Design of immunogens for eliciting antibody responses that may protect against SARS-CoV-2 variants

Eric Wang[1] **and Arup K. Chakraborty**[1,2,3,4,5*]

***For correspondence:**
arupc@mit.edu (AC)

[1]Institute for Medical Engineering and Science, MIT, Cambridge, MA 02139; [2]Department of Chemical Engineering, MIT, Cambridge, MA 02139; [3]Department of Physics, MIT, Cambridge, MA 02139; [4]Department of Chemistry, MIT, Cambridge, MA 02139; [5]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139

Abstract The rise of SARS-CoV-2 variants and the history of outbreaks caused by zoonotic coronaviruses point to the need for next-generation vaccines that confer protection against variant strains. Here, we combined analyses of diverse sequences and structures of coronavirus spikes with data from deep mutational scanning to design SARS-CoV-2 variant antigens containing the most significant mutations that may emerge. We trained a neural network to predict RBD expression and ACE2 binding from sequence, which allowed us to determine that these antigens are stable and bind to ACE2. Thus, they represent viable variants. We then used a computational model of affinity maturation (AM) to study the antibody response to immunization with different combinations of the designed antigens. The results suggest that immunization with a cocktail of the antigens is likely to promote evolution of higher titers of antibodies that target SARS-CoV-2 variants than immunization or infection with the wildtype virus alone. Finally, our analysis of 12 coronaviruses from different genera identified the S2′ cleavage site and fusion peptide as potential pan-coronavirus vaccine targets.

1 Introduction

Vaccines that protect against severe acute respiratory distress coronavirus 2 (SARS-CoV-2) are highly effective. However, variants that impact vaccine efficacy are emerging. For example, the Beta (B.1.351) variant significantly reduces neutralizing antibody titers (***Choi et al.*** (***2021***); ***Bates et al.*** (***2021***)), and the Delta (B.1.617.2) variant both increases transmission and reduces neutralizing antibody titers (***Li et al.*** (***2021***); ***Liu et al.*** (***2021a***); ***Edara et al.*** (***2021b***); ***Planas et al.*** (***2021b***); ***Choi et al.*** (***2021***)). Although current vaccines still remain effective at preventing severe illness upon infection with these variants (***Abu-Raddad et al.*** (***2021***)), their appearance signals that others more capable at evading antibodies elicited by the current vaccines may emerge. Indeed, available data on the Omicron variant suggest that this is true (***Schmidt et al.*** (***2021***); ***Planas et al.*** (***2021a***)). Moreover, other pathogenic coronaviruses could evolve in the future due to zoonosis. Thus, vaccines that protect against potential SARS-CoV-2 variants and those that protect against potential zoonotic coronaviruses would be beneficial for curtailing future outbreaks.

Neutralizing antibodies isolated from convalescent COVID-19 donors commonly target the spike's receptor-binding domain (RBD) (***Robbiani et al.*** (***2020***); ***Rogers et al.*** (***2020***); ***Zost et al.*** (***2020***)), which is located in the S1 domain and is responsible for facilitating viral entry by binding to angiotensin-converting enzyme 2 (ACE2) (***Zamorano Cuervo and Grandvaux*** (***2020***)). Following ACE2 binding,

the spike is cleaved at the S1/S2 site by furin or transmembrane protease serine 2 (TMPRSS2) (***Örd et al.*** (***2020***); ***Hoffmann et al.*** (***2020***)), which separates the S1 and S2 domains. Further cleavage at the S2′ site rearranges S2 conformation and releases the fusion peptide, which then binds to the host membrane and facilitates fusion between the host and viral membranes. This allows the viral genome to enter the host cell's cytosol. Alternatively, following ACE2 binding, the virus may be endocytosed and S1/S2 cleavage occurs through endosomal cathepsins (***Zhao et al.*** (***2021***)).

In response to antigen (whether from vaccination or natural infection), antibodies are generated by a Darwinian process known as affinity maturation (AM) that occurs in secondary lymphoid organs (***Victora and Nussenzweig*** (***2012***)). Activated germline B cells seed structures known as germinal centers (GC), where they undergo multiple cycles of expansion, mutation, and selection based on the binding affinity of their B cell receptor (BCR) to antigen. Through this process, B cells increase their binding affinity to antigen up to 1000-fold or more. B cells can differentiate into plasma cells that secrete antibodies, which are soluble and modified forms of the BCR that inhibit pathogens through neutralization or various other effector functions (***van Erp et al.*** (***2019***)).

Previous computational models of AM have focused on the response to single antigens (***Kepler and Perelson*** (***1993***); ***Oprea*** (***1997***); ***Brink*** (***2007***); ***Chan et al.*** (***2013***); ***Meyer-Hermann et al.*** (***2012***); ***Celada and Seiden*** (***1996***)) or the development of broadly neutralizing antibodies (bnAbs) for influenza and human immunodeficiency virus (HIV) upon immunization with variant antigens (***Wang et al.*** (***2015***); ***Sprenger et al.*** (***2020***); ***Shaffer et al.*** (***2016***); ***Amitai et al.*** (***2020***); ***Chaudhury et al.*** (***2014***); ***Childs et al.*** (***2015***); ***Luo and Perelson*** (***2015***); ***Ganti and Chakraborty*** (***2021***); ***Sachdeva et al.*** (***2020***)). The latter studies aimed to study strategies for induction of bnAbs that target the conserved residues of the viral spike. They have identified a number of factors that influence bnAb development, such as the number of variant antigens and the mutational distances between them, antigen concentration, antigen geometry, and the temporal pattern of antigen administration.

However, given that SARS-CoV-2 appears to mutate more slowly than influenza and HIV ((***van Dorp et al.*** (***2020***); ***Zhao et al.*** (***2004***))), a different strategy is feasible for protection against its variants. Instead of targeting strictly conserved regions, an immunization scheme that generates an appropriate polyclonal response against variable regions could protect against mutant strains. Antibodies targeting the variable class 1 and class 2 RBD epitopes of SARS-CoV-2 tend to be neutralizing, while antibodies that target the conserved class 4 epitope can be non-neutralizing because of their inability to compete with ACE2 binding (***Jette et al.*** (***2021***)). Thus, targeting variable regions may also be more likely to generate neutralizing antibodies against SARS-CoV-2 variants.

At least two significant questions need to be answered in order to design such a vaccine that can protect against SARS-CoV-2 variants,

1. Which antigens should be used, and in what combination/order should they be administered, to optimally produce the desired antibody response?
2. How does previous infection or immunization with SARS-CoV-2 affect antibody evolution upon administering the chosen antigens?

In this study, we aimed to address these questions pertinent to vaccination that protects against SARS-CoV-2 variants and beyond. We first developed a method of calculating conservation at different sites of the spike protein by analyzing both structural and sequence data. By applying it to sarbecovirus spike proteins and combining the results with previous deep mutational scanning results, we designed 6 antigens that may protect against the most significant RBD escape mutations in SARS-CoV-2. After confirming that these antigens are stable and bind ACE2 using neural networks trained to predict RBD expression and ACE2 binding from sequence, we use a computational model of AM to study the antibody response to these antigens. Thus, we identified an immunization scheme which may produce higher titers against current and potential variants than immunization with the WT antigen alone. But, such a vaccination scheme produces lower anti-WT titers than the WT vaccine does, illustrating the fact that a strain-specific vaccine is generally the

most effective at protecting against a particular strain. However, a vaccine that adequately protects a population when diverse new variants emerge can serve as a shield and provide time for a strain-specific vaccine to be developed. Our results also highlight factors that impact titers such as previous exposure to WT antigen and the fraction of GC-seeding cells that are memory cells generated during past exposure. Finally, by applying the conservation analysis to 12 diverse coronaviruses, we identify the S2' cleavage site and fusion peptide as conserved regions, pointing to a potential target for a pan-coronavirus vaccine.

2 Methods

2.1 Spike conservation

In order to identify conserved and variable residues in coronavirus spike proteins, we leveraged both sequence and structural data. Structural data can be used to determine structural conservation (as described below) in a way that does not strongly depend on the number of insertions or deletions. In our method, each residue of the spike was assigned a conservation fraction ranging from 0 to 1, which was the average of the structural conservation fraction and the biochemical conservation fraction. An overview of the method that we use to determine these conservation scores is provided in Figure 1.

2.1.1 Structural conservation fraction

Using the structures for the trimeric spike of various coronaviruses (Table S1), atoms were removed from each trimer so that only a monomer remained. Thus, the structural conservation fraction was calculated using a monomer rather than the full trimer, but the monomer retained the conformation that it had in the trimer. The STRAL-SV server (***Zemla et al.*** (***2011***)) was used to align these monomeric spike protein structures. The STRAL-SV server addresses the fact that multi-domain proteins can adopt multiple conformations, such as the spike protein's RBD being oriented down in one structure and up in another structure, which would have confounded a standard global alignment. In STRAL-SV, a sliding-window approach divided the monomer structures into overlapping fragments of 90 amino acids, and structural alignments were performed between the fragments. This process identified residues across different coronavirus spikes that structurally overlap in the alignment. This data was used to construct a structure-based sequence alignment. A table is first constructed (see Figure 1) where the residues of the spike proteins are the rows, the column names correspond to coronaviruses (such as MHV, HKU1, and HKU2), and the entries of the table are the 20 amino acids or a placeholder entry of "." The placeholder entry of "." means that the coronavirus spike protein was not able to align at that residue.

The structural conservation fraction was calculated from the structure-based sequence alignment using a particular coronavirus as a reference. For each residue, the structural conservation fraction was the fraction of other coronaviruses that are able to structurally align at that residue. We illustrate this with an example from Figure 1's table, assuming for simplicity that the first row corresponds to residue 1. If we select residue 1 and MHV as a reference coronavirus, the other coronaviruses are HKU1 and HKU2. Only HKU1 is able to structurally align, so the structural conservation fraction is 0.5. It is also possible that the reference coronavirus did not align at the chosen residue, in which case the structural conservation fraction is 0. For example, consider residue 1 using HKU2 as a reference coronavirus. Since the entry is ".", then the structural conservation fraction is 0. Note that Figure 1's table is truncated and is shown for illustration purposes only.

2.1.2 Biochemical conservation fraction

For each coronavirus, a set of several hundred spike sequences were gathered (a complete list of sequences is available at https://github.com/ericzwang/sars2-vaccine/tree/main/data/aligned-cov-sequences.gz) and aligned to the PDB structure's sequence using the ClustalW algorithm in MEGAX (***Thompson et al.*** (***1994***); ***Kumar et al.*** (***2018***)). All coronavirus alignments were then concatenated together into a single multiple sequence alignment as illustrated in Figure 1 figure supplement 1.

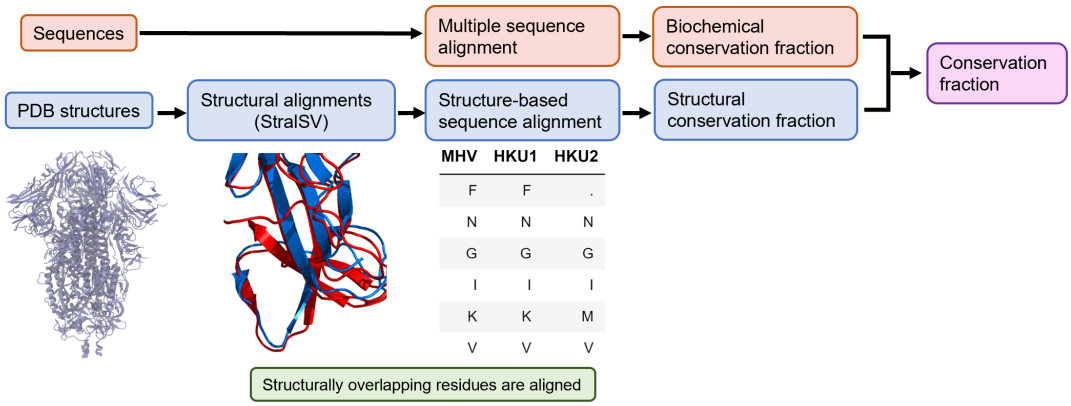

Figure 1. Overview of conservation analysis workflow. Schematic of conservation fraction calculation. Structure analysis (blue) and sequence analysis (red) are independently carried out and averaged to obtain the final conservation fraction.

Figure 1–Figure supplement 1. Illustration of the concatenation of individual coronavirus alignments into a single multiple sequence alignment.

The biochemical conservation fraction was calculated from the multiple sequence alignment using a particular coronavirus as a reference. For each residue, the biochemical conservation fraction was the fraction of sequences that have an amino acid in the same class (hydrophobic, polar, positive, or negative) as the reference coronavirus's consensus (most common) amino acid. Thus, this calculation tolerates mutations if they are biochemically similar.

To illustrate this procedure using Figure 1 figure supplement 1, assuming that the first row corresponds to residue 1 for simplicity, suppose we wish to calculate the biochemical conservation fraction of residue 1 using MHV as a reference coronavirus. The consensus amino acid from the MHV alignment is F, which is a hydrophobic amino acid. The biochemical conservation fraction is the fraction of amino acids in the first row of the concatenated alignment that are also hydrophobic.

2.1.3 Conservation fraction

The conservation fraction was the average of the structural conservation fraction and the biochemical conservation fraction. For the design of antigens that may protect against SARS-CoV-2 variants, only SARS and SARS-CoV-2 data were considered. Since SARS and SARS-CoV-2 are closely related, the threshold conservation fraction below which a residue was identified as variable was chosen to be 0.99.

For pan-coronavirus vaccine target identification, we considered 12 coronaviruses from different genera (Table S1), which constituted an analysis of 12 spike protein structures and ~4000 spike protein sequences. We set a threshold conservation fraction of 0.8, which was chosen because it was the largest value at which we observed a conserved region comparable in size to an antibody footprint.

