Deep generative models predict SARS-CoV-2 Spike infectivity and foreshadow neutralizing antibody escape

Noor Youssef[1,2], Fadi Ghantous[3], Sarah Gurev[1,2], Kelly Brock[4], Javier A. Jaimes[5], Ann Dauphin[5], Leonid Yurkovetskiy[5], Daria Soto[5], Ralph Estanboulieh[1], Ben Kotzen[6], Matteo Bosso[5], Jacob Lemieux[6,7]*, Jeremy Luban[2,5,7,8]*, Michael S. Seaman[3]*, Debora S. Marks[1,2]*

[1]Harvard Medical School, Boston, Massachusetts, USA
[2]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
[3]Center for Virology and Vaccine Research Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA.
[4]Kernal Biologics, Cambridge, Massachusetts, USA.
[5]Program in Molecular Medicine, University of Massachusetts Chan Med School, Worcester, Massachusetts, USA.
[6]Massachusetts General Hospital, Boston, Massachusetts, USA.
[7]Massachusetts Consortium on Pathogen Readiness, Boston, Massachusetts, USA.
[8]Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts, USA

*Corresponding authors
lemieux@broadinstitute.org, jeremy.luban@umassmed.edu, mseaman@bidmc.harvard.edu, debbie@hms.harvard.edu

Abstract

Recurrent waves of SARS-CoV-2 infection, driven by the periodic emergence of new viral variants, highlight the need for vaccines and therapeutics that remain effective against future strains. Yet, our ability to proactively evaluate such therapeutics is limited to assessing their effectiveness against previous or circulating variants, which may differ significantly in their antibody escape from future viral evolution. To address this challenge, we developed deep learning methods to predict the effect of mutations on fitness and escape from neutralizing antibodies and used this information to engineer a set of 68 unique SARS-CoV-2 Spike proteins. The designed constructs, which incorporated novel combinations of up to 46 mutations relative to the ancestral strain, were infectious and evaded neutralization by nine well-characterized panels of human polyclonal anti-SARS-CoV-2 immune sera. Designed constructs on previous SARS-CoV-2 strains anticipated the antibody neutralization escape of variants seen subsequently during the COVID-19 pandemic. We demonstrate that designed Spike constructs using data available at the time of the implementation of the 2022 bivalent mRNA booster vaccine foretold the level of neutralizing antibody escape observed in the most recently emerging variants. Our approach provides extensive datasets of antigenically diverse escape variants to evaluate the protective ability of vaccines and therapeutics to inhibit future variants. This approach is generalizable to other viral pathogens.

Introduction

The emergence of viral variants that evade protective immunity from prior infection, vaccines, and therapeutics is a challenge for the control of viral spread. This issue has come to the forefront during the COVID-19 pandemic. While anti-SARS-CoV-2 vaccines and therapeutics have mitigated the severity of COVID-19, the effectiveness of many interventions has been progressively eroded by new variants of SARS-CoV-2. For instance, the FDA authorizations of all monoclonal antibody therapies have been revoked based on their loss of efficacy against emerging variants of concern (VOCs)[1], and the ability of first-generation vaccines and boosters to prevent infection has waned against subsequent variants[2,3]. Furthermore, immunity elicited by 2022 bivalent mRNA booster vaccines is declining against the most recent variants, such as XBB, and CH.1.1[4].

Current protocols for evaluating the effectiveness of vaccines and therapeutics primarily assess neutralization potency of antisera against previous or currently circulating SARS-CoV-2 strains, or against SARS-like betacoronaviruses (sarbecoviruses) with distant phylogenetic relationships[5–10]. These assessments, although valuable, serve as proxies for anticipating the performance of therapeutics against future viral evolution, which is likely to diverge substantially from previous variants or related viruses. Ideally, more direct methods for anticipating immune-evasive mutations can generate panels of sequences that foreshadow future viral variation against which the breadth and potency of vaccine-elicited neutralizing antibodies and therapeutics can be tested (Fig 1A).

Several experimental approaches have been developed to predict SARS-CoV-2 mutations that evade the immune response, including pseudovirus assays and high-throughput deep mutational scans[9–12]. High-throughput approaches have been used to measure the effects of all possible single mutations in the receptor binding domain (RBD)[2,3,11]. However, variants emerging during the pandemic often include multiple mutations throughout the Spike protein that evade antibody binding[13]. Recently, experiments assaying multi-mutant full Spike proteins were achieved by combining mutations frequently observed during the pandemic[12]. Emerging variants however are likely to include new mutations that evade contemporary immune responses, rather than just combinations of mutations that were already observed during the pandemic. While *in vitro* passaging in the presence of plasma or antibody therapies[9,10] has identified escape mutations, it is impractical to exhaustively assess immune escape mutations across the diverse exposure histories and interventions available. An important drawback to most experimental approaches is that they assay a modest fraction of all possible multi-mutant proteins and these approaches cannot be exploited until after immune sera or monoclonal antibodies are available.

Deep generative models have the potential to bridge these gaps by rapidly assessing large numbers of multi-mutant sequences, before immune sera is available. Computational models trained on natural sequences across evolution accurately predict mutation effects across various proteins by capturing constraints, including complex dependencies among residues[14–18]. Building on this, we recently developed EVEscape[19], which quantifies the potential of viral escape mutations using a deep learning model of sequences from distantly related viruses

combined with biophysical and structural information. EVEscape significantly outperformed previous computational and experimental methods at anticipating Spike mutations observed in the COVID-19 pandemic[19]. In this work, we extend this model framework to design a panel of SARS-CoV-2 Spike proteins representing future antigenic evolution (Fig 1B). The designed constructs can serve as panels of potential SARS-CoV-2 evolution against which vaccines and therapeutics can be evaluated. We focus on the Spike protein since it is the target of most neutralizing antibodies that confer protective immunity[20]. This generative model does not rely on high-throughput experimental scans or on three-dimensional structures of antigen-antibody complexes and is therefore generalizable to other viruses and is applicable on day-zero of a viral outbreak. Using EVEscape to design sets of sequences with high escape propensity provides a novel approach for evaluating the efficacy of vaccines and therapeutics against both existing and future viral variants.

Results

To proactively assess the effectiveness of vaccine elicited antibodies and therapeutics in protecting against rapidly evolving viral variants, it is imperative to generate and test proteins that antigenically resemble future viral strains. To design multi-mutant Spike constructs, we trained variant-specific EVEscape models on five different Spike backgrounds: B.1, BA.4/5, BA.2.12.1, BA.2.75, and XBB. We scored all possible single amino acid mutations (~121,000, Fig 1B) and identified mutations with high escape potential (top 1% EVEscape scores). Then, we calculated the escape potential of all possible double mutants (~140,000; Fig 1B), and combined the double mutants to create full Spike constructs with higher-order mutations (Methods, Supp Fig S1, Fig 1C). For each variant, we updated the EVEscape model to include pandemic sequences that occurred up to the time of emergence of that particular variant[19] (Supp Table S1). Using this approach, we designed 68 distinct Spike proteins predicted to escape polyclonal sera across mutational depths, including both rare and common mutations seen in GISAID (Supp Table S2, Supp Fig S2). Constructs contained up to 10 novel combinations of mutations relative to the nearest variant, and up to 46 mutations relative to the original B.1 strain (Supp Table S3). In total 24 unique mutations, across 21 positions, were designed in different combinations and on different Spike backgrounds. Consistent with previous EVEscape predictions for escape mutations, most designed mutations were located in primary antigenic regions seen in the pandemic[19,20] — 16 (67%) of the mutations were in the RBD and 8 (33%) were in the NTD (Fig 1D).

