfX: Assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. *J. Mol. Biol.* **429**, 426–434 (2017).
- N. Zhang et al., MutaBind2: Predicting the impacts of single and multiple mutations on protein-protein interactions. *iScience* **23**, 100939 (2020).
- X. Liu, Y. Luo, P. Li, S. Song, J. Peng, Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. *PLoS Comput. Biol.* **17**, e1009284 (2021).
- J. K. Leman et al., Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
- H. Park et al., Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
- J. Hansen et al., Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**, 1010–1014 (2020).

45. L. Liu *et al.*, Striking antibody evasion manifested by the Omicron Variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022).
46. D. Planas *et al.*, Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature* **602**, 671–675 (2022).
47. E. Cameroni *et al.*, Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664–670 (2022).
48. S. Cele *et al.*; NGS-SA; COMMIT-KZN Team, Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*, 10.1038/s41586-021-04387-1 (2021).
49. D. Hu *et al.*, Effective optimization of antibody affinity by phage display integrated with high-throughput DNA synthesis and sequencing technologies. *PLoS One* **10**, e0129125 (2015).
50. M. Wang, Z. Cang, G. W. Wei, A topology-based network tree for the prediction of protein-protein binding affinity changes following mutation. *Nat. Mach. Intell.* **2**, 116–123 (2020).
51. J. Schymkowitz *et al.*, The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
52. L. Jiang *et al.*, Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. *Sci. Transl. Med.* **6**, 234ra59 (2014).
53. X. Chi *et al.*, A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020).
54. R. Shi *et al.*, A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).
55. Y. Cao *et al.*, Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84 (2020).
56. J. Audet *et al.*, Molecular characterization of the monoclonal antibodies composing ZMAb: A protective cocktail against Ebola virus. *Sci. Rep.* **4**, 6881 (2014).