2.2 Design of cocktail antigens for protection against SARS-CoV-2 variants

We hypothesized that a cocktail composed of variant SARS-CoV-2 spike sequences can generate a polyclonal response that protects against variants of SARS-CoV-2. The aim is to generate a polyclonal response that is composed of strain-specific antibodies that together can protect against multiple variants. We do not aim to elicit bnAbs that target a particular conserved epitope, as in studies being done in the context of universal influenza and HIV vaccines (***Amitai et al.*** (***2020***); ***Wang et al.*** (***2015***); ***Shaffer et al.*** (***2016***); ***Sprenger et al.*** (***2020***); ***Nachbagauer et al.*** (***2021a***); ***Steichen et al.*** (***2016***); ***Pauthner et al.*** (***2019***); ***Escolano et al.*** (***2016***); ***Mascola and Haynes*** (***2013***); ***Huang et al.*** (***2020***)). To design antigens that can elicit the desired polyclonal response, we first generated a list

of mutations (with respect to the wildtype Wuhan SARS-CoV-2 sequence) that satisfied the following criteria:

1. The mutation was in a residue determined to be variable using the calculation outlined in Section 2.1. The structure and sequence data were restricted to SARS-CoV-2 and SARS, and the conservation fraction threshold was set to 0.99. SARS was included because a calculation exclusively using the structures of SARS-CoV-2 variants and ∼300,000 SARS-CoV-2 sequences classified nearly the entire spike protein as conserved (Figure 7 figure supplement 1). In other words, only a few SARS-CoV-2 mutations have occurred in current SARS-CoV-2 variants, so we consider residues that are different in SARS as such mutations have the potential to emerge because these viruses are closely related. The Omicron variant emerged after our analyses were completed, but its characteristics support our use of SARS data. We considered the 15 RBD mutations in the Omicron variant, and found that 12 out of 15 Omicron RBD mutations are classified as variable residues according to our calculation. Additionally, all RBD mutations in previous variants (Alpha, Beta, Delta, Gamma) are considered variable. Therefore, while few SARS-CoV-2 mutations have already emerged, our choice of including SARS data appears to be predictive of mutations that may emerge in the future.
2. The mutation did not significantly decrease RBD stability. This was determined from deep mutational scanning experiments carried out by Starr et al. (***Starr et al.*** (***2020***)) that measured the effect of nearly every single RBD mutation on protein expression, a correlate of protein stability. A threshold on the change in expression from the WT (expressed as $\log_{10}$(MFI)) was set to -0.2, which was chosen such that mutations in circulating variants would be considered viable. At this stage, multiple-mutation effects were not considered, but they were considered when the final antigens were evaluated using the expression-prediction neural network we developed (described in Section 2.3).
3. The mutation did not significantly decrease ACE2 binding. This was determined from the same deep mutational scanning experiments as in Criterion 2 (***Starr et al.*** (***2020***)), which also measured the effects of single mutations on ACE2 binding. The allowed threshold value of the change in affinity from the WT (expressed as $\log_{10}(K_{D,app})$ where $K_{D,app}$ was in units of M) was set to -0.2, which was chosen such that mutations in circulating variants would be considered viable. As for protein expression, multiple-mutation effects on ACE2 binding were evaluated later using the neural network described in Section 2.3.
4. The mutation significantly abrogated class 1 or class 2 antibody binding. These mutations were found from various deep mutational scanning experiments over 20 class 1/2 antibodies (***Greaney et al.*** (***2021a***,b); ***Starr et al.*** (***2021b***,a,c); ***Dong et al.*** (***2021***)) (Table S2), in which a mutation's ability to abrogate binding was quantified as an escape fraction. In the experiments, many cells are generated which express different RBD sequences, and there are multiple cells that express the same RBD. The cells are sorted into an antibody-escape bin based on their inability to bind a fluorescently tagged antibody. The escape fraction is the fraction of cells expressing a particular RBD that are in the antibody-escape bin. An escape fraction of 0 indicates that no cells expressing a particular RBD are in the antibody-escape bin, and a fraction of 1 indicates that all cells expressing that RBD are in the bin. The 34 mutations with the largest mean escape fractions across all antibodies were selected (see SI for explanation for choosing 34 mutations). Among these 34 mutations, some but not all were also present in circulating strains. The absence of some escape mutations was due to the low mutation rate of SARS-CoV-2 and the boost in transmissibility provided by certain non-escape mutations. An example of a prevalent non-escape mutation is N501Y, which first appeared in the Alpha variant and increased ACE2 binding affinity without significantly affecting antibody binding. Although some of the 34 escape mutations have not arisen yet, they may arise in the future as natural and vaccine-induced responses impose selection pressures. So, they were included in our antigens in order to protect against future variants. Indeed, Q493R, one of the most sig-

nificant class 1/2 escape mutations in the Omicron variant, is found among these 34 escape mutations and has not appeared in previous variants of concern. Note again that Omicron emerged after we completed our analyses.

The procedure outlined above resulted in a set of mutations that were in variable residues and would likely abrogate binding to neutralizing class 1 and 2 antibodies circulating in vaccinated and naturally infected persons, and did not diminish ACE2 binding or decrease spike stability and so would likely be viable viruses. This list of mutations are candidates for inclusion in antigens that may elicit a broadly protective polyclonal antibody response.

Additionally, a survey of ~300,000 SARS-CoV-2 sequences from the GISAID database (sequences available at https://github.com/ericzwang/sars2-vaccine/tree/main/data/gisaid_RBD_sequences.tar.gz) revealed several mutations prevalent in circulating strains. Among these, 3 mutations (K417T, K417N, and T478K) were the most prevalent mutations that also escape antibodies, but they were not among the 34 mutations with the largest escape fractions from deep mutational scanning. Nonetheless, these 3 mutations were also considered in order to protect against current circulating variants. In total, the list contained 37 mutations split among 10 residues.

Multiple mutations from the list generated using the procedure described above were spatially close together on the RBD. Multiple such mutations on the same RBD was undesirable as it could lead to coupled effects that affect ACE2 binding or RBD stability. Although such effects were later accounted for using the neural networks described in Section 2.3, the data used to train those networks mostly contained RBDs with spatially separated mutations. So, we grouped the mutated residues such that the residues within a group would be maximally separated in space on the RBD. To accomplish this, we divided the 10 mutated residues into all possible combinations of 2 groups of 5 residues, and our goal was to select the combination with the maximum separation between residues. This was accomplished as follows:

1. For a particular combination of two groups, two numbers, V_1 and V_2, were calculated. To determine V_1, consider the 5 residues in the first group positioned on the structure of the RBD. The centers of mass of the 5 residues formed a polygon, and the volume of that polygon was V_1. Similarly, the volume of the polygon formed by the residue centers of mass of the second group was V_2. The polygons were calculated using the ConvexHull utility in SciPy (***Virtanen et al.*** (***2020***)).
2. The combination was assigned a score, which was calculated as the minimum of V_1 and V_2 ($\min(V_1, V_2)$)
3. From all possible combinations, the combination with the largest value of $\min(V_1, V_2)$) was chosen. Selecting the score as $\min(V_1, V_2)$), instead of $V_1 + V_2$, avoided choosing cases with severe imbalance between the groups, which may have deleteriously impacted the stability of half of our antigens.

For each of the two groups thus selected, we generated 3 sequences with 5 mutations per sequence. Since there are 2 groups, there are 6 sequences total. For each residue, the amino acid mutations were chosen starting from those with the largest escape fractions (thus, those that escape antibodies most). However, the mutations were also chosen such that the amino acids were dissimilar, meaning that a potential mutation of the same class (hydrophobic, polar, positive, negative) as a previously selected mutation was only selected if no dissimilar mutations were available. For example, if the mutation with the largest escape fraction at residue 490 was F490K, then the first sequence used F490K. For the second sequence, the mutation with the second-largest escape fraction at residue 490 was F490R. But, F490R was skipped because R and K are biochemically similar amino acids. The second sequence then moved to the mutation with the third-largest escape fraction, which was F490E, and this mutation was used because E is not biochemically similar to K. This process repeated for the third sequence and then moved on to another residue. After all 5 residues in the first group were thus examined, the process was repeated for the 5 residues in the second group.

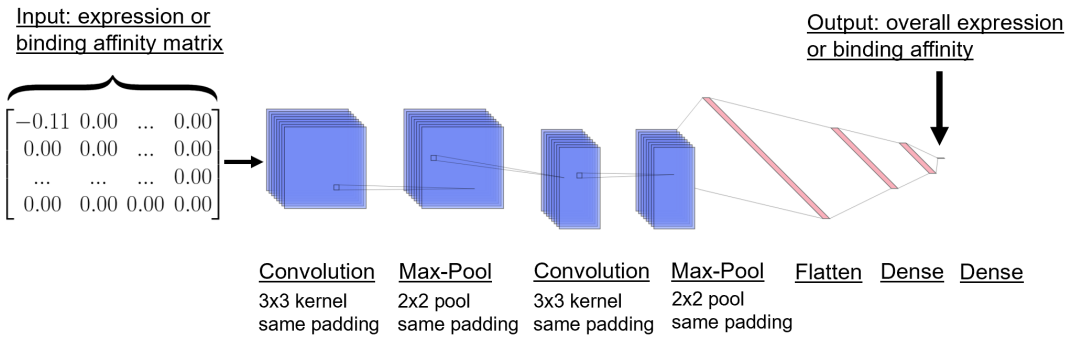

Figure 2. Overview of neural network architecture. Neural network architecture used to predict overall expression or ACE2 binding affinity. The input is either an expression or binding affinity matrix, which is the element-wise product of the one-hot encoded RBD sequence and the matrix of single-mutation changes. The output is the overall expression or binding affinity.

Since our antigen design method relied on deep mutational scanning experiments that only studied RBD substitutions, insertions or deletions in the RBD and mutations outside the RBD were not considered.

2.3 Neural network-based prediction of RBD expression and ACE2 binding affinity

The published data for selecting stable mutations was based on single mutation effects (***Starr et al.*** (***2020***)), but the designed sequences possessed 5 mutations each. Therefore, it was not known whether the designed sequences were still stable or maintained ACE2 binding. We noted that the deep mutational scanning experiments provided two datasets – one with measured changes in expression compared to WT for multi-mutant RBDs, and another with measured changes in ACE2 binding affinity compared to WT for multi-mutant RBDs. After removing sequences with no mutations or missing labels, the expression dataset contained ∼169,000 mutants with 1-7 mutations, and the binding affinity dataset contained ∼135,000 mutants with 1-7 mutations.

For the expression dataset, every sequence was one-hot encoded for the 20 amino acids, with unmutated residues represented by a zero vector. This one-hot encoding matrix was element-wise multiplied with a matrix of the single-mutation expression changes to yield a final matrix termed the expression matrix. In one approach, single-mutation effects were assumed to be additive, and the overall change in expression was the sum of the expression matrix. By training a neural network on the dataset, additivity of single-mutations effects did not need to be assumed. The network architecture is illustrated in Figure 2. In this approach, the expression matrix was passed into a 1D convolutional layer (32 filters, kernel size 3, same padding, and ReLu activation), a max pooling layer (pool size of 2, stride of 1, and same padding), a 1D convolutional layer (16 filters, kernel size 3, same padding, and ReLu activation), a max pooling layer (pool size of 2, stride of 1, and same padding), a flatten layer, a dense layer (64 neurons, ReLu activation), another dense layer (32 neurons, ReLu activation), and an output layer of 1 neuron.

We performed 10-fold cross validation with a 90% train, 5% validation, and 5% test split. Optimization used Adam, a learning rate of 0.001, a batch size of 100, a mean-squared error loss, up to 100 epochs, and early stopping on the validation set with a patience of 3. The model was implemented using Keras version 2.4.0 (***Gulli and Pal*** (***2017***)) with the TensorFlow version 2.3.1 backend (***Abadi et al.*** (***2015***)). We calculated the test-set Pearson correlation coefficient on the change in expression.

The same approach was used to train a separate network that predicted changes in ACE2 binding affinity. The only differences were that the binding affinity network was trained using the binding affinity dataset rather than the expression dataset, and the input matrix was the binding affinity matrix instead of the expression matrix. The neural networks were then used to predict the expres-

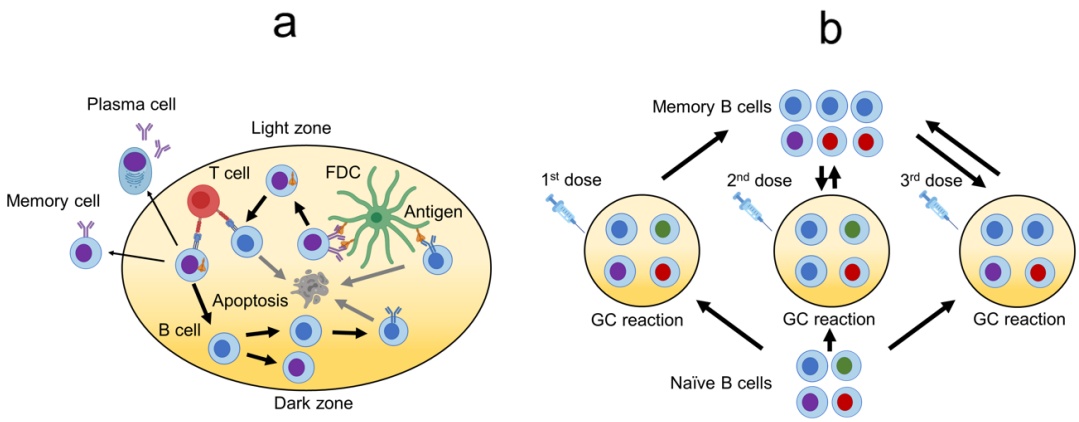

Figure 3. Overview of the affinity maturation model. (a) Schematic of AM within a GC. Activated B cells first undergo expansion and somatic hypermutation (SHM), and then they undergo selection based on their ability to bind and internalize antigen presented on FDCs and receive T cell help. A small fraction of positively selected B cells exits as antibody-secreting plasma cells and memory cells. The others undergo another round of mutation and selection. This cycle repeats until a termination condition is met. (b) Schematic of seeding new GCs upon new immunizations. For the first immunization, the seeding cells come solely from naive B cells. For subsequent immunizations, seeding cells come from both the naive and memory cell populations, and newly generated memory cells join the memory cell population.

sion and ACE2 binding properties of our designed antigens.

2.4 Modeling affinity maturation upon immunization with designed antigens

We simulated the processes that occur during AM using a coarse-grained stochastic model in order to study how antibodies develop in response to immunization with our designed antigens. An overview of the model is illustrated in Figure 3. This model builds on our past work modeling the affinity maturation response to variant antigens (***Sprenger et al.*** (***2020***); ***Wang et al.*** (***2015***); ***Shaffer et al.*** (***2016***)).

2.4.1 Binding free energy representation

The BCR paratope and antigen epitope were each represented by a vector of residues. Both the paratope and epitope were 50 residues long, a length which was chosen based on the number of residues in the class 1 and class 2 epitopes (***Jette et al.*** (***2021***); ***Barnes et al.*** (***2020***); ***Scheid et al.*** (***2021***); ***Liu et al.*** (***2021c***)). For antibodies that did not have the epitope residues reported, the residues were found from the RBD-antibody structure using the PISA server (***Krissinel and Henrick*** (***2007***); ***Krissinel*** (***2010***); ***Krissinel and Henrick*** (***2005***)). This procedure allows identification of RBD residues at the RBD-antibody interface with buried surface areas greater than 0 $Å^2$. In order to choose residues that were consistently part of the class 1 or class 2 epitope, only residues that were bound by 3 or more class 1 or class 2 antibodies were included. Using the conservation analysis described above, 9 of the 50 residues had conservation fractions above 0.99 and were assigned as conserved. The remaining 41 residues were assigned as variable.

For a seeding B cell, each residue, k, was associated with a number PAR(k), which was initially sampled from a uniform distribution between -0.18 and 0.90. After AM began, the value of PAR (k) changed due to mutations, and the bounds on these values were -1 and 1.5. The epitope residues were represented by a number EPI(k), with values of +1 for WT residues and negative values for mutated residues depending on how biochemically different the mutated residue is from the WT. For a mutation in the same class (such as hydrophobic to hydrophobic), the epitope residue was -1. For a mutation from hydrophobic to polar, polar to charged, or vice-versa, the epitope residue was -2. For a mutation from hydrophobic to charged or vice-versa, the epitope residue was -3. For a mutation from positively charged to negatively charged or vice-versa, the epitope residue was -4.

To illustrate the construction of the epitope vector, we will use the Delta variant, which is defined using the mutations L452R and T478K, as an example. L452R is a hydrophobic to positive mutation (corresponding to a value of -3), and T478K is a polar to positive mutation (corresponding to a value of -2). Therefore, one residue of the epitope vector will have a value of -3, another will have a value of -2, and the remaining residues will have values of +1. The order of the residues does not matter in our simple model.

The binding free energy was calculated according to

$$E = \sum_k \text{PAR}(k) * \text{EPI}(k) \tag{1}$$

where PAR(k) was the paratope residue at position k and EPI(k) was the epitope residue at position k. Essentially, the binding free energy is a dot product between the paratope and epitope vectors. Larger values of E corresponded to stronger binding. This coarse-grained representation of the binding free energy was not based on structural considerations and had limitations: (1) the contribution of each residue to the binding free energy was weighted equally; (2) the order of the residues did not matter; (3) only RBD substitutions could be considered; (4) biochemically similar mutations were treated equivalently. This simplified representation should be adequate for achieving our goal of obtaining qualitative results regarding the efficacy of different antigens and associated mechanistic insights.