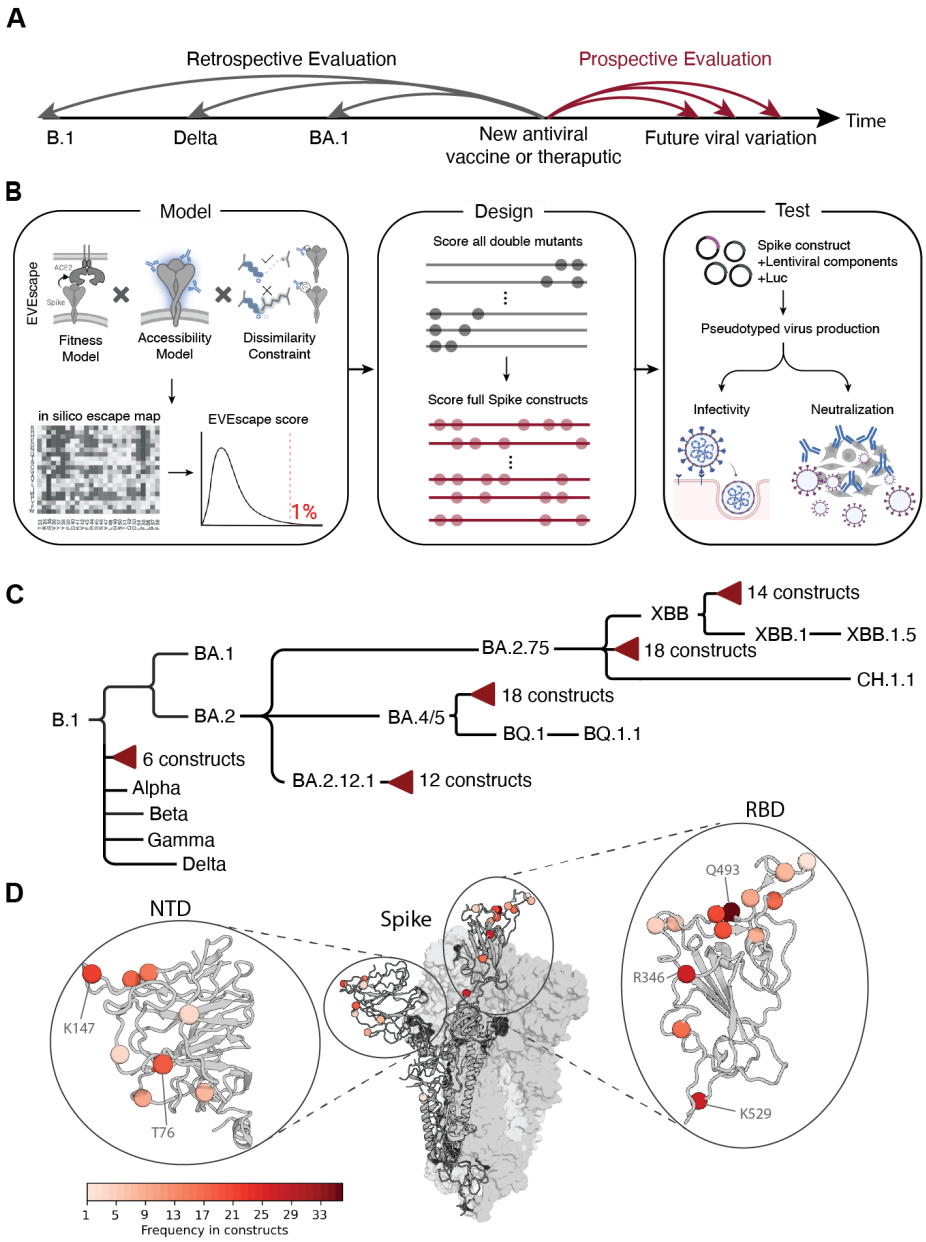


Figure 1: Proactive evaluation of vaccines and therapeutics against computationally designed future variation. (**A**) New antiviral vaccines or therapeutics can either be evaluated retroactively (against past variants or evolutionarily related viruses) or proactively (against future designed variation). For a new vaccine, retroactive evaluation would be insufficient to detail the extent to which the vaccine will remain effective in the future. We propose a proactive evaluation approach where the immunity elicited by vaccines or therapeutics are tested against computationally-predicted variants. (**B**) Future variation can be computationally designed using the EVEscape model, and the resulting multi-mutant Spike designs can then be tested for their infectivity and neutralization using Spike pseudotyped viruses. (**C**) Phylogenetic tree of variants of concern and designed constructs. Branch lengths are proportional to time. (**D**) Designed mutations across all 68 designed constructs mapped onto a representative Spike 3D structure (PDB ID: 7BNN), with coloring indicating the frequency with which a given residue was mutated in designs.

Computationally-designed multi-mutant Spikes maintain infectivity in vitro

A challenge with generating multi-mutant proteins is that most mutated proteins lose their functionality as they deviate progressively further away from the wildtype. For instance, randomly mutating the SARS-CoV-2 receptor binding domain (RBD) using error-prone PCR resulted in only ~2% of sequences with eight or more mutations being expressed[21]. Note that this is likely an overestimate of the proportion of viable full Spike sequences since mutations that are stable and well-folded in the isolated RBD may not be accommodated in the Spike trimmer.

To test infectivity of the designed Spike sequences, we used lentiviral-based pseudoviruses in single-round infection assays, as previously described[22]. Among the 68 designed constructs, 94% (64/68) were infectious (Fig 2A, Supp Table S4, Supp Fig S3A). The proportion of functional proteins achieved by computational design is higher than that achieved by combining mutations observed during the pandemic[12], where only 79% of tested Spike variants were infectious (Fig 2A). Our model provides consistent performance across mutational depths, highlighting the advantage of using deep generative models for designing multi-mutant Spike constructs.

The four designed constructs that were not infectious had complex combinations of L452R, F490R, Q493S, and Q498Y, though constructs containing some of these mutations individually, or in pairs, were infectious. For example, constructs designed on BA.2.12.1 containing the following pairs L452R and Q498Y (construct 2a), L452R and Q493S (2b), and F490R and Q498Y (7a) were infectious (Supp Fig S4A). The proximity of L452, F490, and Q493 in the three-dimensional structure suggests that steric clash between them contributes to the lack of infectivity (PDB 7BNN). This is also consistent with the observation that no variants were detected during the pandemic that contained such closely-spaced triplet mutations (Supp Fig S4B-C). This finding provides valuable insights into the enhancement of the design algorithm.

To further evaluate the impact of designed mutations on functional activity, we compared the infectivity of each variant relative to its parent variant (Fig 2B). The initial Omicron variants BA.1 and BA.2 had over 30 Spike mutations relative to B.1, concurrent with a 43.5-fold and a 25.4-fold decrease in infectivity relative to B.1 (Fig 2B, Supp Fig S3, $P < 0.05$, Mann-Whitney rank-sum test). Subsequent BA.2 sub-lineages, BA.2.12.1 (2 additional Spike mutations) and BA.2.75 (6 additional Spike mutations) recovered infectivity with a 5.6-fold and a 28.9-fold increase relative to parent variant BA.2 (Supp Fig S3, Fig 2B, $P < 0.05$, Mann-Whitney rank-sum test). Designed constructs retained comparable levels of infectivity relative to the background variant on which they were designed despite containing up to 10 mutations. On average, designed constructs had a 1.2-fold reduction in infectivity relative to their parent variant (Supp Fig S3). Collectively, these results underscore the efficacy of this approach in producing multi-mutant Spike constructs that retain *in vitro* infectivity.

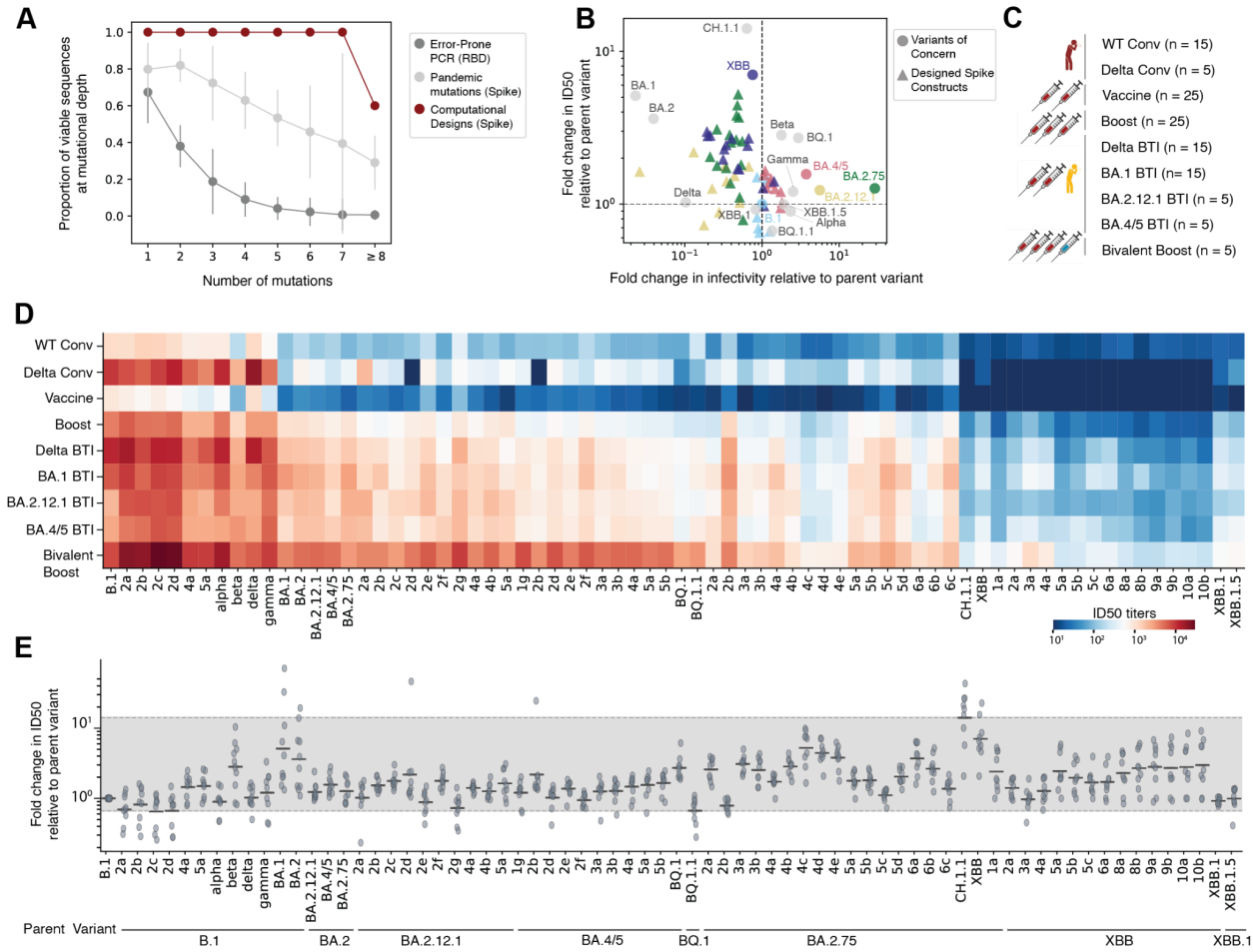


Figure 2: Designed constructs have similar levels of neutralization escape as variants of concern. (**A**) Proportion of viable sequences across mutational depths. Designed constructs are more likely to be viable than sequences resulting from error-prone PCR within the RBD (expression, Starr et al. 2020) or from combinations of pandemic mutations in Spike (infectivity, Dadonite et al. 2022). For the experimental datasets, the proportion of functional sequences was estimated by performing 10,000 samples of n sequences, where n is the number of computationally designed sequences, at a given mutational depth. Error bars represent one standard deviation. (**B**) Relationship between infectivity and neutralization. Most SARS2 variants and designed constructs that demonstrated higher levels of escape were less infectious. (**C**) Summary of serum panels from individuals who were either convalescent (Conv.), vaccinated, boosted, or experienced breakthrough infection (BTI). A total of 23 serum pools were created by grouping 5 individual patient samples per pool. (**D**) Neutralizing ID50 titers for each VOC and designed construct across nine sera panels. Reported are the geometric mean titers across all sera pools within a panel. (**E**) Fold change in ID50 relative to tested parent variant for each VOC and background variant for each design. Each point represents the geometric mean fold change for a given sera panel (nine panels in total). Lines indicate geometric mean across sera panels. The gray band indicates the observed changes in neutralization sensitivity across variants of concern.

Spikes designed on early variants exhibit profiles of neutralization resistance similar to subsequent SARS-CoV-2 variants

The continued efficacy of vaccines and therapeutics in part depends on the ability of elicited antibodies to retain the capacity to neutralize newly emerging viral variants. To assess the relevance of the designed Spike constructs as representative of future viral variation, we performed pseudovirus neutralization assays to evaluate sensitivity profiles when tested against nine panels of polyclonal serum pools representative of diverse SARS-CoV-2 immune histories: (1) individuals with convalescent infection prior to the availability of vaccines, (2) unvaccinated individuals with convalescent Delta infection, (3) recipients of two doses of mRNA-1273 or BNT162b2 vaccines, (4) individuals who received three doses of an mRNA vaccine, (5) vaccinated individuals administered a bi-valent booster (B.1 and BA.4/5), and vaccinated patients who experienced breakthrough infections (BTI) with (6) Delta, (7) BA.1, (8) BA.2.12.1, or (9) BA.4/5 variants (Fig 2C, Supp Table S5). Serum panels were tested against pseudoviruses expressing 56 of the designed Spike constructs, 16 SARS-CoV-2 variant Spike proteins, and the SARS-CoV-1 Spike protein (Fig 2D, Supp Table S6).

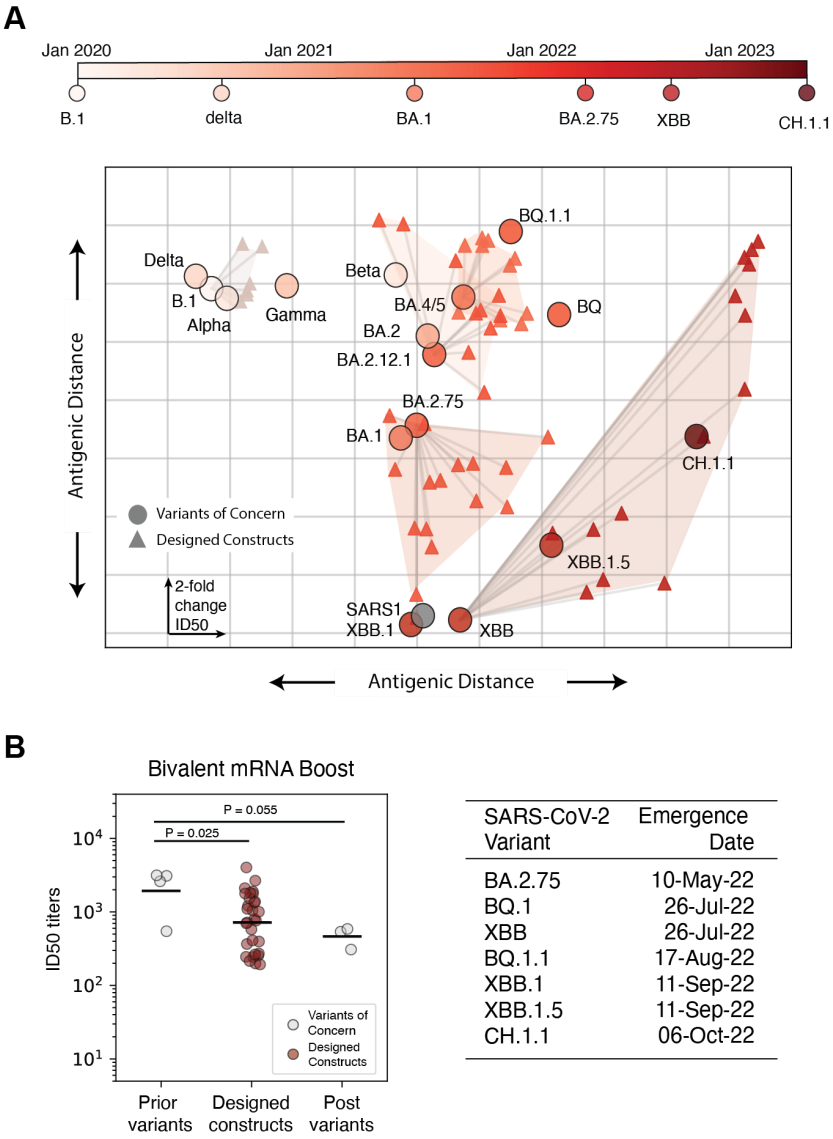
To characterize the degree of immune-evasion observed throughout the pandemic, we compared the half-maximal neutralizing antibody titers (ID_{50}) for each variant relative to the parent variant that preceded it (Fig 2E). As expected, most variants were less susceptible to neutralization compared to its parent variant, with the exception of alpha and BQ.1.1 which had a comparable geometric mean ID_{50} titer across sera panels to their parent variants B.1 and BQ.1, respectively ($P > 0.2$, Mann-Whitney rank-sum test). The XBB and CH.1.1 variants displayed the highest level of escape relative to their parent variant BA.2.75, with a 7.02-fold ($P < 0.01$) and 14.10-fold ($P < 0.01$) reduction, respectively, across sera panels. On average emerging variants exhibited a 1.96-fold reduction in geometric mean ID_{50} titers relative to their parent variant, and a range from 0.67 to 14.10 (gray region, Fig 2E). Generally, variants that had higher levels of antibody escape tended to be less infectious than their parent variant (Fig 2B).

Pseudoviruses expressing designed Spikes had comparable levels of antibody escape as SARS-CoV-2 variants that have evolved throughout the pandemic. On average, designed Spikes had a 1.68-fold reduction (range of 0.65 to 5.24) in geometric mean ID_{50} titer compared to the parent variant on which they were designed (Fig 2E), comparable to the 1.96-fold reduction seen amongst VOCs in relation to their parent variant. Designs on specific backgrounds had similar levels of neutralization escape as SARS-CoV-2 variants evolved from those same backgrounds. For instance, the BA.4/5-2b designed construct, containing the R346T and S494R mutations on the BA.4/5 background, had a 2.18-fold reduction in geometric ID_{50} titers, which is comparable to the level of escape observed for BQ.1 (2.71-fold reduction relative to BA.4/5). Similarly, BA.2.75-4c, a variant designed on BA.2.75 containing four mutations (G339D, L452R, Q493R, K529L), had a 5.24-fold reduction in geometric mean ID_{50} titer relative to BA.2.75, slightly lower than the level of escape observed by XBB (7.02-fold reduction in geometric mean titers relative to BA.2.75). Taken together these results demonstrate that Spike constructs designed using EVEScape exhibit a similar level of antibody escape as variants that have naturally evolved under immune pressure throughout the

pandemic and can therefore serve as useful proxies for future SARS-CoV-2 evolution and guid early evaluation of vaccines and therapeutics.