2.4.2 B cell expansion and mutation in the dark zone

Upon initial immunization, 10 naive B cells seeded a new GC. Increasing the number of seeding cells does not change our qualitative results (Figure 4 figure supplement 1). The B cells then expanded without mutation or selection until the population reached 5120 cells. In AM, activation-induced cytidine deaminase (AID) introduces somatic hypermutations (SHMs) into the BCR. This was modeled by mutating the paratope residues at a rate of 0.14 per sequence per division (B cells divided twice per GC cycle). SHMs had a 0.5 probability of being lethal, 0.3 probability of being silent, and 0.2 probability of modifying the binding energy (***Berek and Milstein*** (***1987***)). For mutations that modified the binding free energy, a random residue on the paratope was chosen, and the change in binding free energy for that residue was sampled from a shifted log-normal distribution:

$$\Delta E = \epsilon - \exp(\mu + \sigma r) \text{ and } -\delta < \Delta E < \delta \tag{2}$$

where r was a standard normal random variable, ϵ was a shift parameter, μ was the mean of the log-normal distribution, σ was the standard deviation of the log-normal distribution, and δ limited the maximum ΔE. We used the same parameters as Sprenger et. al ($\epsilon = 3$, $\mu = 1.9$, $\sigma = 0.5$, $\delta = 1$), which were set to match experimental distributions of changes in binding free energies between proteins due to single-residue mutations (***Kumar*** (***2006***)) as well as GC dynamics after immunization with a single antigen (***Tas et al.*** (***2016***)).

2.4.3 Selection in the light zone

In the light zone of the GC, B cells underwent selection based on their binding affinity to antigen. First, each B cell had a probability of successfully internalizing antigen according to

$$P_{internalize} = \frac{\sum_i c_i e^{e_{scale}(E_i - E_{act})}}{1 + \sum_i c_i e^{e_{scale}(E_i - E_{act})}} \tag{3}$$

where i was a sum over all antigens that the B cell encounters, c_i was the concentration of antigen i, E_i was the binding free energy to antigen i, E_{act} was a reference free energy, and e_{scale} was a scaling parameter. We used the same parameters as Sprenger et. al ($e_{scale} = 0.08$, $E_{act} = 9k_bT$). For a particular immunization scheme, the concentration that produces the highest antibody titers was chosen. The sum in Eq. 3 did not apply for WT doses that only contain a single antigen. For cocktails, $P_{internalize}$ depended on whether the B cell encounters one antigen at a time or all antigens

on the FDC. If the B cell encountered one antigen at a time in each cycle, then a random antigen was chosen, and there was no summation. If the B cell encountered all antigens during each cycle, then the sum included all antigens i. In any given cycle, a B cell has a few chances to be positively selected. So, if the antigen concentration on FDCs is sufficiently high, then the sum over all antigens is likely to be more realistic.

B cells that did not internalize antigen underwent apoptosis and were removed from the population, while B cells that internalized antigen competed for T cell help. To model T cell selection if B cells encountered one antigen per cycle, the B cells were ranked according to their binding free energy to the encountered antigen in the last cycle, and the top $F_{cut,help}$ fraction were positively selected. If B cells instead encountered all antigens per cycle, then the binding free energy ranking was based on all B cell – antigen pairs. The top $F_{cut,help}$ fraction of pairs were selected, and each B cell survived with a probability equal to the frequency that it appeared in the top $F_{cut,help}$ pairs. For example, suppose there are 6 antigens and thus 6 B cell – antigen pairs for a particular B cell. If there are 5000 B cells, then there are 30,000 possible B cell – antigen pairs in total. If 4 of the B cell – antigen pairs are in the top $F_{cut,help} * 30000$ pairs, then that B cell survives with a probability of 4/6. Following Sprenger et. al, we set $F_{cut,help} = 0.7$.

2.4.4 Recycling, exit, GC reaction termination, and seeding a new GC

Positively selected B cells had a 0.7 probability of being recycled for subsequent cycles and a 0.3 probability of exiting the GC as either a plasma or memory cell. The GC reaction stopped if the B cell population exceeded the initial population of 5120 cells, if the number of cycles exceeded 250, or if all B cells died. The first condition is a proxy for the B cells internalizing all the antigen, the second is a proxy for antigen decay over time, and the third is GC extinction. Upon a second immunization, a new GC was seeded using 10 B cells, which were a mixture of memory and naive cells. The fraction of memory cells was 1 (all memory cells), except in Section 3.2.5 when we studied the effect of varying this fraction.

2.4.5 Titer calculation

The B cell population at the end of each GC reaction was analyzed for antibody titers. Identical B cells were grouped together as clones, and the binding free energy against an antigen sequence was calculated for each clone. The antibody titer was the total number of B cells that bind to the antigen with a free energy above $E_{th} = 17\ k_b T$. While the titers quantitatively depended on E_{th}, the rank order of titers for different immunization schemes that we studied did not. So, the qualitative results did not depend on E_{th} (Figure 4 figure supplement 2). The simulation was carried out $N = 200$ times, and the average antibody titer for antigen i was calculated as

$$\text{titer} = \frac{1}{N} \sum^{N} \sum_{\text{clone}_j} \text{size}(j) * \text{H}(E_j^i - E_{th})) \tag{4}$$

where size(j) was the number of B cells belonging to clone j and H represents the step function. Error bars were standard deviations across 10 sets of N simulations. The choice of N was determined such that different immunization schemes could be distinguished based on the standard deviation. The titers thus obtained may be considered to represent the titers generated on average in a typical vaccinated person. This model only considered titers upon completion of each GC reaction. In reality, the titers would be defined by the plasma cells exiting the GC secreting antibodies continuously during the GC reaction.

2.4.6 Mean panel titer calculation

To analyze the coverage provided by antibodies against multiple different antigens, the binding free energy of each B cell clone with a test panel of 100 antigens was calculated. Increasing the panel size to 1000 antigens did not change the results (Figure 4 figure supplement 3). The antibody

Table 1. RBD substitutions, changes in expression, and changes in ACE2 binding affinity (calculated using the neural networks) for designed antigens and circulating variants.

Antigen	RBD substitutions	Change in expression from WT	Change in ACE2 binding affinity from WT
Sequence 1	K417F V483R F486P F490Q S494K	-0.12±0.04	-0.28±0.07
Sequence 2	K417H V483K F486P F490E S494R	-0.09±0.05	-0.53±0.14
Sequence 3	K417Y V483E F486P F490K S494P	-0.07±0.05	-0.18±0.05
Sequence 4	L452R T478K E484N Q493K N501A	-0.06±0.03	+0.01±0.27
Sequence 5	L452W T478K E484R Q493R N501A	-0.13±0.03	-0.20±0.07
Sequence 6	L452E T478K E484K Q493R N501A	-0.11±0.03	-0.15±0.06
Alpha Variant	N501Y	-0.11±0.02	-0.18±0.06
Beta Variant	K417N E484K N501Y	-0.08±0.02	-0.36±0.08
Delta Variant	L452R T478K	-0.01±0.02	+0.21±0.23
Gamma Variant	K417T E484K N501Y	-0.06±0.04	-0.42±0.12

titer against each panel antigen was calculated, and the mean panel titer was the mean titer across all 100 panel antigens.

Mean panel titers for several antigen panels were calculated. In the most general panel, residues took values of either +1 or -1, and the probability that the residue is +1 was the conservation fraction. This panel produced almost always resulted in mean panel titers of 0 because the panel antigens had too many mutations at the same time, which are unlikely to emerge in a variant antigen. So, we instead considered panels in which the antigens are restricted to N_{mut} mutations and varied N_{mut}. We considered cases in which the N_{mut} mutations occur in any variable residue and in which the mutations only occur in the 10 residues that are also mutated in the designed antigens. We also considered panels in which mutated residues take on different values (-1, -2, -3, and -4).

3 Results

3.1 Designed antigens are stable and maintain ACE2 binding based on neural network predictions

Using our conservation analysis on SARS and SARS-CoV-2 spike proteins and applying the process described in Section 2.2, we obtain the final antigens shown in Table 1. When we trained a neural network on a dataset of ∼169,000 mutant RBDs with measured expressions, we obtained a correlation of 0.94±0.01 between measured and calculated values. If the overall change in expression is assumed to be the sum of the single mutation effects, the correlation is instead 0.79±0.01. A separate network trained on a dataset of ∼135,000 mutant RBDs to predict ACE2 binding affinity achieves a correlation of 0.97±0.01, compared to 0.78±0.01 for the additive assumption. Thus, while the additive assumption is correlated to the overall change in expression or binding, using network models offers a significant benefit.

Since these neural networks achieve high correlations on test sets, we compared the predicted expressions and binding affinities for our designed antigens to several circulating variants (Table 1). None of the designed antigens significantly decrease expression or ACE2 binding affinity more than the corresponding most deleterious circulating variants (Alpha for expression and Gamma for ACE2 binding). Since these circulating variants produce viable virions, we conclude that, upon immunization, our designed antigens would generate immune responses that were relevant for viable viral mutants that may emerge in future. Indeed, although our antigens were designed prior to the appearance of the Omicron variant, mutations in residues 417, 478, 484, 493, and 501 are shared between our designed antigens and the Omicron variant. Residue 493 is particularly notable because it was not mutated in previous variants and is one of the most significant escape residues for evading class 1 and 2 antibodies.

3.2 Modeling the affinity maturation response to antigens

We then used our computational model of AM to study how antibodies develop in response to our antigens as well as the WT spike protein. The goals were to identify vaccination schemes using our antigens that may be best at protecting against variants compared to vaccination with the WT spike alone, and to get mechanistic insights into the pertinent AM processes.

3.2.1 Reduced titers against SARS-CoV-2 variants following WT immunization

We simulate AM following one WT immunization, which models infection or one vaccine dose, and two WT immunizations, which models two vaccine doses or infection followed by one vaccine dose. In Figure 4a, the anti-WT titers are high for one immunization (blue bars in various panels) and even higher for two immunizations (orange bars in various panels). In contrast, the titers against variants are reduced depending on the mutational distance between the variant and WT. For example, following two WT immunizations, anti-Beta/Gamma titers are reduced ∼6-fold relative to anti-WT titers, anti-Delta titers are reduced ∼2.5-fold, and anti-Alpha titers are nearly equivalent to anti-WT titers. This is due to the mutational distances of these variants as the Beta and Gamma variants have 3 RBD substitutions, the Delta variant has 2 RBD substitutions, and the Alpha variant has 1 RBD substitution. This also recapitulates experimental data demonstrating that the Alpha variant minimally reduces neutralizing titers while Beta, Delta, and Gamma significantly reduce titers (***Choi et al.*** (***2021***); ***Bates et al.*** (***2021***); ***Planas et al.*** (***2021b***); ***Liu et al.*** (***2021b***); ***Edara et al.*** (***2021c***,a); ***Zhou et al.*** (***2021***); ***Wang et al.*** (***2021a***,b,c)). These studies report differences between anti-Beta and anti-Gamma titers due to the importance of specific mutations or mutations outside of the RBD, but this coarse-grained model is not structurally detailed enough to recapitulate those differences.

3.2.2 Titers are higher for immunization with a cocktail of designed antigens than WT and depend on the choice of antigens in the cocktail

We then used our model to identify an optimal immunization scheme with our 6 designed antigens. We simulated a single immunization with all 6 antigens (Seq1-6), two immunizations with all 6 antigens (Seq1-6 | Seq1-6), a single immunization with sequences 1-3 (Seq1-3) and an immunization of sequences 1-3 followed by an immunization with sequences 4-6 (Seq1-3 | Seq4-6). We first assume that the antigens are homogeneously distributed on the FDCs at high concentrations. So, a B cell encounters all antigens in each cycle. The titers are reported in Figure 4a using different colored bars.

For anti-WT titers, WT immunizations provide greater protection than cocktail immunizations, which is expected because the cocktail antigens are mutated from the WT. However, the boosted cocktails (Seq1-6 | Seq1-6 and Seq1-3 | Seq4-6) still retain moderately high titers against WT, indicating that they may still be protective. The immunizations that start with Seq1-3 (Seq1-3 and Seq1-3 | Seq4-6) have slightly higher anti-WT titers than the analogous immunization starting with Seq1-6 (Seq1-6 and Seq1-6 | Seq1-6). This is because the antigens in Seq1-6 are mutated across 10 residues while the antigens in Seq1-3 are mutated across 5 residues, so the antibodies resulting from immunization with Seq1-3 will have more residues that resemble the WT sequence (more positive in the model). Since more positive residues increases the binding free energy with WT antigen, immunization with Seq1-3 results in higher anti-WT titers. The Alpha variant is similar to the WT, so the same reasoning applies to anti-Alpha titers.

Anti-Seq1 titers are low following WT immunization because sequence 1 contains numerous mutations. Seq1-6 immunization produces lower anti-Seq1 titers than Seq1-3 immunization because sequences 2 and 3 contain mutations in the same residues as sequence 1 while sequences 4-6 contain mutations in different residues from sequence 1. However, immunization with Seq1-6 | Seq1-6 versus Seq1-3 | Seq4-6 are nearly equivalent in anti-Seq1 titers. This is because the anti-Seq1 antibodies provided by the boost with Seq1-6 are much greater than the anti-Seq1 antibodies provided by the boost with Seq4-6, which compensates for the lower anti-Seq1 antibodies generated by priming with Seq 1-6 compared to Seq 1-3. Anti-Seq2 and anti-Seq3 titers showed

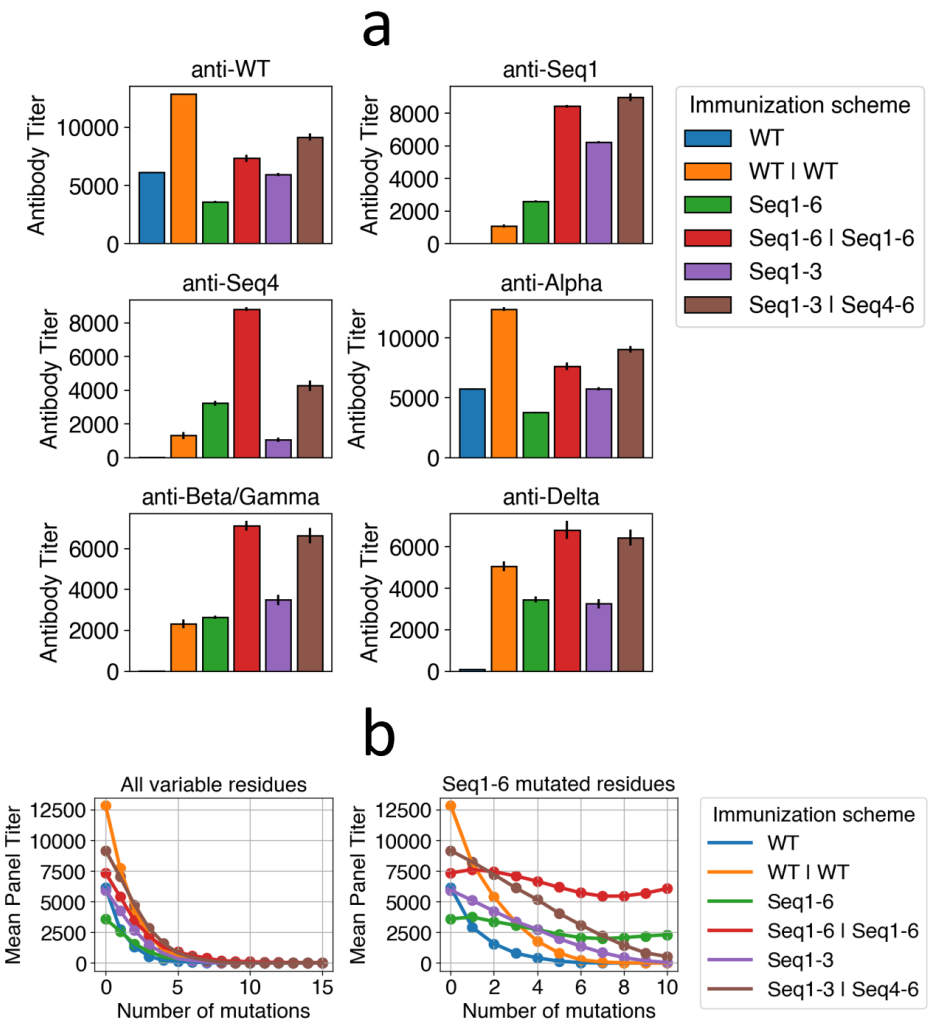

Figure 4. Titers against certain variants are low following WT immunization but high following cocktail immunizations with designed antigens. (a) Titers against WT and variants following different WT or cocktail immunization schemes. Sequences 2 and 3 are similar to sequence 1, so they are omitted for clarity. Sequences 5 and 6 are also omitted because they are similar to sequence 4. Beta and Gamma variants carry biochemically similar RBD mutations, so they are equivalent in this model. (b) Mean panel titers as a function of the number of mutations in panel antigens for different immunization schemes. Two types of panels are considered: one in which panel antigens can be mutated in any variable residue (All variable residues), and one in which panel antigen mutations can only occur in the 10 residues that are also mutated in sequences 1-6 (Seq1-6 mutated residues). It is possible for the panel antigens to have more mutations than in sequences 1-6, which have 5 mutations each. For example, for $n = 8$, every panel antigen has 8 mutations among the 10 residues that were mutated in the designed antigens.