Designed variants foreshadow future antigenic evolution

Thus far we have demonstrated that our designed Spike constructs are comparable to SARS-CoV-2 variants in both infectivity and neutralizing antibody escape. However, it remains unclear whether these constructs accurately reflect the antigenic evolution observed throughout the pandemic. To investigate this, we constructed antigenic maps from the neutralization data using a multidimensional scaling (MDS) algorithm to characterize the space of neutralizing antibody responses across variants (Fig 3A). The resulting antigenic cartography provides a two-dimensional representation such that distance between any pair of variants is related to the similarity in which they are neutralized by the sera pools. Notably, constructs designed on earlier variants consistently exhibited an antigenic resemblance to variants that emerged later in the pandemic. For example, constructs designed on BA.2.75 were antigenically similar to the XBB lineages which emerged nine months later (Fig 3A, shaded area). Similarly, constructs designed on the XBB variant exhibited antigenic responses akin to the CH.1.1 variant, which emerged three months later (Fig 3A, shaded area). This finding holds for all other rounds, with designs on BA.2.12.1 foreshadowing BQ, designs on BA.4./5 foreshadowing BQ.1.1, and designs on B.1 foreshadowing the Alpha and Gamma variants (Fig 3A). Together these findings underscore the model's capability to design constructs with antigenic profiles akin to those of future variants and showcase its potential in predicting antigenic evolution using only data available at the time of VOC emergence.



SARS-CoV-2 Variant	Emergence Date
BA.2.75	10-May-22
BQ.1	26-Jul-22
XBB	26-Jul-22
BQ.1.1	17-Aug-22
XBB.1	11-Sep-22
XBB.1.5	11-Sep-22
CH.1.1	06-Oct-22

Figure 3: Designed constructs foreshadow future SARS-CoV-2 evolution. (**A**) Antigenic cartography map depicting two-dimensional representation of the pairwise antigenic distance between each variant of concern (circles) or designed construct (triangles), as dictated by how each is neutralized by the sera pools. The dimensions of each box represents a 2-fold change in neutralizing ID50 titers. Color indicates time during the pandemic. Constructs designed on earlier variants (triangles of lighter colors) foreshadow the antigenicity of later pandemic variants (circles of darker colors), as indicated by proximity in the antigenic map. (**B**) Neutralizing ID50 titers in sera from individuals who received three doses of mRNA-1273 or BNT162b2 and a fourth dose of the Bi-valent (BA.4/5 and B.1) vaccine (lines indicate geometric mean across variants or designs). The bi-valent booster vaccine was evaluated on its ability to neutralize variants circulating in the fourth months prior to the creation of the bi-valent booster vaccines (BA.2.75, BQ.1, BQ.1.1, and XBB), which was approved on Aug 31, 2022. Emergence dates were estimated from nextstrain.org. Designed Spike constructs on the background of variants circulating prior to the vaccine foreshadowed the levels of antibody escape observed against variants that emerged post the bi-valent booster (XBB.1, XBB.1.5, CH.1.1, Mann-Whitney rank-sum test).

Evaluating the breadth and potency of mRNA vaccine-elicited neutralizing antibodies
To demonstrate the utility of the proposed approach for proactive evaluation of vaccine candidates, we showcase how designed variants could have been utilized to evaluate the ability of neutralizing antibodies elicited by the bi-valent booster vaccines (B.1 & BA.4/5) to protect against future evolving variants. Considering variants circulating in the four months prior to the implementation of the bivalent booster (BA.2.75, BQ.1, BQ.1.1, and XBB), the geometric mean ID_{50} titers for serum panels representative of bi-valent boosting (4 shots) were 1,931 (Fig 3B, Supp Table S8). Constructs designed on the variants circulating at the time of the booster vaccination campaign (designs on BA.2.75 and XBB) demonstrated a range of antibody escape with ID_{50} ranging from 193 to 4,029, and had an average 2.6-fold decrease in geometric mean titers relative to the VOCs present prior to the bi-valent vaccine ($P = 0.025$, Mann-Whitney rank-sum test, Fig 3B). Notably, levels of antibody escape seen in variants that evolved in the four months after the implementation of the bivalent boosters XBB.1, XBB.1.5 and CH.1.1 (ID_{50} titers of 538, 587, and 308 respectively) fall within the distribution of titers observed in the designed constructs (Fig 3B). Therefore, designed constructs on variants circulating at the time of the bivalent B.1 & BA.4/5 booster vaccination campaign could have highlighted the potential for antibody escape observed in subsequent lineages. Furthermore, eight of the designed constructs on XBB had lower ID_{50} titers than CH.1.1, demonstrating the potential for further escape from neutralizing antibodies elicited by bivalent booster vaccines.

Discussion

Driven by the need to develop vaccines and therapeutics that remain effective in the face of viral evolution, almost 400 SARS-CoV-2 vaccine candidates are in preclinical or clinical trials[23]. However, it is currently unclear how to proactively evaluate the ability of candidate vaccines to remain effective against future variants. We demonstrate that deep learning models of escape from neutralizing antibodies can generate panels of Spike constructs that foreshadow the antigenic landscape of future SARS-CoV-2 variants. Panels of designed Spike constructs can be used to assess the ability of candidate vaccines and therapeutics to neutralize future viral variants. We highlight the utility of our approach through a retrospective analysis of the bi-valent booster vaccine and show that the level of immune-escape observed by currently circulating SARS-CoV-2 variants, such as XBB and CH.1.1, could have been anticipated at the time that bi-valent booster vaccines were implemented. Furthermore, the designed constructs demonstrate the potential for even further immune escape from antibodies elicited by the 2022 bi-valent mRNA booster vaccine than observed by the latest SARS-CoV-2 variants.

The approach presented here is generally applicable for evaluating pan-coronavirus vaccines and therapeutics and is applicable to other rapidly evolving viruses. An advantage of the modeling framework is its adaptability across viral families, particularly viruses with limited data, such as Lassa virus[19]. Additionally, the design algorithm can be used to evaluate antibody therapeutics by targeting mutations within a particular epitopes. Moreover, while antibody neutralization is likely a critical driver of initial clinical efficacy[24], ideally a similar approach could be developed for predicting and evaluating longer term T-cell responses. Perhaps most

importantly, an approach could be developed to evaluate not only immune escape but also the efficacy of vaccines and therapeutics for preventing severe disease.

To date, cross-neutralization potential of pan-variant therapeutics[5,25], as well as SARS-CoV-2 boosters[26] and other antivirals[27], have been evaluated against previous or contemporary SARS-CoV-2 variants or against distantly-related sarbecoviruses[7,8]. Such approaches are limited to retrospectively testing the diversity seen in the pandemic or throughout evolution, rather than proactively measuring protection against future viral strains. The vaccine and therapeutic evaluation approach presented here addresses these issues by anticipating mutations that will escape current antibody repertoires, generating pseudotyped viruses containing complex mutations, and evaluating the ability of vaccines and therapeutics to elicit antibodies that not only neutralize previous and circulating variants but also future viral evolution.

Acknowledgement

The authors would like to thank Amy Sherman, Stephen Walsh, and Lindsey Baden at Brigham and Women's Hospital (Boston, MA) for providing serum from vaccinated individuals for this study, Nikki Thadani, Rachel Liang, and members of the Luban, Lemieux, Seaman and Marks; labs. This work was supported by grants from the Coalition for Epidemic Preparedness Innovations (CEPI) and the Massachusetts Consortium on Pathogen Readiness.

Methods

Computational Spike protein design

Multiple Sequence Alignments for fitness models:
We designed Spike constructs on the backgrounds of five VOCs: the canonical wildtype SARS-CoV-2 Spike protein (Uniprot ID: P0DTC2) with D614G mutation as the query sequence, and the BA.4/5, BA.2.12.1, BA.2.75, and XBB Spike sequences. We define the pandemic lineages based on the list of mutations described in outbreak.info [28]). For each round of construct design, we built multiple sequence alignments for Spike with both pre-pandemic sequences and pandemic sequences seen before the emergence of the variant used as the background for the design. To generate the pre-pandemic alignment, we used 5 iterations of the jackhmmer iterative HMM-based homology search alignment tool [29], searching against the Uniref100 dataset [30] downloaded on March 16, 2022. To account for the growing number of pandemic sequences, we also included unique pandemic sequences seen more than 100 times in the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database (www.gisaid.org; [31] at the time of emergence of the variant (Supp Table S1). The resulting pandemic sequences and evolutionary alignments were combined and aligned with super5 [32], version 5.1). We remove sequences that are aligned to less than 50% of the query sequence and remove residue positions that contain more than 70% gap characters.

Modeling approach:
We use the EVEscape model [19], which combines fitness, accessibility, and dissimilarity calculations, to predict a mutant's ability to escape antibody recognition. For the fitness component we trained an EVE model [15], which is a Bayesian variational autoencoder (VAE), to predict the fitness scores. We down weighted redundant sequence clusters by assigning each protein sequence i a weight $w(i) = 1/T$, where to T is the number of sequences in the alignment within a given hamming distance cutoff of t. We combine the pre-pandemic and pandemic alignments and use a t value of 0.01. We estimate the fitness of each sequence as the difference in the negative Evidence Lower Bound (ELBO), an approximation to the log-likelihood, of that sequence relative to the query sequence. We estimate ELBO values using twenty thousand Monte Carlo samples of the latent space.