Figure 4–Figure supplement 1. Mean panel titers using 10 and 20 seeding cells.

Figure 4–Figure supplement 2. Anti-WT titers for various energy thresholds.

Figure 4–Figure supplement 3. Mean panel titers for 100-antigen and 1000-antigen panels.

Figure 4–Figure supplement 4. GC collapse rate for Seq1-3 | Seq4-6 and Seq1-6 | Seq1-6.

Figure 4–Figure supplement 5. Mean panel titers for different values of mutated residues.

Figure 4–Figure supplement 6. Mean panel titers for different B cell-FDC encounters.

Figure 4–Figure supplement 7. Titers as a function of concentration for different B cell-FDC encounters.

similar results to anti-Seq1 titers, so they are omitted for brevity. Anti-Seq2 and anti-Seq3 titers are similar, even though sequence 2 and 3 have different mutations from sequence 1, because the mutations are in the same residues. Although our simulation modeled different mutations in the same residues, the resulting antibodies are relatively similar.

For anti-Seq4 titers, immunization with Seq1-6 produces higher titers than immunization with Seq1-3 because sequences 1-3 are all mutated in different residues from sequence 4. Immunization with Seq1-6 | Seq1-6 produces higher anti-Seq4 titers than immunization with Seq1-3 | Seq4-6, which is somewhat unexpected because one might expect that a boost with Seq4-6 would produce higher anti-Seq4 titers than a boost with Seq1-6. The origin of this result is that, because Seq 4-6 are very different from Seq 1-3, the memory B cells produced by priming with Seq 1-3 are poorly adapted to Seq 4-6. Therefore, the boost with Seq4-6 results in many memory B cells in the secondary GC dying (Figure 4 figure supplement 4), which reduces the anti-Seq4 titer from the boost. Similar to how anti-Seq2 and anti-Seq3 titers are similar to anti-Seq1 titers, anti-Seq5 and anti-Seq6 titers are similar to anti-Seq4 titers and are not shown for brevity.

For anti-Beta/Gamma and anti-Delta titers, cocktail immunization universally produces higher titers than WT immunization with the same number of immunizations (e.g. Seq1-6 is higher than WT and Seq1-6 | Seq1-6 is higher than WT | WT). The titers resulting from immunizations with different cocktails are roughly equivalent, such as the titers resulting from Seq1-6 | Seq1-6 versus Seq1-3 | Seq4-6.

Figure 4b shows the mean panel titer as a function of the number of mutations in the panel antigens from immunization with WT and cocktails. Mutated residues change from 1 to -4, which is chosen because the vaccine should ideally protect against emergent strains with biochemically dissimilar mutations. Using values such as -3, -2, or -1 instead of -4 does not change the qualitative results (Figure 4 figure supplement 5). If the panel mutations occur in any variable residue (Figure 4b, left panel), then the mean panel titer decreases at a similar rate for the cocktails as for the WT immunization. This is because there are few conserved residues, so the contribution of conserved regions to the binding free energy is not sufficient to generate bnAbs that can protect against strains bearing many mutations at other residues. This is consistent with experiments showing that few antibodies targeting class 1 or class 2 epitopes of the RBD are broadly protective against all single RBD substitutions (***Starr et al.*** (***2021a***)) let alone those bearing multiple such mutations.

We next studied whether immunization with our designed antigens can protect against panels of variants with mutations only in the residues that are also mutated in the designed antigens (Figure 4b, right panel). Note that the actual amino acids at the 10 residues can be different from those in our designed antigens. In this case, the mean panel titers do not significantly drop for immunization with Seq1-6 and Seq1-6 | Seq1-6. That is, vaccination using these schemes protects against variants with up to 10 mutations in particular variable residues. The mean panel titers for Seq1-3 | Seq4-6 do drop as the number of mutations increases, but less so than WT immunization. These results suggest that Seq1-6 | Seq1-6 may be an effective immunization scheme that could protect against current variants and some that may emerge in the future.

3.2.3 Titers do not depend on the number of variant antigens encountered in each B cell - FDC interaction

We also simulated the case in which antigens are heterogeneously distributed on the FDC, so a B cell encounters only one randomly chosen antigen in each cycle. In Figure 4 figure supplement 6, the mean panel titers are not significantly different from the all-antigen case. The main difference between the all-antigen and one-antigen case is that the optimal concentration is lower for the all-antigen case (Figure 4 figure supplement 7). An optimal concentration exists because low concentrations induce GC collapse while high concentrations are incapable of discriminating between low-affinity and high-affinity clones.

In the one-antigen case, a B cell that has high affinity to a particular antigen will likely survive a cycle if it encounters that antigen on the FDC, but it will likely die if it encounters a different antigen

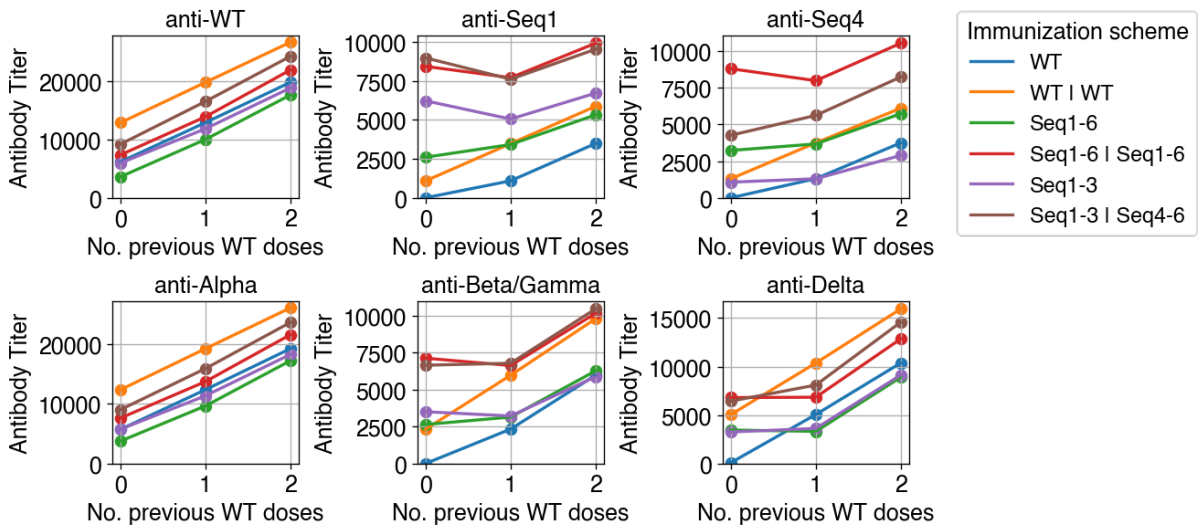

Figure 5. Titers against variants are affected by previous WT immunizations. Titers against various antigens (indicated in panel title) as a function of the number of previous WT immunizations. Sequences 2 and 3 are similar to sequence 1, so they are omitted for clarity. Sequences 5 and 6 are also omitted because they are similar to sequence 4. Beta and Gamma variants carry biochemically similar RBD mutations, so they are equivalent in this model.

for which the affinity is below a threshold. Over many cycles, the B cell is likely to encounter a different antigen at some point and die. Increasing the concentration counteracts this because a higher concentration will decrease the probability of low-affinity B cells dying. Thus, the optimal concentration is higher for the one-antigen case. It is important to note that our goal in this study is not to design antigens and immunization schemes that result in bnAbs that focus their binding footprint only on the conserved residues and can protect against a large set of variants. Having a high concentration of antigen in the one-antigen case would result in a polyclonal response of strain-specific antibodies, just like the all-antigen case, not bnAbs. Our goal in this study is to elicit a polyclonal antibody response that can protect against a limited set of variants. For this situation, manipulating antigen concentration can mitigate the difference between heterogeneous and homogeneous distribution of antigens on FDCs.

3.2.4 Titers against variants depend on previous exposure to WT and the number of mutations in the variant

Many individuals have been previously exposed to WT SARS-CoV-2 either through vaccination or natural infection, which might affect the antibody response generated upon immunization with a cocktail of our designed antigens. This is because memory B cells generated during prior exposure may compete with naïve B cells that seed GCs. To study this process, we use our model to predict antibody response given previous WT immunizations. Figure 5 graphs antibody titer as a function of the number of previous immunizations with WT for different immunization schemes. Although naive cells are likely involved in seeding new GCs in vivo (***Mesin et al.*** (***2020***)), here we assume that only memory cells seed new GCs to maximally illustrate the effects of memory. Section 3.2.5 studies the effect of including naive cells.

First, we compare WT versus WT | WT and WT | WT | WT immunization as well as cocktail versus WT | cocktail and WT | WT | cocktail immunization. Intuitively, anti-WT and anti-Alpha titers steadily increase as the number of previous immunizations with WT increase. However, titers against variants with numerous mutations are not always higher for WT | cocktail and WT | WT | cocktail compared to just cocktail immunization. For example, the anti-Seq1 titers decrease from 0 to 1 previous WT immunization for Seq1-6 | Seq1-6 and Seq1-3 | Seq4-6. This is because the mutational distance between the previous WT immunizations and the cocktails induces GC extinction,

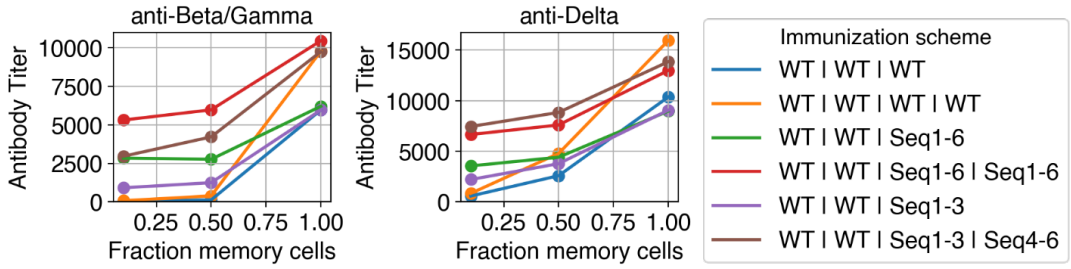

Figure 6. Titers against current variants are affected by the fraction of GC-seeding cells that are memory cells. Titers against Beta/Gamma and Delta variants as a function of the fraction of GC-seeding cells that are memory cells. Beta and Gamma variants carry biochemically similar RBD mutations, so they are equivalent in this model.

thereby limiting antibody development. Nonetheless, the decrease is not large even though GCs are seeded entirely by memory cells.

Next, we compare WT followed by WT immunization versus WT followed by cocktail immunization (for example, WT | WT versus WT | cocktail). For variants with numerous mutations, immunization with the cocktails after the first exposure to WT show higher titers than boosting with WT. For example, anti-Seq1 and anti-Seq4 titers are significantly higher in the WT | Seq1-6 | Seq1-6 and WT | Seq1-3 | Seq4-6 immunizations than the WT | WT immunization. This is because Seq1 and Seq4 each have numerous mutations, so the WT immunizations are not protective against Seq1 and Seq4. Anti-Delta, anti-Alpha, and anti-WT titers are higher for WT | WT immunization than WT | cocktail immunization because of the low number of mutations in these variants. Since the Beta and Gamma variants have more mutations than the Delta and Alpha variants, the anti-Beta/Gamma titers are equivalent between WT | cocktail and WT | WT immunizations.

3.2.5 Sequential immunization of WT and cocktail produces significantly higher titers against current variants than sequential immunization of WT and WT when naive B cells also seed new GCs.

Anti-Beta/Gamma and anti-Delta titers are not higher for cocktail immunizations compared to WT immunizations under specific conditions (assuming 1-2 previous immunizations with WT have been given and new GCs are seeded entirely by memory cells). However, if naive cells also seed GCs, then the titers for cocktail immunization following immunizations with WT become significantly higher. Illustrating this, Figure 6 shows titers as a function of the memory cell fraction given 2 previous WT immunizations. At a memory cell fraction of 0.5, anti-Beta/Gamma titers are around 10 to 15-fold higher for cocktail immunizations than WT immunizations. This is a large difference compared to the memory cell fraction of 1, where the titers are nearly equivalent, suggesting that cocktail immunization is less sensitive to the memory cell fraction. In the WT | WT | cocktail immunization, the WT and cocktail antigens are sufficiently different so that naïve cells can better develop against Beta and Gamma variants than in the WT | WT | WT immunization. Thus, decreasing the memory cell fraction (increasing naïve cell fraction) increases the difference between WT | WT | cocktail immunization and WT | WT | WT immunization. A similar, but less pronounced, effect occurs for anti-Delta titers, as the titers resulting from WT | WT | cocktail immunization are around 1.5 to 2-fold higher than WT | WT | WT immunization. This decrease occurs but is less pronounced for anti-Delta titers because the Delta variant has 2 RBD mutations whereas the Beta and Gamma variants have 3 RBD mutations.

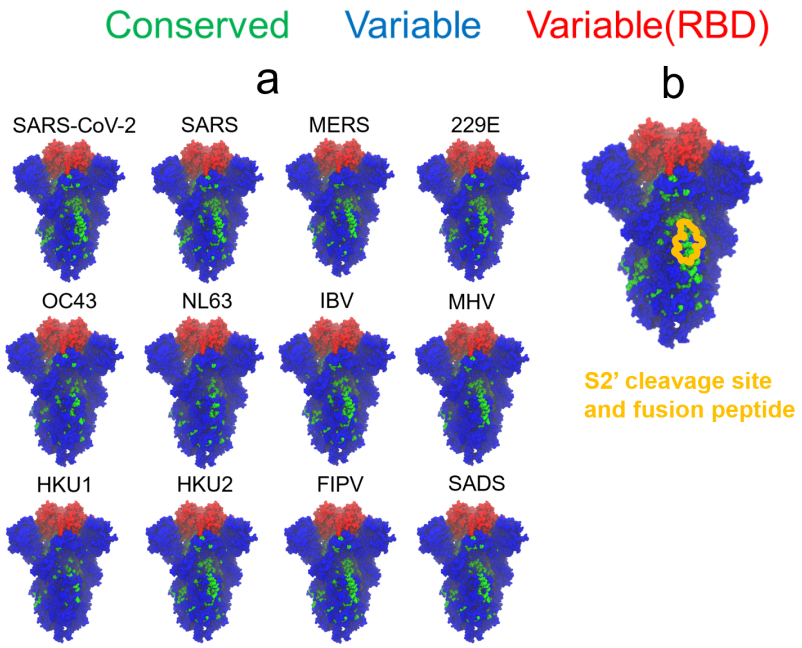

Figure 7. A conserved epitope in the S2 domain overlaps the S2' cleavage site and fusion peptide. (a) Spike protein structures colored by conservation fraction using each coronavirus as a reference. Green residues have conservation fractions above 0.8, blue residues have conservation fractions below 0.8 and are not in the RBD, and red residues have conservation fractions below 0.8 and are in the RBD. (b) Spike structure colored by conservation fraction using SARS-CoV-2 as a reference. The residues corresponding to the S2' cleavage site and fusion peptide are outlined in yellow. This structure is shown in a separate panel in order to avoid covering up some residues with the outline.