For the accessibility component, we selected Spike structures from the RCSB Protein Data Bank (PDB) representative of both "open" and "closed" configurations (PDB IDs: 6VXX, 6VYB, 7CAB, and 7BNN). Using these structures, we calculated the weighted contact number WCN for each residue position i as

$$WCN_i = \sum_{j \neq i} r_{ij}^{-2}$$

where r_{ij} is the distance between the geometric centers of the side chain of the residues occupying sites i and j. Each position is then assigned the minimum weighted contact number across all structures, and we use the negative weighted contact number.

For the dissimilarity component, we use both charge and hydrophobicity [33] differences between each pair of amino acids and then assign each pair a dissimilarity value equal to the sum of the standard-scaled differences.

We combine the mutation-level fitness (f), accessibility (a), and dissimilarity (d) using a temperature-scaled logistic function that calibrates the three factors by dividing by a temperature parameter to get a single escape score for each individual mutation. The score of a mutation m is calculated as

$$EVEscape(m) = logistic(T_f^{-1} * F_f^m) * logistic(T_a^{-1} * F_a^m) * logistic(T_d^{-1} * F_d^m)$$

where T_i is the temperature scaling and F_i is the standardized vector for factor i, and we then take the log transform of the product. We modified the temperature parameters by increasing the contribution of the fitness component, $T_f = 1$, $T_a = 8$, $T_d = 16$, compared to the values used previously [19].

Designs:
We designed 68 constructs on the backgrounds of five different Spike VOC sequences: B.1, BA.5, BA.2.12.1, BA.2.75, and XBB. We selected between 2 and 10 mutations to add to each background, by first scoring all possible single amino acid substitutions to Spike. Note that we consider amino acid mutations involving any number of nucleotide changes rather than focusing on only single nucleotide changes which are by far the most common (more than 99%)

mutations seen in GISAID [31], as we are not trying to predict exact future mutations with designs. To design constructs, we considered single amino acid mutations in the top 1% from our complete EVEscape score. We then ranked each of the unique double mutations based on their EVEscape score alone. Ultimately, this resulted in designs, spanning 24 unique mutations across 21 positions, in different combinations and on different VOC backgrounds (Supp Table S1-S2).

Testing Infectivity

DNA sequence design and plasmid:
We designed a pDMJ2 plasmid where Spike expression is under the control of a cytomegalovirus immediate early promoter (CMV-IE) fused to the human T lymphotropic virus type 1 (HTLV-1) 5' UTR [34]. Transcription termination is directed by a bovine growth hormone polyadenylation signal (bGH-PolyA). The pDMJ2 plasmid carries an origin of replication (ori) from ColE1 and an Ampicillin resistance gene. Spike sequences were inserted between the *XbaI* and *BamHI* restriction sites (Supp Table S3; Supp Fig S5).

An ancestral B.1 with D614G Spike sequence was codon optimized by Twist Biosciences with a deletion of the terminal 18 amino acids to remove the endoplasmic reticulum retention signal [35]. This sequence was used as the base sequence on which other constructs were designed to minimize effects of nucleotide sequence on protein synthesis efficiency. Specifically, to generate Spike sequences of VOCs and designed constructs, we mutated individual codon positions to the most commonly used codon in human cells encoding the mutant amino acid. We repeated this for all mutations in a given sequence. If a restriction site was introduced, the second most common codon was used and so on.

Infectivity assay:
HIV-1-derived virions bearing a luciferase reporter gene, and pseudotyped with SARS-CoV-2 Spike, were produced by transfection of HEK293T cells, as previously reported [22]. The supernatant from the transfected cells was collected 72 hrs post-transfection and filtered (0.45 µM, Avantor™) to remove cellular debris. Virion yield in each transfection supernatant was normalized using our in-house RT assay [22,36]. Vector supernatants were then placed on HEK293T cells that were previously transduced with puromycin- and blasticidin-resistant lentivectors to stably express the human angiotensin converting enzyme 2 (ACE2) and the transmembrane serine protease 2 (TMPRSS2) [22], respectively. Luciferase activity on target cells was measured as an indication of transduction efficiency, using the Steady-Glo® Luciferase Assay System (Promega Corporation) read on a Promega GloMax Discover machine. All plasmids were deposited to Addgene (Supp Table S4).

Comparison to experimental assays:
We used data from recent deep mutation scans to quantify the likelihood of obtaining functional sequences with multiple mutations compared to a wild type sequence. We analyzed the RBD data from [21] and the full Spike data from [12]. In the RBD dataset, mutations were introduced using error-prone PCR resulting in 135,386 unique mutant RBDs with up to 11 amino acid mutations.

The expression of each sequence was measured as the change in mean fluorescence intensity, Δlog(MFI), relative to the unmutated SARS-CoV-2 RBD. We used sequences containing synonymous mutations (labeled viable) and sequences with premature stop codons (labeled nonviable) to train a logistic regression for classifying missense variants as either viable or nonviable. The fitted model had an intercept of 4.66 and a coefficient of 1.39. The decision threshold estimated for expression scores was -1.56 Δlog(MFI) with variants having expression values less than this boundary classified as non-viable, and variants with expression scores greater than the boundary classified as viable. [37] used an expression threshold of -1 Δlog(MFI) to classify variants that have as large an expression deficit as mutations to core disulfide residues as nonviable. More recently, [11] used an expression threshold of -1.5 Δlog(MFI). Our estimated decision threshold is in-line with previous biologically-informed cutoffs.

We repeated the logistic regression analysis described above on the full Spike dataset from [12]. In this dataset, only function-preserving mutations seen in the pandemic were introduced to the full Spike in higher-order combinations. The dataset included 191,418 unique variants containing up to 38 mutations. For each variant, a functional score was estimated based on its relative frequency in the Spike- versus VSV-G-pseudotyped libraries. The fitted model had an intercept of 6.11 and a coefficient of 3.91. The decision threshold was estimated to be a functional score of -3.34.

Testing Neutralization

Sera Collection and Pooling:
We collected 115 individual serum samples from individuals who were either convalescent, vaccinated, boosted, or experienced breakthrough infection (Supp Table S6). Samples were acquired from the POSITIVES study whose collections and primary outcomes have been reported elsewhere or through local Boston collaborations following institutional IRB approved protocols and with patient informed consent [38,39]. For patients experiencing breakthrough infection, whole genome sequencing was performed on a nasal swab collected at the time of diagnosis to confirm infecting SARS-CoV-2 variant. Serum samples were acquired approximately 2-4 weeks after vaccination or convalescence. In order to maximize the number of designed Spike variant pseudoviruses that can be tested against the same immune serum panels, a total of 23 serum pools were created by grouping 5 individual patient samples per pool. The pooled sera are representative of the different populations that have existed throughout the pandemic: natural infection, primary vaccination, boosted, bivalent boosted, and breakthrough infection. When sufficient samples were available, individual serum samples were pooled based on neutralization titers resulting in low, medium, and high neutralization pools.

Neutralization Assay:
Neutralizing activity against SARS-CoV-2 pseudovirus was measured using a single-round infection assay in 293T/ACE2 target cells. Pseudotyped virus particles were produced in 293T/17 cells (ATCC) by co-transfection of plasmids encoding codon-optimized SARS-CoV-2 Spike (containing G at position 614), packaging plasmid pCMV R8.2, and luciferase reporter plasmid pHR' CMV-Luc. Packaging and luciferase plasmids were kindly provided by Dr. Barney

Graham (NIH, Vaccine Research Center). The 293T cell line stably overexpressing the human ACE2 cell surface receptor protein was kindly provided by Drs. Michael Farzan and Huihui Ma (The Scripps Research Institute). For neutralization assays, serial dilutions of patient serum samples were performed in duplicate followed by addition of pseudovirus. Pooled serum samples from convalescent COVID-19 patients or pre-pandemic normal healthy serum (NHS) were used as positive and negative controls, respectively. Plates were incubated for 1 hour at 37 C followed by addition of 293/ACE2 target cells ($1x10^4$ /well). Wells containing cells + pseudovirus (without sample) or cells alone acted as positive and negative infection controls, respectively. Assays were harvested on day 3 using Promega BrightGlo luciferase reagent and luminescence detected with a Promega GloMax luminometer. Titers are reported as the dilution of serum that inhibited 50% or 80% virus infection (ID_{50} and ID_{80} titers, respectively).

Antigenic cartography:
We use Racmacs [40] to perform multidimensional scaling (MDS) on our titration data measuring the neutralization sensitivity of all VOCs and designs for each serum pool. MDS is used to create a 2-dimensional representation of the pairwise antigenic distance between each VOC or design, as dictated by how each is neutralized by the different sera pools. Each box represents a distance between two VOCs or designs corresponding to a 2-fold change in ID_{50}. This allows us to visualize the trajectory of evolution in antigenic space, and showcase antigenic similarity between designs and later pandemic strains.

References

1. Kip, K. E. *et al.* Evolving Real-World Effectiveness of Monoclonal Antibodies for Treatment of COVID-19 : A Cohort Study. *Ann. Intern. Med.* **176**, 496–504 (2023).
2. Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).
3. Cao, Y. *et al.* BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **608**, 593–602 (2022).
4. Scheaffer, S. M. *et al.* Bivalent SARS-CoV-2 mRNA vaccines increase breadth of neutralization and protect against the BA.5 Omicron variant in mice. *Nat. Med.* **29**, 247–257 (2023).
5. Liu, Z. *et al.* A pan-sarbecovirus vaccine based on RBD of SARS-CoV-2 original strain elicits potent neutralizing antibodies against XBB in non-human primates. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2221713120 (2023).
6. Li, D. *et al.* Breadth of SARS-CoV-2 neutralization and protection induced by a nanoparticle vaccine. *Nat. Commun.* **13**, 6309 (2022).
7. Cohen, A. A. *et al.* Mosaic RBD nanoparticles protect against challenge by diverse sarbecoviruses in animal models. *Science* **377**, eabq0839 (2022).
8. Cohen, A. A. *et al.* Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021).
9. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369**, 1014–1018 (2020).
10. Schmidt, F. *et al.* Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, (2020).

11. Greaney, A. J. *et al.* Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **29**, 44–57.e9 (2021).
12. Dadonaite, B. *et al.* A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *bioRxiv* 2022.10.13.512056 (2022) doi:10.1101/2022.10.13.512056.
13. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
14. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018).
15. Frazer, J. *et al.* Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021).
16. Rives, A. *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
17. Hopf, T. A. *et al.* Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
18. Shin, J.-E. *et al.* Protein design and variant prediction using autoregressive generative models. *Nat. Commun.* **12**, 2403 (2021).
19. Thadani, N. N. *et al.* Learning from pre-pandemic data to forecast viral escape. *bioRxiv* 2022.07.21.501023 (2023) doi:10.1101/2022.07.21.501023.
20. Piccoli, L. *et al.* Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).
21. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295–1310.e20 (2020).
22. Yurkovetskiy, L. *et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739–751.e8 (2020).
23. COVID-19 vaccine tracker and landscape. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines.
24. Pascal, K. E. *et al.* Pre- and postexposure efficacy of fully human antibodies against Spike protein in a novel humanized mouse model of MERS-CoV infection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8738–8743 (2015).
25. Tan, C.-W. *et al.* Pan-Sarbecovirus Neutralizing Antibodies in BNT162b2-Immunized SARS-CoV-1 Survivors. *N. Engl. J. Med.* **385**, 1401–1406 (2021).
26. Chalkias, S. *et al.* A Bivalent Omicron-Containing Booster Vaccine against Covid-19. *N. Engl. J. Med.* **387**, 1279–1291 (2022).
27. Pitts, J. *et al.* Remdesivir and GS-441524 Retain Antiviral Activity against Delta, Omicron, and Other Emergent SARS-CoV-2 Variants. *Antimicrob. Agents Chemother.* **66**, e0022222 (2022).
28. Gangavarapu, K. *et al.* Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nat. Methods* **20**, 512–522 (2023).
29. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
30. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
31. Khare, S. *et al.* GISAID's Role in Pandemic Response. *China CDC Wkly* **3**, 1049–1051 (2021).
32. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* **13**, 6968 (2022).
33. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 140–144 (1984).
34. Takebe, Y. *et al.* SR alpha promoter: an efficient and versatile mammalian cDNA expression

system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. *Mol. Cell. Biol.* **8**, 466–472 (1988).
35. Ujike, M., Huang, C., Shirato, K., Makino, S. & Taguchi, F. The contribution of the cytoplasmic retrieval signal of severe acute respiratory syndrome coronavirus to intracellular accumulation of S proteins and incorporation of S protein into virus-like particles. *J. Gen. Virol.* **97**, 1853–1864 (2016).
36. Lai, R. P. J. *et al.* Nef decreases HIV-1 sensitivity to neutralizing antibodies that target the membrane-proximal external region of TMgp41. *PLoS Pathog.* **7**, e1002442 (2011).
37. Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476.e6 (2021).
38. Boucau, J. *et al.* Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection. *N. Engl. J. Med.* **387**, 275–277 (2022).
39. Seaman, M. S. *et al.* Vaccine breakthrough infection leads to distinct profiles of neutralizing antibody responses by SARS-CoV-2 variant. *JCI Insight* **7**, (2022).
40. Wilks, S. Racmacs: R antigenic cartography macros. Preprint at (2021).

A

	B.1						BA.4/5										BA.2.12.1										BA.2.75													XBB											
	2a	2b	2c	2d	4a	5a	2a	2b	2c	2d	2e	2f	3a	3b	4a	5a	5b	2a	2b	2c	2d	2e	2f	2g	4a	4b	5a	7a	8a	9a	2a	2b	3a	3b	4a	4b	4c	4d	4e	5a	5b	5c	5d	6a	6b	6c	8a	9a	2a	3a	4a	5a	5b	5c	6a	8a	8b	9a	9b	10a	10b
T19																																																														
A27										L						L									L																		L									L								
T76					I	I					I	I		I	I	I	I								I	I									I					I			I								I	I							
K147				N																			N				N	N	N	N						N					N	N			N	N		N					N	N	N	N	N	N	N	N
M153					S	S						S		S	S	S	S								S	S									S					S			S								S	S							
S155																								R			R	R	R	R																		R	R					R	R	R	R	R	R	R	R
Q183																																																			V							V		V
V213																																																	G	G	G				G	G	G	G	G	G
G339																																D		D			D	D				D		D	D				D	D	D				D	D	D	D	D	D
R346	T	T	T				T	T												T								T		T	T		T					T	T		T				T	T	T	T						K	K		K	K	K	K
K417			V	V																												V									V	V		V	V	V															
V445																																																		E	E					E			E	
L452																		R	R										R	R						R	R		R					R		R	R	R				R	R	R	R	R	R	R	R
Y473					F	F			F				F		F	F	F								F	F								F					F			F								F	F								
T478	K																																																											
N481						T							T				T									T														T												T								
F490																					R							R	R	R									R								R	R	R											
Q493					R	R			R	K	R		R	R	R	R	R		S										S	S			R	R	R	R	R		R		R	R	R	R		S	S			R	R	R	R	R	R	R	R	R	R
S494		R						R							R	R	R	R			R	R	R				R	R	R	R	R		R								R						R	R												
Q498																		Y									Y	Y	Y	Y																	Y	Y												
K529																						L					L	L	L	L			L	L		L	L	L	L		L	L		L	L	L	L	L					L	L	L	L	L	L	L	L