Figure 7–Figure supplement 1. Spike structure colored by conservation fraction exclusively using SARS-CoV-2 data.

3.3 The S2' cleavage site is structurally conserved and is a potential pan-coronavirus vaccine target.

Finally, we consider the more general problem of protecting against any coronavirus. To identify a pan-coronavirus vaccine target, we used the conservation analysis outlined in Section 2.1 to calculate spike conservation across 12 coronaviruses from different genera. Figure 7a illustrates conservation using a threshold fraction of 0.8. None of the coronaviruses are conserved in the RBD, which is simply because different coronaviruses bind different receptors. For example, SARS-CoV-2 and SARS to bind ACE2, while MERS binds DPP4 (Table S1).

In contrast to this, the S2 domain of the spike possesses residues that are structurally conserved across many coronaviruses. In particular, 9 out of 12 of the coronaviruses share a conserved patch of residues in the S2 subunit despite coming from diverse lineages. For example, SARS-CoV-2 and IBV both share this conserved region even though SARS-CoV-2 is a betacoronavirus that infects humans and IBV is a gammacoronavirus that infects birds.

These residues are conserved in many coronaviruses because they overlap with the S2' cleavage site and fusion peptide (Figure 7b). Following S1/S2 cleavage, further cleavage at the S2' site facilitates the release of the fusion peptide, which fuses the viral and host membranes so that the virus can infect the cell. Since the cleavage site and fusion peptide are functionally important, they must be conserved. It follows that if the virus is subject to evolutionary pressure by S2'-targeting antibodies, it is more difficult to develop escape mutations because potential mutations must also preserve S2' site function.

Lastly, enzymes such as furin, TMPRSS2, and cathepsin (***Örd et al.*** (***2020***); ***Hoffmann et al.*** (***2020***); ***Zhao et al.*** (***2021***)) must be able to access cleavage sites, which means that the cleavage sites are

exposed and accessible to proteins. Furthermore, this limits the virus's ability to develop spike densities that sterically hinder antibodies from accessing the S2' site. Electron tomography estimates that SARS-CoV-2 virions have 25-127 spikes and a diameter of 80 nm (***Klein et al.*** (***2020***)). Even at the dense estimate of 127 spikes, this produces a density of 0.6 spikes per 100 nm^2, which is half as dense as influenza (1.2 spikes per 100 nm^2) (***Harris et al.*** (***2013***); ***Amitai*** (***2020***)). Thus, compared to the broadly neutralizing epitope on the stem of the influenza hemagglutinin spike, from a steric point of view, generating antibodies to the stem of the coronavirus spike may be more feasible. Taken together, these attributes suggest that targeting the S2' site could serve as a potential pan-coronavirus vaccine strategy.

4 Discussion

SARS-CoV-2 has evolved variants within a few years since its emergence. Current vaccines that immunize with the WT spike show reduced efficacy for some of these variants, especially the newly emergent Omicron variant. This suggests that vaccines that can protect against multiple variant strains are becoming increasingly important. However, while SARS-CoV-2 is mutable, it is less so than influenza and HIV. Therefore, while universal vaccines against influenza and HIV largely focus on redirecting antibody responses to elicit bnAbs that target conserved regions only, such an approach may not be necessary to protect against SARS-CoV-2 variants. Targeting conserved regions may have disadvantages as well, as antibodies targeting the conserved class 4 RBD epitope can be non-neutralizing (***Jette et al.*** (***2021***)). In this paper, based on analyses of structures and sequences, we designed antigens that may protect against the most significant antibody-evading variants. Using a neural network, we check that these antigens are viable and can bind ACE2, and so may emerge in the future. We also use a computational model of AM to show that a cocktail of these antigens provides stronger protection against current variants and certain potential variants than the WT vaccination does. In particular, a specific cocktail (sequence 1-6) is predicted to be optimal.

Our cocktail vaccination schemes result in lower titers for the WT antigen than vaccination with the WT antigen alone. It is important to emphasize that our goal is to generate a response that is protective against different variants that may emerge. Such a vaccine will be less protective for any given variant compared to a strain-specific vaccine. However, a vaccine that provides adequate coverage to diverse potential variants would confer protection to a population, which may sufficiently restrict transmission on its own or buy time for a strain-specific vaccine to be developed without the need for strict non-pharmaceutical interventions.

The antibodies elicited by our cocktail vaccine are not bnAbs as they are not designed to target conserved regions. Our AM simulations suggest that the antibodies are broadly protective against multiple mutations that occur in the residues that are mutated in the designed antigens, but not against mutations that occur outside of those residues. Thus, variants that escape our cocktail vaccine may arise. However, note that the residues mutated in the designed antigens are predicted by our conservation analyses to be the most likely to emerge . Also, our AM model weights all residues equally in order to calculate the binding free energy, which is a limitation of the model. In actuality, certain residues easily abrogate antibody binding upon mutation (specifically, class 1 and 2 antibodies elicited from exposure to the WT sequence), and these escape mutations are the ones included in our designed antigens. Thus, our cocktail vaccine should protect against WT escape mutations. The mutations required to evade antibodies generated by our designed vaccine would be complementary to these WT escape mutations. It may be possible that an individual with a previous exposure to the WT sequence could have protection against the mutations that escape our cocktail vaccine.

A recent study created chimeric spike mRNA vaccines, in which the mRNA sequence contains segments from multiple sarbecoviruses, which induced higher neutralizing titers against various sarbecoviruses compared to the WT SARS-CoV-2 sequence (***Martinez et al.*** (***2021***)). However, the chimeras induced lower neutralizing titers against SARS-CoV-2 variants of concern than the WT

SARS-CoV-2 sequence. This is not surprising because the chimeras are designed using sarbecoviruses, which are more mutated from the variants of concern than WT SARS-CoV-2 is. If one aims to protect against just SARS-CoV-2 variants, the chimeric vaccine is not better than the WT SARS-CoV-2 vaccine. Our designed antigens are predicted to generate higher titers against SARS-CoV-2 variants than the WT by exploiting data specifically for variants. Therefore, our antigens would be more effective at addressing the immediate threat of vaccine-evading SARS-CoV-2 variants, while chimeric sarbecovirus vaccines would be more effective at addressing the long-term threat of future sarbecoviruses.

Looking beyond sarbecoviruses, a vaccine that protects against any coronavirus would be beneficial. To develop such a pan-coronavirus vaccine, identifying a conserved region of the coronavirus spike to target is necessary. However, previous conservation analysis of coronavirus spikes has not employed available structural data (***Prakash et al.*** (***2021***)), or has not considered diverse sets of coronaviruses (***Srinivasan et al.*** (***2020***); ***Guruprasad*** (***2020***); ***Hofmann et al.*** (***2006***)). In this work, we applied our conservation analysis of spike proteins from diverse coronaviruses and found that the S2′ cleavage site and fusion peptide may serve as pan-coronavirus vaccine targets. In support of this idea, previous work has shown that children who have never been infected with SARS-CoV-2 tend to possess antibodies targeting the S2 domain, which correlates with their lower susceptibility to SARS-CoV-2 (***Ng et al.*** (***2020***)), and antibodies targeting residues 810-824 (containing the S2′ site) showed particularly potent cross-reactivity. Since the position of the S2′ site is analogous to the position of the stem epitope of influenza hemagglutinin, previous approaches designed to target hemaglutinin's stem epitope (***Amitai et al.*** (***2020***); ***Nachbagauer et al.*** (***2021b***); ***Boyoglu-Barnum et al.*** (***2021***); ***Yassine et al.*** (***2015***); ***Bommakanti et al.*** (***2010***); ***Krammer et al.*** (***2013***)) can be applied to coronaviruses (***Joyce et al.*** (***2021***)). A limitation of this approach is that the S2′ site is not perfectly conserved, so targeting it would likely yield a vaccine that is less effective against a particular coronavirus than a strain-specific vaccine.

Taken together, the bulk of our results predict that a specific cocktail of variant antigens used as an immunogen would offer protection against variants of SARS-CoV-2 that may emerge in the future. We hope that experimental efforts to test this prediction will follow. We also suggest one vaccination scheme that may help protect against diverse coronaviruses.

5 Acknowledgments

Financial support was provided by the National Science Foundation Graduate Research Fellowship Grant No. 1745302, NIH Grant No. 1-R61-AI161805-01, and the Ragon Institute of MGH, MIT, and Harvard. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing HPC resources that have contributed to the research results reported within this work.

6 Competing interests

AKC is a member of the board and consultant for FL77 (a Flagship Pioneering company) and a member of the SAB of Evozyne.

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. http://tensorflow.org/, software available from tensorflow.org.

Abu-Raddad LJ, Chemaitelly H, Butt AA. Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants. New England Journal of Medicine. 2021 Jul; 385(2):187–189. http://www.nejm.org/doi/10.1056/NEJMc2104974, doi: 10.1056/NEJMc2104974.

Amitai A. Viral surface geometry shapes influenza and coronavirus spike evolution through antibody pressure. Biophysics; 2020.

Amitai A, Sangesland M, Barnes RM, Rohrer D, Lonberg N, Lingwood D, Chakraborty AK. Defining and Manipulating B Cell Immunodominance Hierarchies to Elicit Broadly Neutralizing Antibody Responses against Influenza Virus. Cell Systems. 2020; 11(6):573–588.e9. https://www.sciencedirect.com/science/article/pii/S2405471220303355, doi: https://doi.org/10.1016/j.cels.2020.09.005.

Barnes CO, Jette CA, Abernathy ME, Dam KMA, Esswein SR, Gristick HB, Malyutin AG, Sharaf NG, Huey-Tubman KE, Lee YE, Robbiani DF, Nussenzweig MC, West AP, Bjorkman PJ. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. Nature. 2020 Dec; 588(7839):682–687. http://www.nature.com/articles/s41586-020-2852-1, doi: 10.1038/s41586-020-2852-1.

Bates TA, Leier HC, Lyski ZL, McBride SK, Coulter FJ, Weinstein JB, Goodman JR, Lu Z, Siegel SAR, Sullivan P, Strnad M, Brunton AE, Lee DX, Adey AC, Bimber BN, O'Roak BJ, Curlin ME, Messer WB, Tafesse FG. Neutralization of SARS-CoV-2 variants by convalescent and BNT162b2 vaccinated serum. Nature Communications. 2021 Dec; 12(1):5135. https://www.nature.com/articles/s41467-021-25479-6, doi: 10.1038/s41467-021-25479-6.

Berek C, Milstein C. Mutation Drift and Repertoire Shift in the Maturation of the Immune Response. Immunological Reviews. 1987 Apr; 96(1):23–41. https://onlinelibrary.wiley.com/doi/10.1111/j.1600-065X.1987.tb00507.x, doi: 10.1111/j.1600-065X.1987.tb00507.x.

Bommakanti G, Citron MP, Hepler RW, Callahan C, Heidecker GJ, Najar TA, Lu X, Joyce JG, Shiver JW, Casimiro DR, ter Meulen J, Liang X, Varadarajan R. Design of an HA2-based Escherichia coli expressed influenza immunogen that protects mice from pathogenic challenge. Proceedings of the National Academy of Sciences. 2010 Aug; 107(31):13701–13706. http://www.pnas.org/cgi/doi/10.1073/pnas.1007465107, doi: 10.1073/pnas.1007465107.

Boyoglu-Barnum S, Ellis D, Gillespie RA, Hutchinson GB, Park YJ, Moin SM, Acton OJ, Ravichandran R, Murphy M, Pettie D, Matheson N, Carter L, Creanga A, Watson MJ, Kephart S, Ataca S, Vaile JR, Ueda G, Crank MC, Stewart L, et al. Quadrivalent influenza nanoparticle vaccines induce broad protection. Nature. 2021 Apr; 592(7855):623–628. http://www.nature.com/articles/s41586-021-03365-x, doi: 10.1038/s41586-021-03365-x.

Brink R. Germinal-Center B Cells in the Zone. Immunity. 2007 May; 26(5):552–554. https://linkinghub.elsevier.com/retrieve/pii/S1074761307002531, doi: 10.1016/j.immuni.2007.05.002.

Celada F, Seiden PE. Affinity maturation and hypermutation in a simulation of the humoral immune response. European Journal of Immunology. 1996 Jun; 26(6):1350–1358. https://onlinelibrary.wiley.com/doi/10.1002/eji.1830260626, doi: 10.1002/eji.1830260626.

Chan C, Billard M, Ramirez SA, Schmidl H, Monson E, Kepler TB. A Model for Migratory B Cell Oscillations from Receptor Down-Regulation Induced by External Chemokine Fields. Bulletin of Mathematical Biology. 2013 Jan; 75(1):185–205. http://link.springer.com/10.1007/s11538-012-9799-9, doi: 10.1007/s11538-012-9799-9.

Chaudhury S, Reifman J, Wallqvist A. Simulation of B Cell Affinity Maturation Explains Enhanced Antibody Cross-Reactivity Induced by the Polyvalent Malaria Vaccine AMA1. The Journal of Immunology. 2014 Sep; 193(5):2073–2086. http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1401054, doi: 10.4049/jimmunol.1401054.

Childs LM, Baskerville EB, Cobey S. Trade-offs in antibody repertoires to complex antigens. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015 Sep; 370(1676):20140245. https://royalsocietypublishing.org/doi/10.1098/rstb.2014.0245, doi: 10.1098/rstb.2014.0245.

Choi A, Koch M, Wu K, Dixon G, Oestreicher J, Legault H, Stewart-Jones GBE, Colpitts T, Pajon R, Bennett H, Carfi A, Edwards DK. Serum Neutralizing Activity of mRNA-1273 against SARS-CoV-2 Variants. Microbiology; 2021.

Dong J, Zost SJ, Greaney AJ, Starr TN, Dingens AS, Chen EC, Chen RE, Case JB, Sutton RE, Gilchuk P, Rodriguez J, Armstrong E, Gainza C, Nargi RS, Binshtein E, Xie X, Zhang X, Shi PY, Logue J, Weston S, et al. Genetic and structural basis for recognition of SARS-CoV-2 spike protein by a two-antibody cocktail. bioRxiv. 2021; https://www.biorxiv.org/content/early/2021/03/01/2021.01.27.428529, doi: 10.1101/2021.01.27.428529.

van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT, Balloux F. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infection, Genetics and Evolution. 2020 Sep; 83:104351. https://linkinghub.elsevier.com/retrieve/pii/S1567134820301829, doi: 10.1016/j.meegid.2020.104351.

Edara VV, Norwood C, Floyd K, Lai L, Davis-Gardner ME, Hudson WH, Mantus G, Nyhoff LE, Adelman MW, Fineman R, Patel S, Byram R, Gomes DN, Michael G, Abdullahi H, Beydoun N, Panganiban B, McNair N, Hellmeister K, Pitts J, et al. Infection- and vaccine-induced antibody binding and neutralization of the B.1.351 SARS-CoV-2 variant. Cell Host & Microbe. 2021 Apr; 29(4):516–521.e3. https://linkinghub.elsevier.com/retrieve/pii/S1931312821001372, doi: 10.1016/j.chom.2021.03.009.

Edara VV, Pinsky BA, Suthar MS, Lai L, Davis-Gardner ME, Floyd K, Flowers MW, Wrammert J, Hussaini L, Ciric CR, Bechnak S, Stephens K, Graham BS, Bayat Mokhtari E, Mudvari P, Boritz E, Creanga A, Pegu A, Derrien-Colemyn A, Henry AR, et al. Infection and Vaccine-Induced Neutralizing-Antibody Responses to the SARS-CoV-2 B.1.617 Variants. New England Journal of Medicine. 2021 Aug; 385(7):664–666. http://www.nejm.org/doi/10.1056/NEJMc2107799, doi: 10.1056/NEJMc2107799.

Edara VV, Pinsky BA, Suthar MS, Lai L, Davis-Gardner ME, Floyd K, Flowers MW, Wrammert J, Hussaini L, Ciric CR, Bechnak S, Stephens K, Graham BS, Bayat Mokhtari E, Mudvari P, Boritz E, Creanga A, Pegu A, Derrien-Colemyn A, Henry AR, et al. Infection and Vaccine-Induced Neutralizing-Antibody Responses to the SARS-CoV-2 B.1.617 Variants. New England Journal of Medicine. 2021 Aug; 385(7):664–666. http://www.nejm.org/doi/10.1056/NEJMc2107799, doi: 10.1056/NEJMc2107799.

van Erp EA, Luytjes W, Ferwerda G, van Kasteren PB. Fc-Mediated Antibody Effector Functions During Respiratory Syncytial Virus Infection and Disease. Frontiers in Immunology. 2019; 10:548. https://www.frontiersin.org/article/10.3389/fimmu.2019.00548, doi: 10.3389/fimmu.2019.00548.

Escolano A, Steichen JM, Dosenovic P, Kulp DW, Golijanin J, Sok D, Freund NT, Gitlin AD, Oliveira T, Araki T, Lowe S, Chen ST, Heinemann J, Yao KH, Georgeson E, Saye-Francisco KL, Gazumyan A, Adachi Y, Kubitz M, Burton DR, et al. Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. Cell. 2016 Sep; 166(6):1445–1458.e12. https://linkinghub.elsevier.com/retrieve/pii/S009286741630976X, doi: 10.1016/j.cell.2016.07.030.

Ganti RS, Chakraborty AK. Mechanisms underlying vaccination protocols that may optimally elicit broadly neutralizing antibodies against highly mutable pathogens. Phys Rev E. 2021 May; 103:052408. https://link.aps.org/doi/10.1103/PhysRevE.103.052408, doi: 10.1103/PhysRevE.103.052408.

Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, Gaebler C, Cho A, Agudelo M, Finkin S, Wang Z, Poston D, Muecksch F, Hatziioannou T, Bieniasz PD, Robbiani DF, Nussenzweig MC, Bjorkman PJ, Bloom JD. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nature Communications. 2021 Dec; 12(1):4196. http://www.nature.com/articles/s41467-021-24435-8, doi: 10.1038/s41467-021-24435-8.

Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, Hilton SK, Huddleston J, Eguia R, Crawford KHD, Dingens AS, Nargi RS, Sutton RE, Suryadevara N, Rothlauf PW, Liu Z, Whelan SPJ, Carnahan RH, Crowe JE, Bloom JD. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host & Microbe. 2021 Jan; 29(1):44–57.e9. https://linkinghub.elsevier.com/retrieve/pii/S1931312820306247, doi: 10.1016/j.chom.2020.11.007.

Gulli A, Pal S. Deep learning with Keras. Packt Publishing Ltd; 2017.

Guruprasad L. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 spike proteins for host receptor recognition. Proteins: Structure, Function, and Bioinformatics. 2020 Nov; 88(11):1387–1393. https://onlinelibrary.wiley.com/doi/10.1002/prot.25967, doi: 10.1002/prot.25967.

Harris AK, Meyerson JR, Matsuoka Y, Kuybeda O, Moran A, Bliss D, Das SR, Yewdell JW, Sapiro G, Subbarao K, Subramaniam S. Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. Proceedings of the National Academy of Sciences. 2013; 110(12):4592–4597. https://www.pnas.org/content/110/12/4592, doi: 10.1073/pnas.1214913110.

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr; 181(2):271–280.e8. https://linkinghub.elsevier.com/retrieve/pii/S0092867420302294, doi: 10.1016/j.cell.2020.02.052.

Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pohlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. Proceedings of the National Academy of Sciences. 2005 May; 102(22):7988–7993. http://www.pnas.org/cgi/doi/10.1073/pnas.0409465102, doi: 10.1073/pnas.0409465102.

Hofmann H, Simmons G, Rennekamp AJ, Chaipan C, Gramberg T, Heck E, Geier M, Wegele A, Marzi A, Bates P, Pohlmann S. Highly Conserved Regions within the Spike Proteins of Human Coronaviruses 229E and NL63 Determine Recognition of Their Respective Cellular Receptors. Journal of Virology. 2006 Sep; 80(17):8639–8652. https://journals.asm.org/doi/10.1128/JVI.00560-06, doi: 10.1128/JVI.00560-06.

Huang D, Tran JT, Olson A, Vollbrecht T, Tenuta M, Guryleva MV, Fuller RP, Schiffner T, Abadejos JR, Couvrette L, Blane TR, Saye K, Li W, Landais E, Gonzalez-Martin A, Schief W, Murrell B, Burton DR, Nemazee D, Voss JE. Vaccine elicitation of HIV broadly neutralizing antibodies from engineered B cells. Nature Communications. 2020 Dec; 11(1):5850. http://www.nature.com/articles/s41467-020-19650-8, doi: 10.1038/s41467-020-19650-8.

Hulswit RJG, Lang Y, Bakkers MJG, Li W, Li Z, Schouten A, Ophorst B, van Kuppeveld FJM, Boons GJ, Bosch BJ, Huizinga EG, de Groot RJ. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proceedings of the National Academy of Sciences. 2019; 116(7):2681–2690. https://www.pnas.org/content/116/7/2681, doi: 10.1073/pnas.1809667116.

Jette CA, Cohen AA, Gnanapragasam PNP, Muecksch F, Lee YE, Huey-Tubman KE, Schmidt F, Hatziioannou T, Bieniasz PD, Nussenzweig MC, West AP, Keeffe JR, Bjorkman PJ, Barnes CO. Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. bioRxiv. 2021; https://www.biorxiv.org/content/early/2021/04/26/2021.04.23.441195, doi: 10.1101/2021.04.23.441195.

Joyce MG, Chen WH, Sankhala RS, Hajduczki A, Thomas PV, Choe M, Martinez EJ, Chang WC, Peterson CE, Morrison EB, Smith C, Chen RE, Ahmed A, Wieczorek L, Anderson A, Case JB, Li Y, Oertel T, Rosado L, Ganesh A, et al. SARS-CoV-2 ferritin nanoparticle vaccines elicit broad SARS coronavirus immunogenicity. Cell Reports. 2021 Dec; 37(12):110143. https://linkinghub.elsevier.com/retrieve/pii/S2211124721016399, doi: 10.1016/j.celrep.2021.110143.

Kepler TB, Perelson AS. Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation. Immunology Today. 1993 Aug; 14(8):412–415. https://linkinghub.elsevier.com/retrieve/pii/016756999390145B, doi: 10.1016/0167-5699(93)90145-B.

Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, Stanifer ML, Boulant S, Bartenschlager R, Chlanda P. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nature Communications. 2020 Dec; 11(1):5885. http://www.nature.com/articles/s41467-020-19619-7, doi: 10.1038/s41467-020-19619-7.

Krammer F, Pica N, Hai R, Margine I, Palese P. Chimeric Hemagglutinin Influenza Virus Vaccine Constructs Elicit Broadly Protective Stalk-Specific Antibodies. Journal of Virology. 2013 Jun; 87(12):6542–6550. https://journals.asm.org/doi/10.1128/JVI.00641-13, doi: 10.1128/JVI.00641-13.

Krissinel E. Crystal contacts as nature's docking solutions. Journal of Computational Chemistry. 2010 Jan; 31(1):133–143. https://onlinelibrary.wiley.com/doi/10.1002/jcc.21303, doi: 10.1002/jcc.21303.

Krissinel E, Henrick K. Detection of Protein Assemblies in Crystals. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, R Berthold M, Glen RC, Diederichs K, Kohlbacher O, Fischer I, editors. *Computational Life Sciences*, vol. 3695 Berlin, Heidelberg: Springer Berlin Heidelberg; 2005.p. 163–174. http://link.springer.com/10.1007/11560500_15, doi: 10.1007/11560500_15.

Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology. 2007 Sep; 372(3):774–797. https://linkinghub.elsevier.com/retrieve/pii/S0022283607006420, doi: 10.1016/j.jmb.2007.05.022.

Kumar MDS. PINT: Protein-protein Interactions Thermodynamic Database. Nucleic Acids Research. 2006 Jan; 34(90001):D195–D198. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkj017, doi: 10.1093/nar/gkj017.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution. 2018 05; 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096, doi: 10.1093/molbev/msy096.

Li B, Deng A, Li K, Hu Y, Li Z, Xiong Q, Liu Z, Guo Q, Zou L, Zhang H, Zhang M, Ouyang F, Su J, Su W, Xu J, Lin H, Sun J, Peng J, Jiang H, Zhou P, et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. medRxiv. 2021; https://www.medrxiv.org/content/early/2021/07/23/2021.07.07.21260122, doi: 10.1101/2021.07.07.21260122.

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003 Nov; 426(6965):450–454. http://www.nature.com/articles/nature02145, doi: 10.1038/nature02145.

Li Z, Tomlinson AC, Wong AH, Zhou D, Desforges M, Talbot PJ, Benlekbir S, Rubinstein JL, Rini JM. The human coronavirus HCoV-229E S-protein structure and receptor binding. eLife. 2019 Oct; 8:e51230. https://elifesciences.org/articles/51230, doi: 10.7554/eLife.51230.

Liu C, Ginn HM, Dejnirattisai W, Supasa P, Wang B, Tuekprakhon A, Nutalai R, Zhou D, Mentzer AJ, Zhao Y, Duyvesteyn HME, López-Camacho C, Slon-Campos J, Walter TS, Skelly D, Johnson SA, Ritter TG, Mason C, Costa Clemens SA, Gomes Naveca F, et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. Cell. 2021 Aug; 184(16):4220–4236.e13. https://linkinghub.elsevier.com/retrieve/pii/S0092867421007558, doi: 10.1016/j.cell.2021.06.020.

Liu C, Ginn HM, Dejnirattisai W, Supasa P, Wang B, Tuekprakhon A, Nutalai R, Zhou D, Mentzer AJ, Zhao Y, Duyvesteyn HME, López-Camacho C, Slon-Campos J, Walter TS, Skelly D, Johnson SA, Ritter TG, Mason C, Costa Clemens SA, Gomes Naveca F, et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. Cell. 2021 Aug; 184(16):4220–4236.e13. https://linkinghub.elsevier.com/retrieve/pii/S0092867421007558, doi: 10.1016/j.cell.2021.06.020.

Liu H, Yuan M, Huang D, Bangaru S, Zhao F, Lee CCD, Peng L, Barman S, Zhu X, Nemazee D, Burton DR, van Gils MJ, Sanders RW, Kornau HC, Reincke SM, Prüss H, Kreye J, Wu NC, Ward AB, Wilson IA. A combination of cross-neutralizing antibodies synergizes to prevent SARS-CoV-2 and SARS-CoV pseudovirus infection. Cell Host & Microbe. 2021 May; 29(5):806–818.e6. https://linkinghub.elsevier.com/retrieve/pii/S1931312821001815, doi: 10.1016/j.chom.2021.04.005.

Luo S, Perelson AS. Competitive exclusion by autologous antibodies can prevent broad HIV-1 antibodies from arising. Proceedings of the National Academy of Sciences. 2015 Sep; 112(37):11654–11659. http://www.pnas.org/lookup/doi/10.1073/pnas.1505207112, doi: 10.1073/pnas.1505207112.

Martinez DR, Schäfer A, Leist SR, la Cruz GD, West A, Atochina-Vasserman EN, Lindesmith LC, Pardi N, Parks R, Barr M, Li D, Yount B, Saunders KO, Weissman D, Haynes BF, Montgomery SA, Baric RS. Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice. Science. 2021; 373(6558):991–998. https://www.science.org/doi/abs/10.1126/science.abi4506, doi: 10.1126/science.abi4506.

Mascola JR, Haynes BF. HIV-1 neutralizing antibodies: understanding nature's pathways. Immunological Reviews. 2013 Jul; 254(1):225–244. https://onlinelibrary.wiley.com/doi/10.1111/imr.12075, doi: 10.1111/imr.12075.

Mesin L, Schiepers A, Ersching J, Barbulescu A, Cavazzoni CB, Angelini A, Okada T, Kurosaki T, Victora GD. Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting. Cell. 2020; 180(1):92–106.e11. https://www.sciencedirect.com/science/article/pii/S0092867419313170, doi: https://doi.org/10.1016/j.cell.2019.11.032.

Meyer-Hermann M, Mohr E, Pelletier N, Zhang Y, Victora G, Toellner KM. A Theory of Germinal Center B Cell Selection, Division, and Exit. Cell Reports. 2012 Jul; 2(1):162–174. https://linkinghub.elsevier.com/retrieve/pii/S2211124712001350, doi: 10.1016/j.celrep.2012.05.010.

Nachbagauer R, Feser J, Naficy A, Bernstein DI, Guptill J, Walter EB, Berlanda-Scorza F, Stadlbauer D, Wilson PC, Aydillo T, Behzadi MA, Bhavsar D, Bliss C, Capuano C, Carreño JM, Chromikova V, Claeys C, Coughlan L, Freyn AW, Gast C, et al. A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. Nature Medicine. 2021 Jan; 27(1):106–114. http://www.nature.com/articles/s41591-020-1118-7, doi: 10.1038/s41591-020-1118-7.

Nachbagauer R, Feser J, Naficy A, Bernstein DI, Guptill J, Walter EB, Berlanda-Scorza F, Stadlbauer D, Wilson PC, Aydillo T, Behzadi MA, Bhavsar D, Bliss C, Capuano C, Carreño JM, Chromikova V, Claeys C, Coughlan L, Freyn AW, Gast C, et al. A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. Nature Medicine. 2021 Jan; 27(1):106–114. http://www.nature.com/articles/s41591-020-1118-7, doi: 10.1038/s41591-020-1118-7.

Ng KW, Faulkner N, Cornish GH, Rosa A, Harvey R, Hussain S, Ulferts R, Earl C, Wrobel AG, Benton DJ, Roustan C, Bolland W, Thompson R, Agua-Doce A, Hobson P, Heaney J, Rickman H, Paraskevopoulou S, Houlihan CF, Thomson K, et al. Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. Science. 2020 Dec; 370(6522):1339–1343. https://www.science.org/doi/10.1126/science.abe1107, doi: 10.1126/science.abe1107.

Oprea M. Somatic mutation leads to efficient affinity maturation when centrocytes recycle back to centroblasts. Immunology Letters. 1997 May; 56(1-3):43. https://linkinghub.elsevier.com/retrieve/pii/S0165247897870009, doi: 10.1016/S0165-2478(97)87000-9.

Pauthner MG, Nkolola JP, Havenar-Daughton C, Murrell B, Reiss SM, Bastidas R, Prévost J, Nedellec R, von Bredow B, Abbink P, Cottrell CA, Kulp DW, Tokatlian T, Nogal B, Bianchi M, Li H, Lee JH, Butera ST, Evans DT, Hangartner L, et al. Vaccine-Induced Protection from Homologous Tier 2 SHIV Challenge in Nonhuman Primates Depends on Serum-Neutralizing Antibody Titers. Immunity. 2019 Jan; 50(1):241–252.e6. https://linkinghub.elsevier.com/retrieve/pii/S107476131830493X, doi: 10.1016/j.immuni.2018.11.011.

Planas D, Saunders N, Maes P, Guivel-Benhassine F, Planchais C, Buchrieser J, Bolland WH, Porrot F, Staropoli I, Lemoine F, Péré H, Veyer D, Puech J, Rodary J, Baela G, Dellicour S, Raymenants J, Gorissen S, Geenen C, Vanmechelen B, et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. Nature. 2021 Dec; p. d41586–021–03827–2. https://www.nature.com/articles/d41586-021-03827-2, doi: 10.1038/d41586-021-03827-2.

Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Le Guen J, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Courtellemont L, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature. 2021 Aug; 596(7871):276–280. https://www.nature.com/articles/s41586-021-03777-9, doi: 10.1038/s41586-021-03777-9.

Prakash S, Srivastava R, Coulon PG, Dhanushkodi NR, Chentoufi AA, Tifrea DF, Edwards RA, Figueroa CJ, Schubl SD, Hsieh L, Buchmeier MJ, Bouziane M, Nesburn AB, Kuppermann BD, BenMohamed L. Genome-Wide B Cell, CD4 $^{+}$, and CD8 $^{+}$ T Cell Epitopes That Are Highly Conserved between Human and Animal Coronaviruses, Identified from SARS-CoV-2 as Targets for Preemptive Pan-Coronavirus Vaccines. The Journal of Immunology. 2021 Jun; 206(11):2566–2582. http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.2001438, doi: 10.4049/jimmunol.2001438.

Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, Agudelo M, Barnes CO, Gazumyan A, Finkin S, Hägglöf T, Oliveira TY, Viant C, Hurley A, Hoffmann HH, Millard KG, Kost RG, Cipolla M, Gordon K, Bianchini F, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature. 2020 Aug; 584(7821):437–442. http://www.nature.com/articles/s41586-020-2456-9, doi: 10.1038/s41586-020-2456-9.

Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He Wt, Limbo O, Smith C, Song G, Woehl J, Yang L, Abbott RK, Callaghan S, Garcia E, Hurtado J, Parren M, Peng L, Ramirez S, Ricketts J, Ricciardi MJ, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science. 2020 Aug; 369(6506):956–963. https://www.sciencemag.org/lookup/doi/10.1126/science.abc7520, doi: 10.1126/science.abc7520.

Sachdeva V, Husain K, Sheng J, Wang S, Murugan A. Tuning environmental timescales to evolve and maintain generalists. Proceedings of the National Academy of Sciences. 2020 Jun; 117(23):12693–12699. http://www.pnas.org/lookup/doi/10.1073/pnas.1914586117, doi: 10.1073/pnas.1914586117.

Scheid JF, Barnes CO, Eraslan B, Hudak A, Keeffe JR, Cosimi LA, Brown EM, Muecksch F, Weisblum Y, Zhang S, Delorey T, Woolley AE, Ghantous F, Park SM, Phillips D, Tusi B, Huey-Tubman KE, Cohen AA, Gnanapragasam PNP, Rzasa K, et al. B cell genomics behind cross-neutralization of SARS-CoV-2 variants and SARS-CoV. Cell. 2021 Jun; 184(12):3205–3221.e24. https://linkinghub.elsevier.com/retrieve/pii/S0092867421005353, doi: 10.1016/j.cell.2021.04.032.

Schmidt F, Muecksch F, Weisblum Y, Da Silva J, Bednarski E, Cho A, Wang Z, Gaebler C, Caskey M, Nussenzweig MC, Hatziioannou T, Bieniasz PD. Plasma Neutralization of the SARS-CoV-2 Omicron Variant. New England Journal of Medicine. 2021 Dec; p. NEJMc2119641. http://www.nejm.org/doi/10.1056/NEJMc2119641, doi: 10.1056/NEJMc2119641.

Schultze B, Cavanagh D, Herrler G. Neuraminidase treatment of avian infectious bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. Virology. 1992 Aug; 189(2):792–794. https://linkinghub.elsevier.com/retrieve/pii/004268229290608R, doi: 10.1016/0042-6822(92)90608-R.

Shaffer JS, Moore PL, Kardar M, Chakraborty AK. Optimal immunization cocktails can promote induction of broadly neutralizing Abs against highly mutable pathogens. Proceedings of the National Academy of Sciences. 2016 Nov; 113(45):E7039–E7048. http://www.pnas.org/lookup/doi/10.1073/pnas.1614940113, doi: 10.1073/pnas.1614940113.

988 **Sprenger KG**, Louveau JE, Murugan PM, Chakraborty AK. Optimizing immunization protocols to elicit broadly
989 neutralizing antibodies. Proceedings of the National Academy of Sciences. 2020; 117(33):20077–20087. https:
990 //www.pnas.org/content/117/33/20077, doi: 10.1073/pnas.1919329117.

991 **Srinivasan S**, Cui H, Gao Z, Liu M, Lu S, Mkandawire W, Narykov O, Sun M, Korkin D. Structural Genomics
992 of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. Viruses. 2020 Mar;
993 12(4):360. https://www.mdpi.com/1999-4915/12/4/360, doi: 10.3390/v12040360.

994 **Starr TN**, Czudnochowski N, Liu Z, Zatta F, Park YJ, Addetia A, Pinto D, Beltramello M, Hernandez P, Greaney
995 AJ, Marzi R, Glass WG, Zhang I, Dingens AS, Bowen JE, Tortorici MA, Walls AC, Wojcechowskyj JA, De Marco A,
996 Rosen LE, et al. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. Nature. 2021
997 Jul; http://www.nature.com/articles/s41586-021-03807-6, doi: 10.1038/s41586-021-03807-6.

998 **Starr TN**, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, Li JZ, Bloom JD. Prospective mapping
999 of viral mutations that escape antibodies used to treat COVID-19. Science. 2021; 371(6531):850–854. https:
1000 //science.sciencemag.org/content/371/6531/850, doi: 10.1126/science.abf9302.

1001 **Starr TN**, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the mon-
1002 oclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Reports Medicine. 2021 Apr; 2(4):100255.
1003 https://linkinghub.elsevier.com/retrieve/pii/S2666379121000719, doi: 10.1016/j.xcrm.2021.100255.

1004 **Starr TN**, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC,
1005 King NP, Veesler D, Bloom JD. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals
1006 Constraints on Folding and ACE2 Binding. Cell. 2020; 182(5):1295–1310.e20. https://www.sciencedirect.com/
1007 science/article/pii/S0092867420310035, doi: https://doi.org/10.1016/j.cell.2020.08.012.

1008 **Steichen J**, Kulp D, Tokatlian T, Escolano A, Dosenovic P, Stanfield R, McCoy L, Ozorowski G, Hu X, Kalyuzhniy
1009 O, Briney B, Schiffner T, Garces F, Freund N, Gitlin A, Menis S, Georgeson E, Kubitz M, Adachi Y, Jones M,
1010 et al. HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibod-
1011 ies. Immunity. 2016 Sep; 45(3):483–496. https://linkinghub.elsevier.com/retrieve/pii/S1074761316303405, doi:
1012 10.1016/j.immuni.2016.08.016.

1013 **Tas JMJ**, Mesin L, Pasqual G, Targ S, Jacobsen JT, Mano YM, Chen CS, Weill JC, Reynaud CA, Browne EP, Meyer-
1014 Hermann M, Victora GD. Visualizing antibody affinity maturation in germinal centers. Science. 2016 Mar;
1015 351(6277):1048–1054. https://www.science.org/doi/10.1126/science.aad3439, doi: 10.1126/science.aad3439.

1016 **Thompson JD**, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence
1017 alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic
1018 Acids Research. 1994; 22(22):4673–4680. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/22.22.
1019 4673, doi: 10.1093/nar/22.22.4673.

1020 **Tresnan DB**, Levis R, Holmes KV. Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and
1021 human coronaviruses in serogroup I. Journal of Virology. 1996 Dec; 70(12):8669–8674. https://journals.asm.
1022 org/doi/10.1128/jvi.70.12.8669-8674.1996, doi: 10.1128/jvi.70.12.8669-8674.1996.

1023 **Victora GD**, Nussenzweig MC. Germinal Centers. Annual Review of Immunology. 2012; 30(1):429–457. https:
1024 //doi.org/10.1146/annurev-immunol-020711-075032, doi: 10.1146/annurev-immunol-020711-075032, pMID:
1025 22224772.

1026 **Virtanen P**, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser
1027 W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E,
1028 Carey CJ, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;
1029 17:261–272. doi: 10.1038/s41592-019-0686-2.

1030 **Wang N**, Shi X, Jiang L, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X, Arledge KC, Chen YH, Zhang L, Wang X.
1031 Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. Cell Research.
1032 2013 Aug; 23(8):986–993. http://www.nature.com/articles/cr201392, doi: 10.1038/cr.2013.92.

1033 **Wang P**, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR,
1034 Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. Antibody Resistance of
1035 SARS-CoV-2 Variants B.1.351 and B.1.1.7. Immunology; 2021.

1036 **Wang S**, Mata-Fink J, Kriegsman B, Hanson M, Irvine DJ, Eisen HN, Burton DR, Wittrup KD, Kardar M, Chakraborty
1037 AK. Manipulating the Selection Forces during Affinity Maturation to Generate Cross-Reactive HIV Antibod-
1038 ies. Cell. 2015 Feb; 160(4):785–797. https://linkinghub.elsevier.com/retrieve/pii/S0092867415000707, doi:
1039 10.1016/j.cell.2015.01.027.

Wang Z, Muecksch F, Schaefer-Babajew D, Finkin S, Viant C, Gaebler C, Hoffmann HH, Barnes CO, Cipolla M, Ramos V, Oliveira TY, Cho A, Schmidt F, Da Silva J, Bednarski E, Aguado L, Yee J, Daga M, Turroja M, Millard KG, et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. Nature. 2021 Jul; 595(7867):426–431. http://www.nature.com/articles/s41586-021-03696-9, doi: 10.1038/s41586-021-03696-9.

Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature. 2021 Apr; 592(7855):616–622. http://www.nature.com/articles/s41586-021-03324-6, doi: 10.1038/s41586-021-03324-6.

Williams RK, Jiang GS, Holmes KV. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. Proceedings of the National Academy of Sciences. 1991 Jul; 88(13):5533–5536. http://www.pnas.org/cgi/doi/10.1073/pnas.88.13.5533, doi: 10.1073/pnas.88.13.5533.

Yassine HM, Boyington JC, McTamney PM, Wei CJ, Kanekiyo M, Kong WP, Gallagher JR, Wang L, Zhang Y, Joyce MG, Lingwood D, Moin SM, Andersen H, Okuno Y, Rao SS, Harris AK, Kwong PD, Mascola JR, Nabel GJ, Graham BS. Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. Nature Medicine. 2015 Sep; 21(9):1065–1070. http://www.nature.com/articles/nm.3927, doi: 10.1038/nm.3927.

Zamorano Cuervo N, Grandvaux N. ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. eLife. 2020 Nov; 9:e61390. https://elifesciences.org/articles/61390, doi: 10.7554/eLife.61390.

Zemla AT, Lang DM, Kostova T, Andino R, Ecale Zhou CL. StralSV: assessment of sequence variability within similar 3D structures and application to polio RNA-dependent RNA polymerase. BMC Bioinformatics. 2011 6; 12(1). doi: 10.1186/1471-2105-12-226.

Zhao MM, Yang WL, Yang FY, Zhang L, Huang WJ, Hou W, Fan CF, Jin RH, Feng YM, Wang YC, Yang JK. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. Signal Transduction and Targeted Therapy. 2021 Dec; 6(1):134. http://www.nature.com/articles/s41392-021-00558-8, doi: 10.1038/s41392-021-00558-8.

Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, Boerwinkle E, Fu YX. Moderate mutation rate in the SARS coronavirus genome and its implication. BMC Evolutionary Biology. 2004; 4(1):21. http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-4-21, doi: 10.1186/1471-2148-4-21.

Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Tuekprakhon A, Nutalai R, Wang B, Paesen GC, Lopez-Camacho C, Slon-Campos J, Hallis B, Coombes N, Bewley K, Charlton S, Walter TS, Skelly D, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell. 2021 Apr; 184(9):2348–2361.e6. https://linkinghub.elsevier.com/retrieve/pii/S0092867421002269, doi: 10.1016/j.cell.2021.02.037.

Zost SJ, Gilchuk P, Case JB, Binshtein E, Chen RE, Nkolola JP, Schäfer A, Reidy JX, Trivette A, Nargi RS, Sutton RE, Suryadevara N, Martinez DR, Williamson LE, Chen EC, Jones T, Day S, Myers L, Hassan AO, Kafai NM, et al. Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature. 2020 Aug; 584(7821):443–449. http://www.nature.com/articles/s41586-020-2548-6, doi: 10.1038/s41586-020-2548-6.

Örd M, Faustova I, Loog M. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. Scientific Reports. 2020 Dec; 10(1):16944. http://www.nature.com/articles/s41598-020-74101-0, doi: 10.1038/s41598-020-74101-0.

Supplementary Information

Table S1. Coronaviruses used in conservation analysis along with the PDB ID of the spike protein, the genus of the coronavirus, and the coronavirus's host receptor.

Coronavirus Name	PDB ID	Genus	Receptor
SARS-CoV-2	6VXX	Beta	ACE2 (*Zamorano Cuervo and Grandvaux* (*2020*))
SARS	5X58	Beta	ACE2 (*Li et al.* (*2003*))
MERS	5X59	Beta	DPP4 (*Wang et al.* (*2013*))
229E	6U7H	Alpha	APN (*Li et al.* (*2019*))
HKU1	5I08	Beta	Sialic acid (*Hulswit et al.* (*2019*))
OC43	6NZK	Beta	Sialic acid (*Hulswit et al.* (*2019*))
NL63	7KIP	Alpha	ACE2 (*Hofmann et al.* (*2005*))
IBV	6CV0	Gamma	Sialic acid (*Schultze et al.* (*1992*))
MHV	3JCL	Beta	CEACAM1 (*Williams et al.* (*1991*))
FIPV	6JX7	Alpha	APN (*Tresnan et al.* (*1996*))
SADS	6M16	Alpha	Unknown
HKU2	6M15	Alpha	Unknown

Table S2. Class 1 and class 2 antibodies used to identify RBD escape mutations.

Antibody Name	Class	Reference
C105	1	*Greaney et al.* (*2021a*)
COV2-2165	1	*Greaney et al.* (*2021b*)
COV2-2196	1	*Dong et al.* (*2021*)
COV2-2832	1	*Greaney et al.* (*2021b*)
LY-CoV016	1	*Starr et al.* (*2021b*)
REGN10933	1	*Starr et al.* (*2021b*)
S2E12	1	*Starr et al.* (*2021a*)
S2H14	1	*Starr et al.* (*2021a*)
C002	2	*Greaney et al.* (*2021a*)
C121	2	*Greaney et al.* (*2021a*)
C144	2	*Greaney et al.* (*2021a*)
COV2-2050	2	*Greaney et al.* (*2021b*)
COV2-2096	2	*Greaney et al.* (*2021b*)
COV2-2479	2	*Greaney et al.* (*2021b*)
LY-CoV555	2	*Starr et al.* (*2021c*)
S2D106	2	*Starr et al.* (*2021a*)
S2H13	2	*Starr et al.* (*2021a*)
S2H58	2	*Starr et al.* (*2021a*)
S2X16	2	*Starr et al.* (*2021a*)
S2X58	2	*Starr et al.* (*2021a*)

Choice of number of escape mutations considered

We chose the top 34 escape mutations because mutating too many residues was likely to destabilize the RBD. First, we noted that the deep mutational scanning data was based on RBDs with 1-7 mutations, so we could not design RBDs with more than 7 mutations. Furthermore, rather than mutating the maximum number of 7 residues, we chose to mutate 5 residues per RBD to further decrease the probability of destabilizing the RBD. Since the mutations were divided into 2 groups, mutating 5 residues per RBD allowed 10 unique residues to be mutated in total. The top 34 escape mutations was the largest set that contained 10 unique residues (the top 35 escape mutations contained 11 unique residues).

If 7 mutations per RBD were allowed, this corresponded to 14 unique residues. The top 40 escape mutations contained 14 unique residues. However, the total escape of the top 40 escape mutations was only moderately larger than the total escape of the top 34 escape mutations, so choosing the top 34 mutations did not significantly affect the escape potential of our antigens.