B

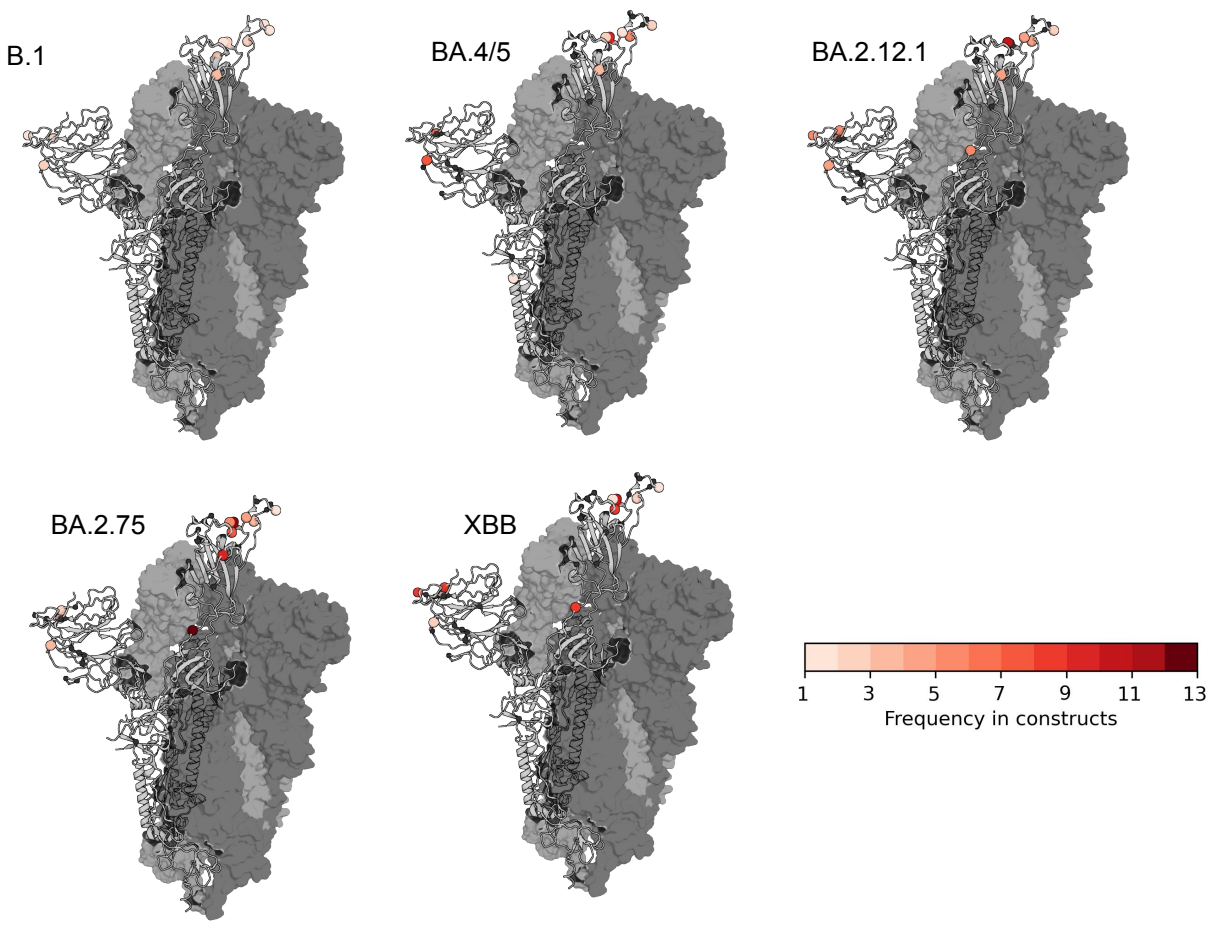


Figure S1: Overview of designed sequences. (A) Summary of designed spike sequences with two or more mutations on five variant backgrounds (B.1, BA.4/5, BA.2.12.1, BA.2.75, and XBB). (**B**) Designed mutations across designed constructs in each of the five rounds mapped onto a representative Spike 3D structure (PDB ID: 7BNN), with coloring indicating the frequency with which a given residue was mutated in that round.

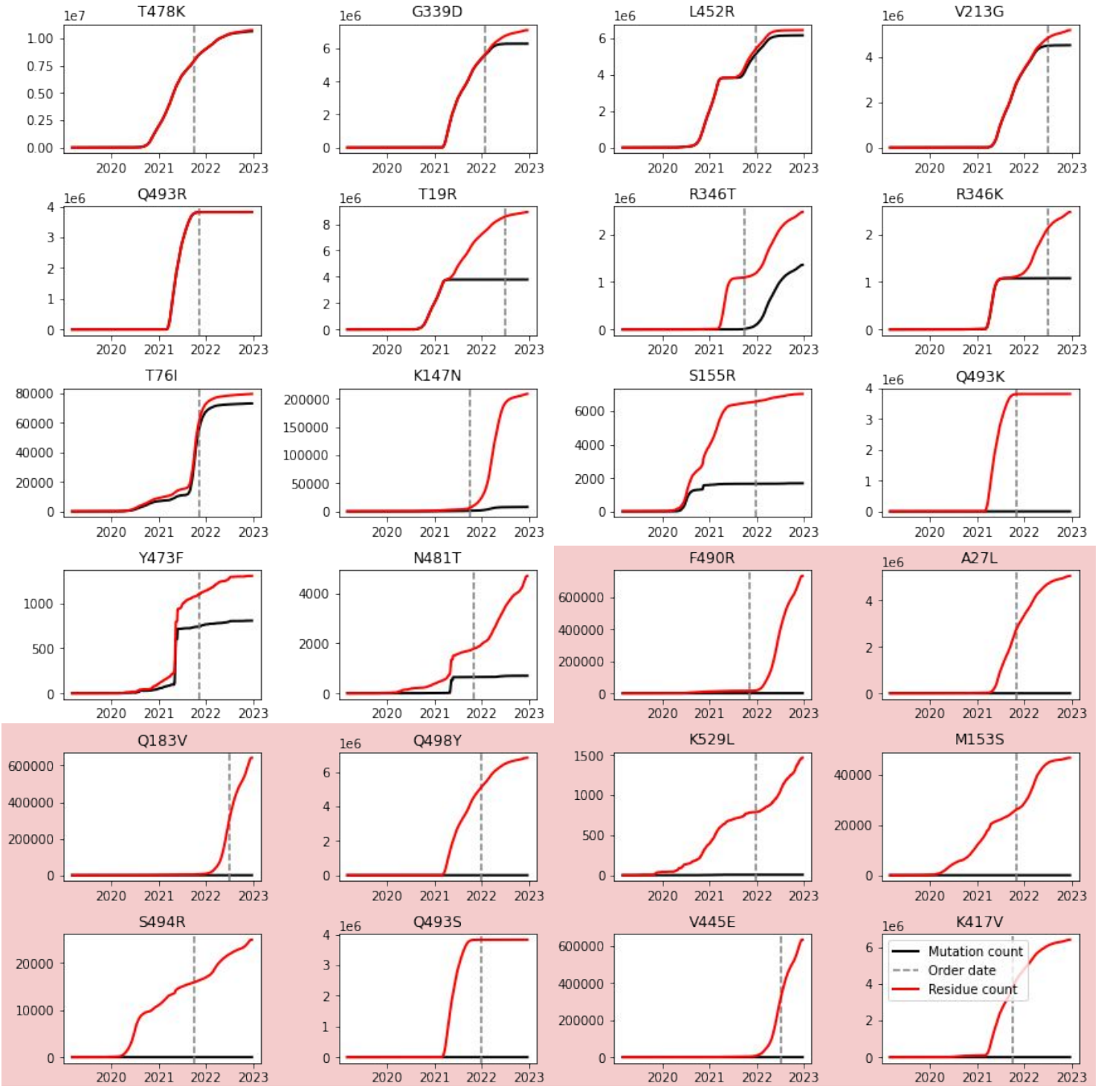
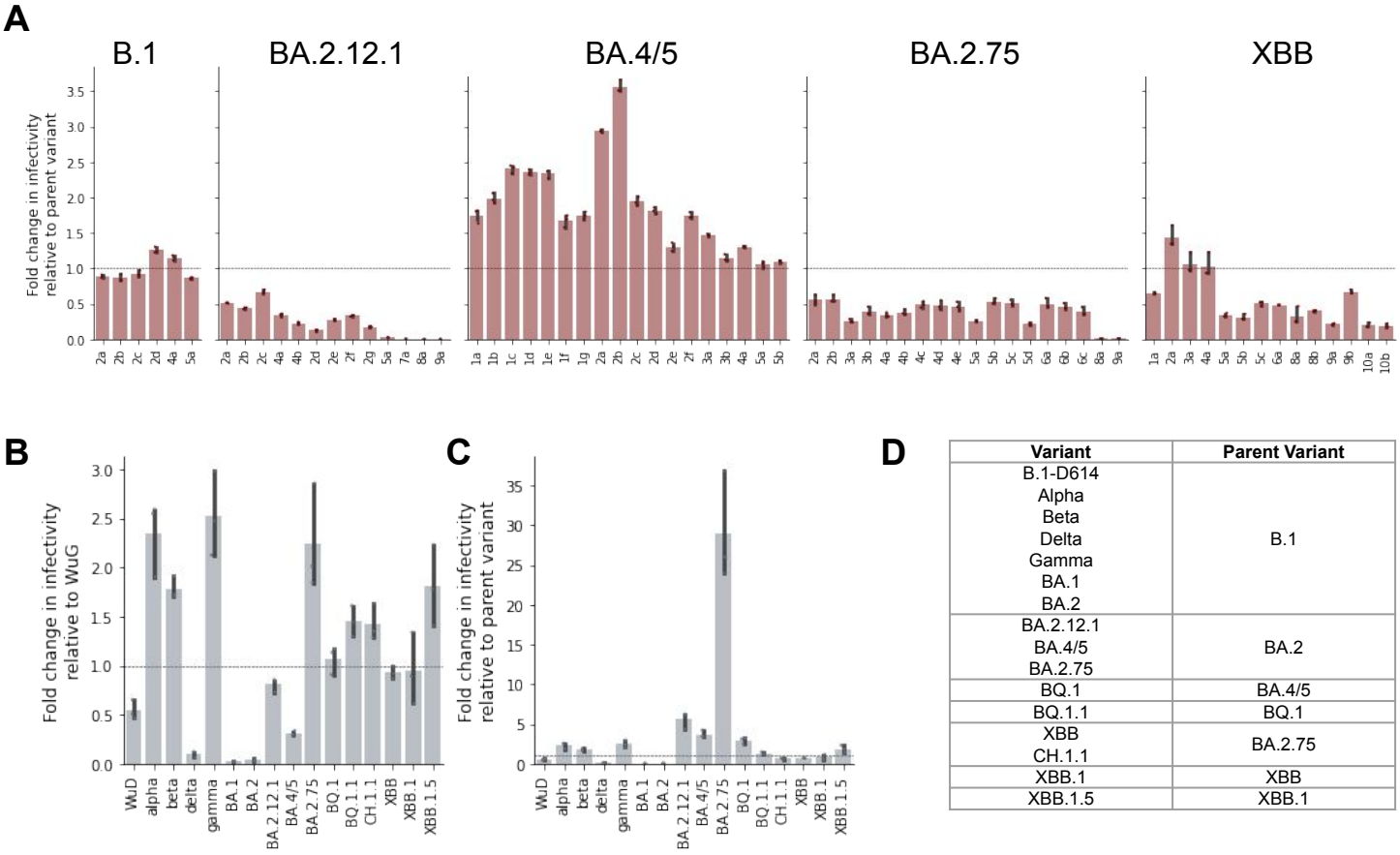


Figure S2: Count of designed mutation throughout the COVID19 pandemic. Black line shows cumulative mutation count in GISAID. Red line shows cumulative count of any mutations at that particular residue. Gray dashed line shows data construct containing that particular mutation was ordered. As expected, designed mutations which necessitate more than one nucleotide change (red background), were rare in the pandemic. However, the corresponding residue positions were highly mutated during the pandemic, highlighting the model's ability to identify mutable positions.



Variant	Parent Variant
B.1-D614 Alpha Beta Delta Gamma BA.1 BA.2	B.1
BA.2.12.1 BA.4/5 BA.2.75	BA.2
BQ.1	BA.4/5
BQ.1.1	BQ.1
XBB CH.1.1	BA.2.75
XBB.1	XBB
XBB.1.5	XBB.1

Figure S3: Infectivity Data. (A) Fold change in infectivity of designed constructs relative to the base variant on which they were designed. **(B)** Fold change in infectivity of variants of concern (VOCs) relative to B.1 . For consistency, we also compare the infectivity of the B.1. **(C)** Fold change in infectivity of variants of concern (VOCs) relative to their patient variant. **(D)** Summary of parent variants for VOCs.

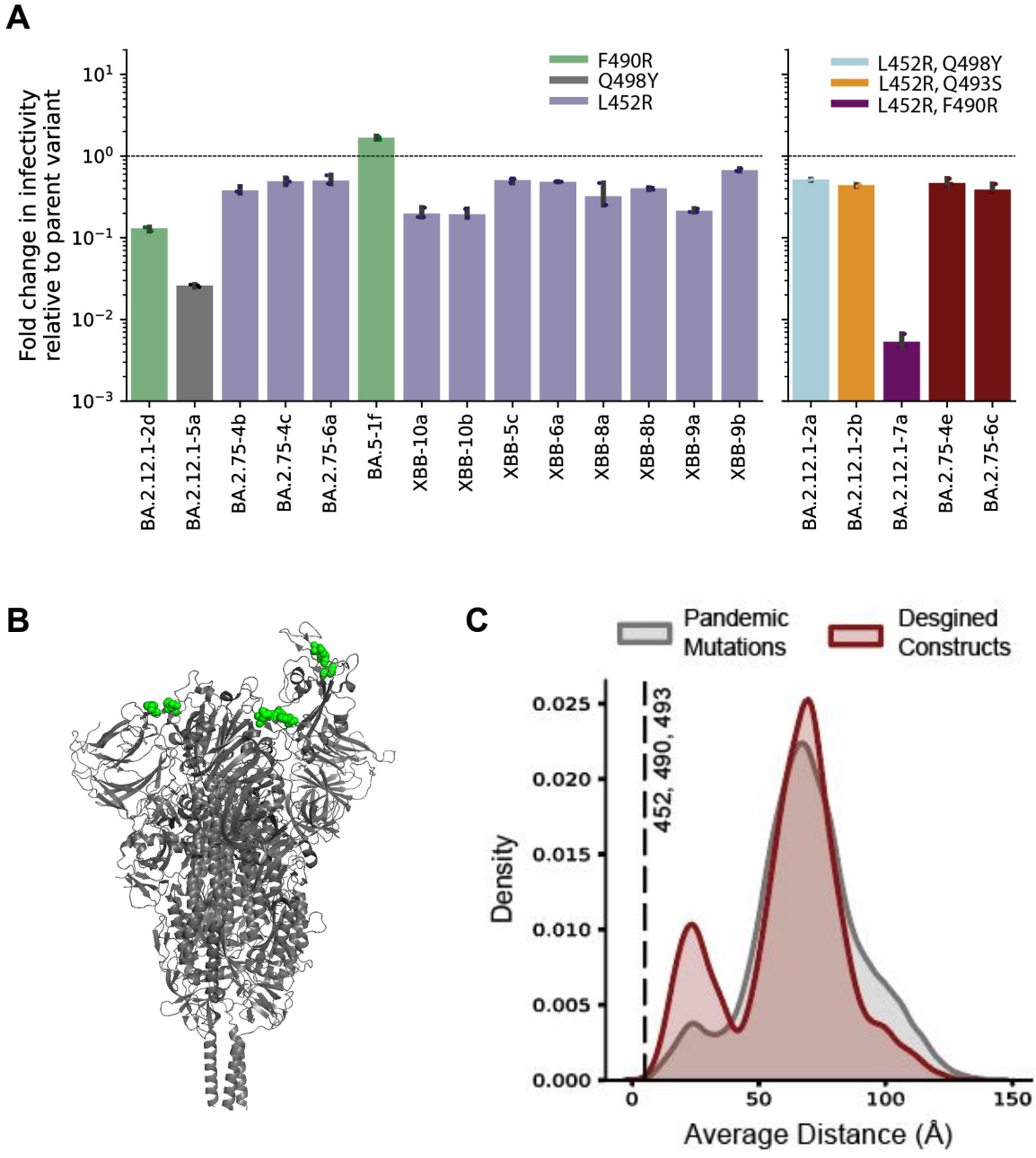


Figure S4: Proximity in three-dimensional structure suggests that the lack of infectivity may be explained by physical residue interactions. (**A**) Constructs containing some of these mutations individually, or in pairs, remain infectious. (**B**) The four designed constructs which were not infectious shared a unique subset of mutations: L452R, F490R, Q493S, and Q498Y. Three proximal residues (452, 490, 493) are shown on Spike structure (PDB ID: 7BNN). (**C**) The three residues 452, 490, and 493 are closer in average distance than the average distance of any triplet of mutations in any pandemic strains or in any other designed constructs.

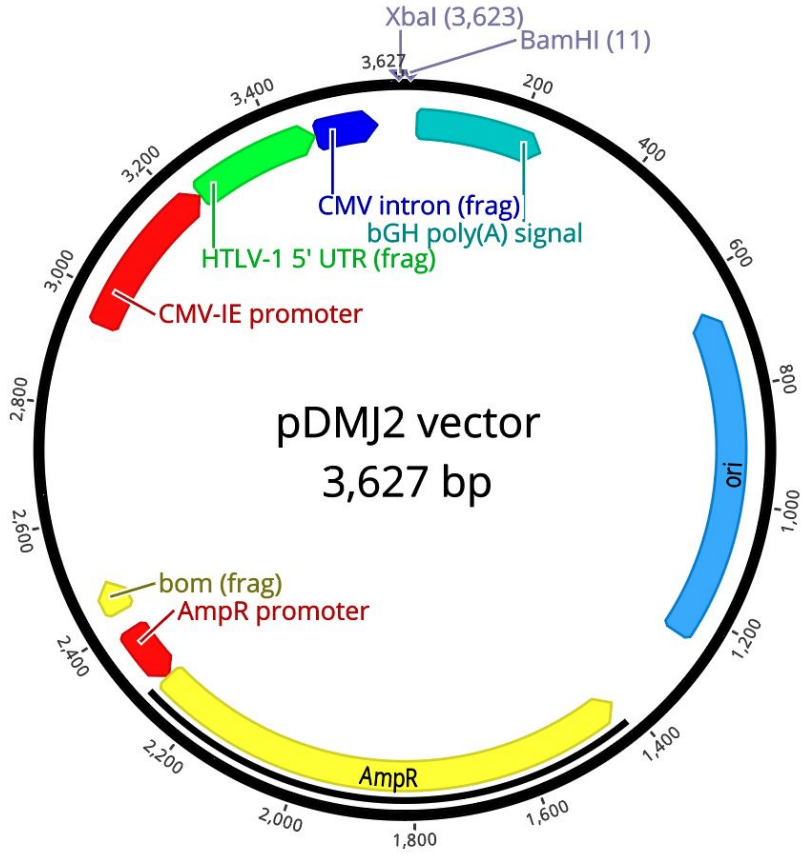


Figure S5: pDMJ2 plasmid map (3,627 bp). Annotations: CMV-IE promoter: Cytomegalovirus immediate early promoter; HTLV-1 5’ UTR (frag): human T lymphotropic virus type 1 (HTLV-1) 5’ UTR fragment, CMV intron (frag): Cytomegalovirus intron fragment; bGH poly(A) signal: bovine growth hormone polyadenylation signal; ori: ColE1 origin of replication; AmpR: Ampicillin resistance gene; AmpR promoter: Ampicillin resistance promoter; bom (frag): ColE1 basis of mobility fragment. *XbaI* and *BamHI* restriction sites.