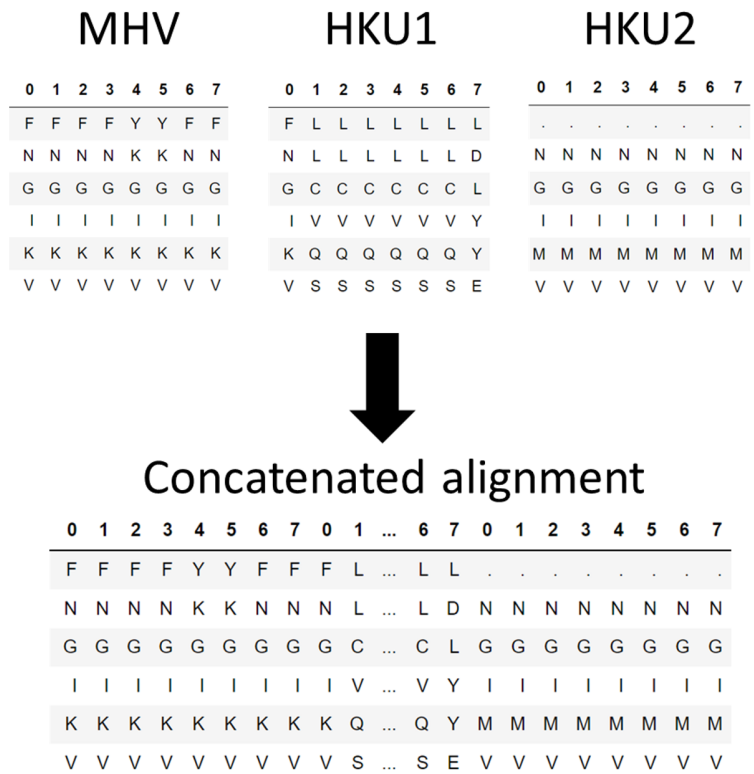

Figure 1–Figure supplement 1. Illustration of the concatenation of individual coronavirus alignments into a single multiple sequence alignment. The number of coronaviruses and size of the alignments have been reduced for clarity and to match the example table from Figure 1.

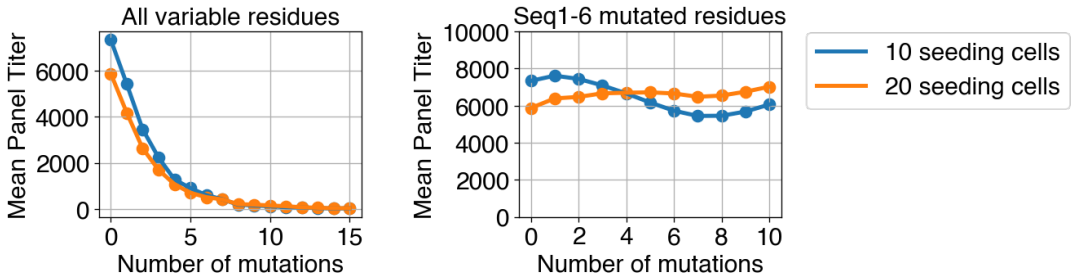

Figure 4–Figure supplement 1. Mean panel titers for the Seq1-6 | Seq1-6 immunization scheme using 10 and 20 seeding cells. Mutations occur either in any variable residues (All variable residues) or in the same residues that are mutated in the sequences 1-6 (Seq1-6 mutated residues). B cells are assumed to encounter all antigens at a time on the FDC.

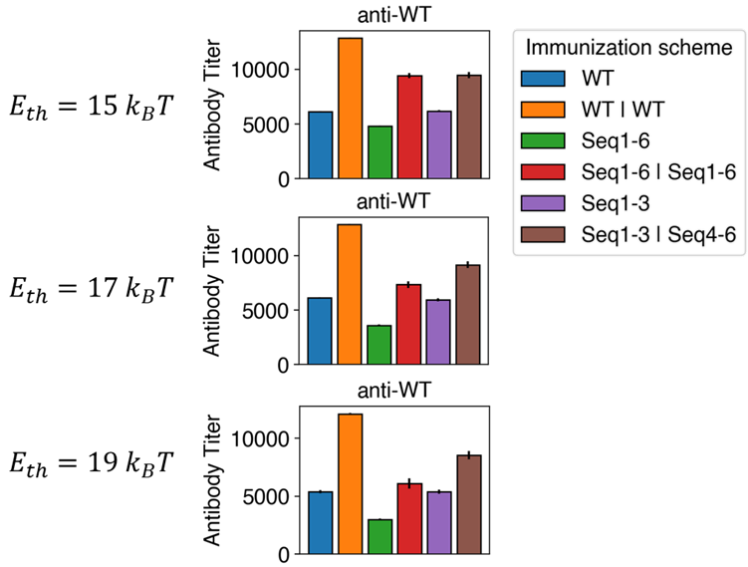

Figure 4–Figure supplement 2. Anti-WT titers for various immunization schemes and energy thresholds for calculating titers (E_{th}).

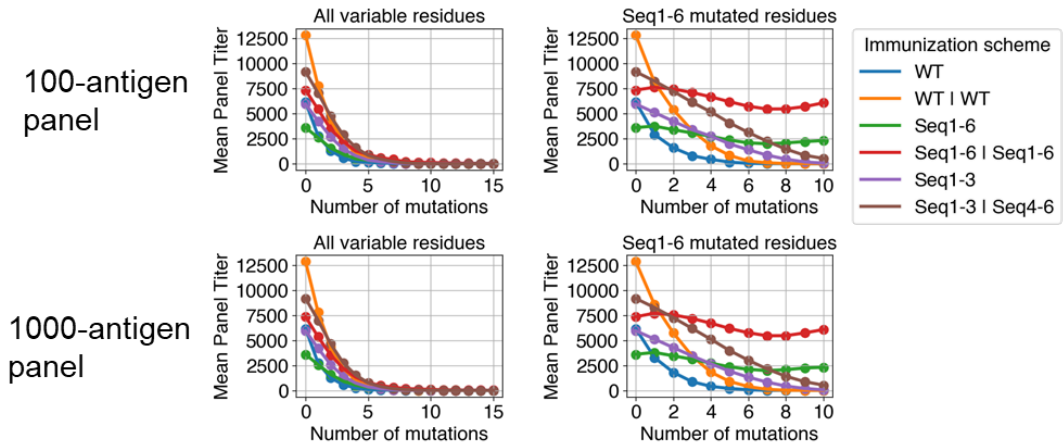

Figure 4–Figure supplement 3. Mean panel titers as a function of the number of mutations in panel antigens. Mutations occur either in any variable residues (All variable residues) or in the same residues that are mutated in the sequences 1-6 (Seq1-6 mutated residues). Panel titers are calculated against panels of 100 antigens and 1000 antigens. Mutated residues take on a value of -4. B cells are assumed to encounter all antigens at a time on the FDC.

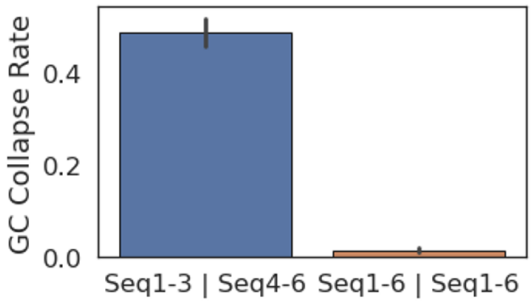

Figure 4–Figure supplement 4. GC collapse rate for Seq1-3 | Seq4-6 and Seq1-6 | Seq1-6.

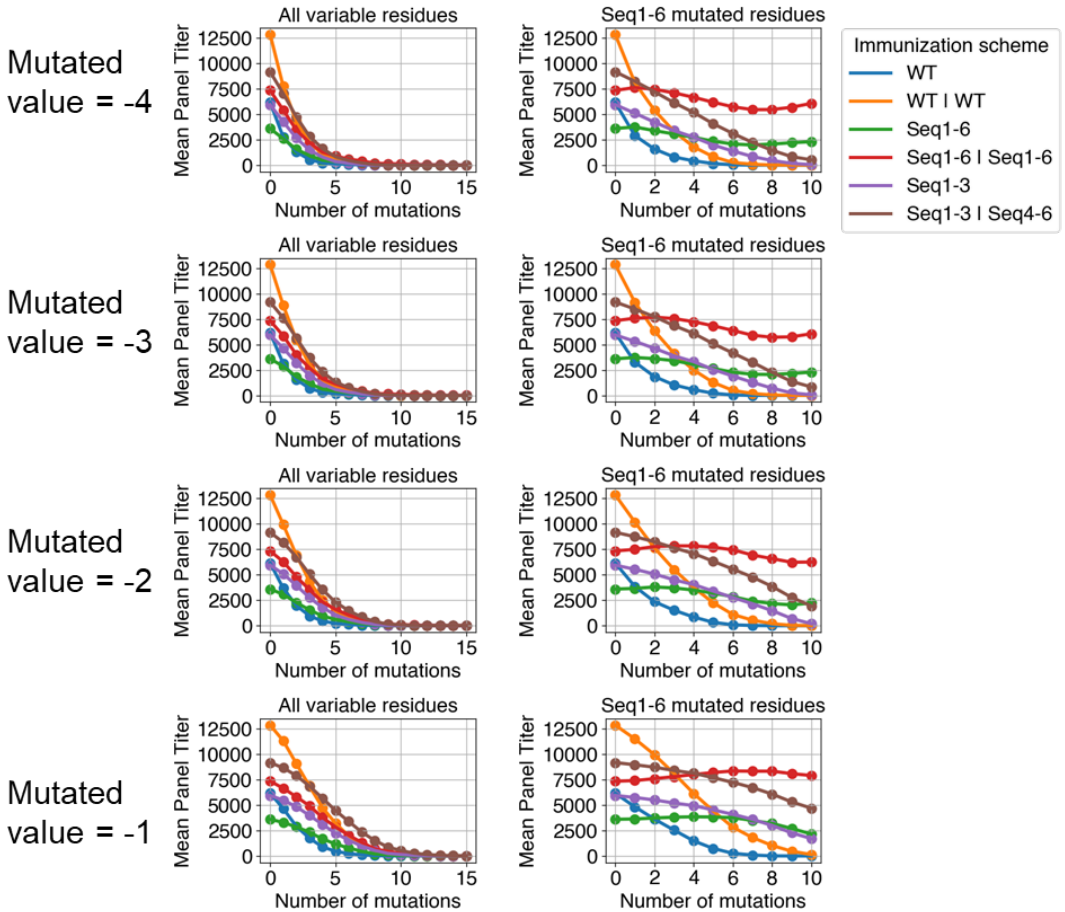

Figure 4–Figure supplement 5. Mean panel titers as a function of the number of mutations in panel antigens. Mutations occur either in any variable residues (All variable residues) or in the same residues that are mutated in the sequences 1-6 (Seq1-6 mutated residues). B cells are assumed to encounter all antigens at a time on the FDC. Mutated residues take on values of -4, -3, -2, or -1, as indicated.

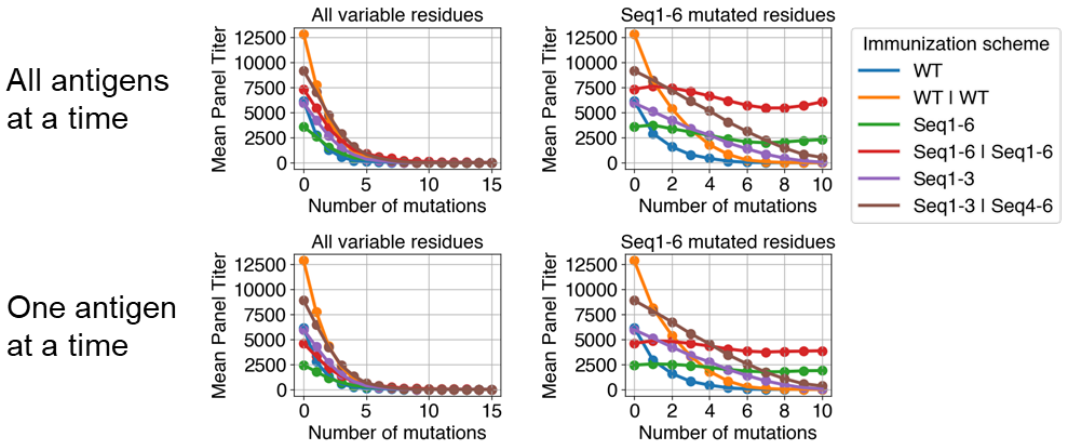

Figure 4–Figure supplement 6. Mean panel titers as a function of the number of mutations in panel antigens. Mutations occur either in any variable residues (All variable residues) or in the same residues that are mutated in the sequences 1-6 (Seq1-6 mutated residues). Mutated residues take on values of -4. B cells encounter either all antigens at a time (All-antigen) or one antigen at a time (One-antigen) on the FDC.

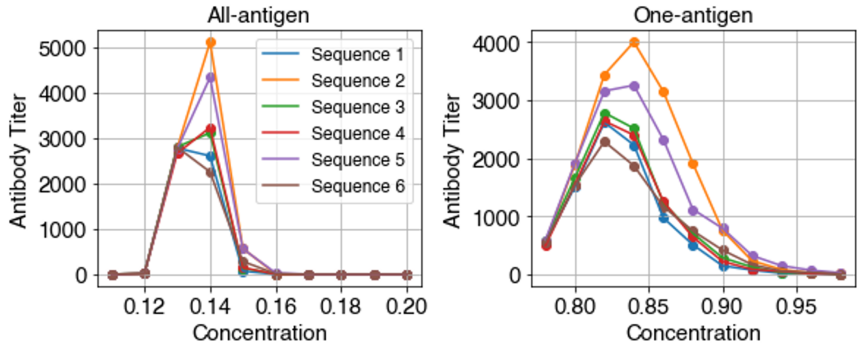

Figure 4–Figure supplement 7. Antibody titers against sequences 1-6 as a function of concentration if B cells encounter all antigens (All-antigen) or one antigen (One-antigen) at a time on the FDC.

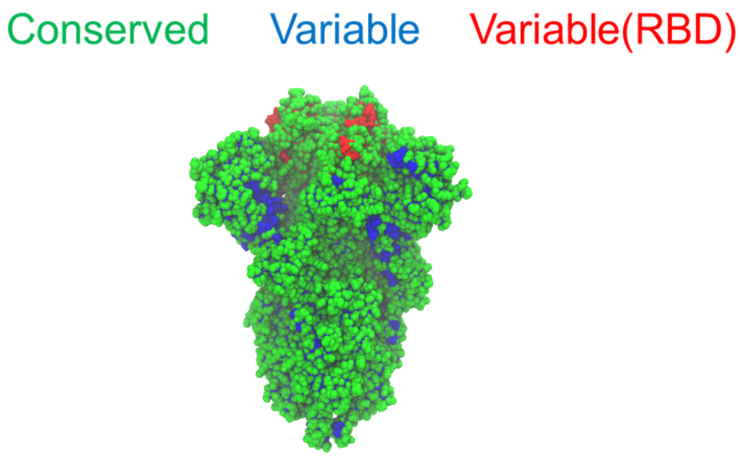

Figure 7–Figure supplement 1. Spike structure colored by conservation fraction exclusively using SARS-CoV-2 data. Green residues have conservation fractions above 0.8, blue residues have conservation fractions below 0.8 and are not in the RBD, and red residues have conservation fractions below 0.8 and are in the RBD. PDB structures used to calculate the structural conservation fraction include the WT structure (PDB ID: 6VXX), the Alpha variant structure (PDB ID: 7LWI), the Beta variant structure (PDB ID: 7LWS), the Delta variant structure (PDB ID: 7V7Q), and the Gamma variant structure (PDB ID: 7M8K). ~300,000 spike sequences were used to calculate the biochemical conservation fraction, which are available at https://github.com/ericzwang/sars2-vaccine/tree/main/data/aligned-cov-sequences.gz.